

# Universidad de Oviedo

Programa de Doctorado en Materiales

---

Desarrollo de un sistema de reconstrucción tomográfica para óptica adaptativa solar mediante redes neuronales artificiales.

---

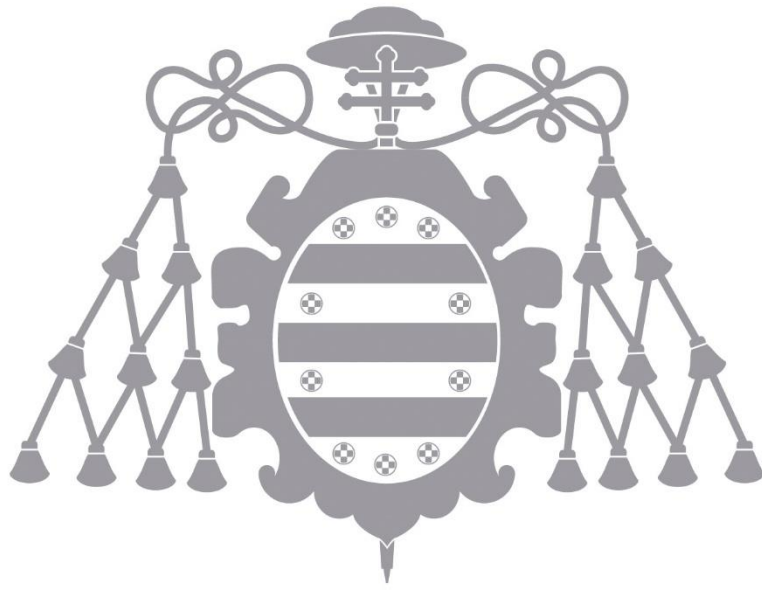
---

Development of a tomographic reconstruction system for solar adaptive optics using artificial neural networks.

---

TESIS DOCTORAL

FRANCISCO GARCÍA RIESGO



# Universidad de Oviedo

Programa de Doctorado en Materiales

---

Desarrollo de un sistema de reconstrucción tomográfica para óptica adaptativa solar mediante redes neuronales artificiales.

---

---

Development of a tomographic reconstruction system for solar adaptive optics using artificial neural networks.

---

TESIS DOCTORAL

Directores:

Jesús Daniel Santos Rodríguez

Sergio Luis Suárez Gómez

Tutor:

Víctor Manuel De La Prida Pidal



Universidad de Oviedo

## RESUMEN DEL CONTENIDO DE TESIS DOCTORAL

1.- Título de la Tesis	
Español/Otro Idioma: Desarrollo de un sistema de reconstrucción tomográfica para óptica adaptativa solar mediante redes neuronales artificiales.	Inglés: Development of a tomographic reconstructions system for solar adaptive optics using artificial neural networks.
2.- Autor	
Nombre: Francisco García Riesgo	DNI/Pasaporte/NIE: <del>74754000-1</del>
Programa de Doctorado: Doctorado en Materiales	
Órgano responsable: Departamento de Materiales.	

### RESUMEN (en español)

La investigación realizada a lo largo de la tesis se enmarca en el campo de la Óptica Adaptativa (OA), técnica que pretende corregir las aberraciones presentes en los frentes de onda recibidos por los telescopios situados en la superficie terrestre. Los frentes de onda provenientes de objetos celestes sufren perturbaciones en el camino desde la fuente emisora hasta ser recogidos por los telescopios, especialmente mientras atraviesan la atmósfera terrestre debido a la presencia de la turbulencia atmosférica. Uno de los principales objetivos de los sistemas OA es corregir las aberraciones en el menor periodo de tiempo posible, tratando de realizarlo en tiempo real.

La tesis se centrará en el entorno de los sistemas reconstructores investigando la viabilidad del empleo de redes neuronales para la reconstrucción en OA, concretamente en OA Solar.

La tesis se presenta mediante compendio de publicaciones, debido a las necesidades de diferentes planes de investigación que financiaron el estudio. Siguiendo las mismas se observa cronológicamente el desarrollo de la investigaciones así como los resultados obtenidos.

El primer paso que se llevó a cabo fue realizar un estudio para determinar qué tipo de red neuronal sería más ventajosa para realizar las investigaciones. Una vez conocida, los siguientes pasos que se llevaron a cabo fue la aplicación de este tipo de red neuronal en diferentes configuraciones de OA Solar.

Con el fin de determinar las configuraciones más óptimas de sistemas reconstructores para el trabajo con redes neuronales, se realizaron dos estudios en los que se analizaron qué datos de los diferentes recibidos por los telescopios serían los más adecuados, así como las configuraciones de los diferentes sensores del telescopio que en combinación con las redes neuronales artificiales optimizaran el funcionamiento del sistema reconstructor.

Todas las investigaciones llevadas a cabo se realizaron trabajando con datos de simulación, que permitía conocer la reconstrucción de la turbulencia atmosférica exacta y de esta forma llevar a cabo el entrenamiento de las redes neuronales.

Paralelamente a los trabajos mencionados anteriormente, se fueron llevando diferentes investigaciones variando parámetros como la configuración del sistema de OA o probando diferentes modelos de redes neuronales artificiales. Los resultados extraídos de estas investigaciones fueron expuestos en diferentes congresos internacionales o preparados para su posterior publicación.

FOR-MAT-VOA-010 (Reg. 2018)



Universidad de Oviedo

La tesis deja abiertas diferentes líneas de investigación continuistas con las investigaciones realizadas, que permitan el perfeccionamiento y nuevas aplicaciones de estos sistemas en diferentes configuraciones de sistemas de OA, así como la comprobación del correcto funcionamiento de los mismo en situaciones con telescopios reales.

#### RESUMEN (en Inglés)

The research carried out throughout the thesis is part of the field of Adaptive Optics (AO), a technique designed to correct the aberrations presented in the wavefronts received by telescopes located on the Earth's surface. The wavefronts coming from celestial objects suffer distortions on the way from the source to the telescope, especially while passing through the Earth's atmosphere due to the presence of atmospheric turbulence. One of the main objectives of AO systems is to correct the aberrations in the shortest possible time, trying to do it in real time.

This thesis is focused on the environment of reconstructor systems, investigating the feasibility of using neural networks for reconstruction in AO, specifically in Solar AO.

The thesis is presented as a compendium of publications, due to the needs of several research plans that financed the study. Following these publications, the development of the research and the results obtained can be followed chronologically.

The first step was to carry out a study to determine which type of neural network would be the most suitable to carry out the research. Once known, the next steps that were carried out were the application of this kind of neural network in different configurations of AO Solar.

In order to determine the most optimal reconstructor system configuration for working with neural networks, two studies were carried out analyzing which of the different data received by the telescopes would be the most suitable, as well as the configurations of the different telescope sensors that in combination with the artificial neural networks would optimize the performance of the reconstructor system.

All the researches have been made working with simulation data, which allowed to know the exact atmospheric turbulence reconstruction at each moment allowing the training of the neural networks.

Parallel to the mentioned researches, several investigations were made varying parameters such as the configuration of the OA system or testing different models of artificial neural networks. The results of these investigations were presented in different international congresses or prepared for later publications.

The thesis leaves open different lines of research that continue with the investigations carried out, allowing the improvement and new applications of these systems in different configurations of OA systems, as well as the verification of the correct operation in situations with real telescopes.

SR. PRESIDENTE DE LA COMISIÓN ACADÉMICA DEL PROGRAMA DE DOCTORADO EN MATERIALES

Oviedo a 28 de MARZO de 2023



## **Agradecimientos**

*Ante todo, me gustaría dedicar esta tesis a Dani, principal responsable de que me haya embarcado en esta aventura. Gracias por confiar en mí desde el principio, por darme la oportunidad y el impulso para iniciarme en este mundo. Me gustaría agradecer en particular su comprensión, su apoyo y su ofrecimiento continuo para orientar a tantos y tantos compañeros. Obviamente, durante cinco años se han compartido momentos buenos y no tan buenos, pero él me ha apoyado en todos ellos.*

*De la mano de Dani va Javi, promotor de mi primer contrato profesional, como del de muchos compañeros que, por su labor y dedicación, han salido adelante.*

*¡Como no!, dedico asimismo esta tesis al grupo de investigación que me ha visto crecer, el MOMA, y a todos sus integrantes, que, cuando han podido, no han dudado en echarme una mano.*

*Gracias, Carlos; gracias, Kike; gracias, Fernando. Mención destacada merece Sergio, compañero en un inicio y hoy amigo y codirector de esta tesis. Ha sido corresponsable, de que este trabajo haya llegado a buen puerto, empleando para ello tiempo, esfuerzo y comprensión, la propia del conocedor de las horas difíciles de la carrera del investigador. Gracias.*

*A mi familia, que siempre sigue de cerca y con ilusión mis progresos: en especial, a mi abuela, siempre pendiente y orgullosa de todos nosotros, y a mis padres, por su apoyo incondicional, por darme esa fuerza y motivación que no siempre se tiene cuando las cosas no salen bien, por enseñarme lo que es el respeto y los valores y por ser ejemplo de cómo avanzar en la vida sin perderlos nunca. No habría llegado a donde estoy ahora sin ellos.*

*A mis amigos, a todos ellos, que siempre están ahí, por tantos buenos momentos de reunión viéndonos avanzar y prosperar, cada uno siguiendo su camino: a Roque, a Rosete, a Guille, a Luis, a Claudia, a Ana, a Lucas, a Efrén, a Andrea, a Esther, a Héctor, a Blanca. Gracias. Para finalizar, pero muy importante, quiero dedicar este trabajo a Lucía, por su apoyo y su paciencia, por obligarme a confiar en mí, por ser ejemplo de cómo luchar por las metas sin desistir, y por siempre sacar tiempo de donde no lo había para conseguir que el camino recorrido estos años se hiciera más fácil. A todos ellos, nuevamente, Gracias.*



# 1 ÍNDICE

---

<b>2 índice de figuras.....</b>	<b>4</b>
<b>3 Introducción .....</b>	<b>5</b>
<b>4 La óptica adaptativa .....</b>	<b>7</b>
4.1 La transmisión de la luz en la atmósfera .....	8
4.1.1 Modelo de Kolmogorov.....	9
4.1.2 Propagación de la luz en la atmósfera.....	14
4.1.3 Variaciones estadísticas del campo.....	15
4.2 Componentes de un sistema de óptica adaptativa.....	17
4.2.1 Sensor de frente de onda .....	17
4.2.2 Sistemas correctores del frente de onda.....	24
4.2.3 Sistemas reconstructores del frente de onda .....	27
4.2.4 Sistemas de control.....	31
4.3 Configuraciones sistemas de óptica adaptativa.....	32
4.3.1 SCAO ( <i>Single Conjugated Adaptive Optics</i> ).....	32
4.3.2 GLAO ( <i>Ground-Layer Adaptive Optics</i> ) .....	33
4.3.3 MCAO ( <i>Multi-Conjugated Adaptive Optics</i> ).....	34
4.3.4 MOAO ( <i>Multi-Object Adaptive Optics</i> ).....	34
4.4 Óptica adaptativa solar .....	35
4.4.1 Cambios en la energía de la turbulencia .....	35
4.4.2 Cambios en la longitud del frente de onda.....	36
4.4.3 El sensor de frente de onda Shack-Hartmann Solar .....	37
4.4.4 Algunos sistemas de OA Solar en funcionamiento actualmente.....	39
4.5 Simulador de óptica adaptativa – Plataforma DASP .....	41
<b>5 Fundamentos de Inteligencia Artificial. ....</b>	<b>43</b>
5.1 Las redes neuronales artificiales (ANN).....	43

5.2	El Perceptron multicapa ( <i>Multi-Layer Perceptron</i> ).....	51
5.3	Redes neuronales convolucionales ( <i>Convolutional Neural Networks, CNN</i> ).....	53
5.4	Redes neuronales completamente convolutivas ( <i>Fully-Convolutional Neural Networks, FCN</i> ).....	57
5.5	Frameworks de redes neuronales artificiales .....	58
<b>6</b>	<b>Objetivos .....</b>	<b>59</b>
<b>7</b>	<b>Discusión de los resultados publicados .....</b>	<b>61</b>
7.1	Comprobación y elección de diferentes modelos de ANN para reconstrucciones de óptica adaptativa nocturna. ....	61
7.1.1	Resultados.....	63
7.1.2	Discusión de resultados.....	65
7.2	Reconstrucción de turbulencia atmosférica utilizando redes completamente convolucionales FCN. Comparación de las reconstrucciones en función de los parámetros de entrada escogidos. ....	66
7.2.1	Resultados.....	67
7.2.2	Discusión de resultados.....	69
7.3	Estudio de diferentes configuraciones de sensores Shack-Hartmann para reconstrucciones de turbulencia atmosférica en OA Solar.....	72
7.3.1	Resultados.....	74
7.3.2	Discusión de resultados.....	76
<b>8</b>	<b>Discusión de los resultados de congresos internacionales .....</b>	<b>79</b>
8.1	Reconstructor convolucional CARMEN para óptica adaptativa nocturna.....	79
8.1.1	Resultados.....	80
8.1.2	Discusión de resultados.....	81
8.2	Reconstrucción de fase turbulenta mediante redes completamente convolucionales en OA Solar.....	82
8.2.1	Resultados.....	82
8.2.2	Discusión de resultados.....	83
8.3	Reconstrucción mediante redes neuronales para un sistema con la configuración GLAO Solar.....	83



8.3.1	Resultados.....	84
8.3.2	Discusión de resultados.....	85
<b>9</b>	<b>Otros resultados obtenidos pendientes de publicación .....</b>	<b>87</b>
9.1	Reconstructor basado en redes neuronales artificiales para OA Solar capaz de generalizar a diferentes regiones solares. ....	87
9.1.1	Resultados.....	88
9.1.2	Discusión de resultados.....	89
<b>10</b>	<b>Conclusiones y líneas futuras .....</b>	<b>91</b>
<b>11</b>	<b>Copia completa de los artículos publicados .....</b>	<b>95</b>
11.1	Overview and Choice of Artificial Intelligence Approaches for Night-Time Adaptive Optics Reconstruction.....	95
11.2	Fully Convolutional Approaches for Numerical Approximation of Turbulent Phases in Solar Adaptive Optics .....	115
11.3	Comparative study of Shack-Hartmann configurations for atmospheric turbulence reconstructions in solar adaptive optics.....	137
<b>12</b>	<b>Informe del factor de impacto .....</b>	<b>149</b>
<b>13</b>	<b>Referencias .....</b>	<b>151</b>

## 2 ÍNDICE DE FIGURAS

---

<i>Figura 1</i> .....	20
<i>Figura 2</i> .....	22
<i>Figura 3</i> .....	23
<i>Figura 4</i> .....	25
<i>Figura 5</i> .....	28
<i>Figura 6</i> .....	33
<i>Figura 7</i> .....	33
<i>Figura 8</i> .....	34
<i>Figura 9</i> .....	35
<i>Figura 10</i> .....	36
<i>Figura 11</i> .....	37
<i>Figura 12</i> .....	38
<i>Figura 13</i> .....	39
<i>Figura 14</i> .....	40
<i>Figura 15</i> .....	47
<i>Figura 16</i> .....	52
<i>Figura 17</i> .....	53
<i>Figura 18</i> .....	54
<i>Figura 19</i> .....	55
<i>Figura 20</i> .....	56
<i>Figura 21</i> .....	56
<i>Figura 22</i> .....	56
<i>Figura 23</i> .....	57
<i>Figura 24</i> .....	58
<i>Figura 25</i> .....	64
<i>Figura 26</i> .....	64
<i>Figura 27</i> .....	64
<i>Figura 28</i> .....	65
<i>Figura 29</i> .....	68
<i>Figura 30</i> .....	69
<i>Figura 31</i> .....	70
<i>Figura 32</i> .....	71
<i>Figura 33</i> .....	75
<i>Figura 34</i> .....	83
<i>Figura 35</i> .....	83
<i>Figura 36</i> .....	85
<i>Figura 37</i> .....	85
<i>Figura 38</i> .....	88
<i>Figura 39</i> .....	88

### 3 INTRODUCCIÓN

---

La presente tesis versa sobre el trabajo y estudio realizado, en el ámbito de la viabilidad del uso de las redes neuronales, como sistemas reconstructores en diferentes configuraciones de óptica adaptativa solar. Para ello, se han desarrollado modelos con diferentes estructuras y topologías, con el objetivo de ser, posteriormente, aplicadas en situaciones de telescopios reales.

La óptica adaptativa es una herramienta fundamental en cualquier sistema de observación actual, ya que se encarga de determinar y reconstruir la turbulencia atmosférica con el objetivo de mejorar la calidad de las imágenes recibidas, eliminando las aberraciones sufridas por los rayos de luz en su camino óptico. Este proceso debe realizarse de la manera más rápida posible, lo más cercano al tiempo real. Debido a esto, con el desarrollo de la tecnología, cada vez se utilizan algoritmos más complejos buscando mayor eficiencia computacionalmente y reconstrucciones más precisas.

El trabajo realizado durante esta tesis ha consistido en el estudio de las redes neuronales artificiales como posible alternativa a los citados algoritmos, determinando las ventajas que podría conllevar su utilización, así como los posibles inconvenientes. La inteligencia artificial y, en concreto, las redes neuronales artificiales, son modelos de matemática aplicada y computación capaces de modelar y predecir situaciones físicas. Estos sistemas ya se han aplicados en diferentes campos, demostrando ser una alternativa eficaz a los algoritmos de clasificación y predicción para la resolución de diferentes problemas. En el caso de la óptica adaptativa en particular, ya han sido capaces de demostrar, en observaciones nocturnas, ser una opción real a los reconstructores comunes, mientras que, para el homólogo solar, aún están pendientes de desarrollo.

Este documento recapitula, en primer lugar, conceptos básicos de óptica adaptativa y de redes neuronales artificiales, introduciendo los diferentes tipos de sistemas y configuraciones que existen en el estudio, dotando así de los fundamentos necesarios para comprender la investigación que se ha realizado. A continuación, se muestran los objetivos que se han perseguido en esta investigación para, posteriormente, exponer los resultados obtenidos y una breve discusión de cada uno de ellos.

La tesis se presenta bajo la modalidad de compendio de publicaciones; los capítulos finales recopilan de forma íntegra los artículos publicados.





## 4 LA ÓPTICA ADAPTATIVA

---

La Óptica Adaptativa (OA) es una rama de la Óptica que se basa en la corrección de las aberraciones que se producen en los frentes de onda recibidos; en el caso de la astronomía, por los telescopios terrestres [1]–[3]. Este conjunto de técnicas multidisciplinares se encuentra incluido dentro de la Óptica, al ser esta la ciencia que estudia la propagación de la luz a lo largo del espacio.

En esencia, cuando se recibe la luz proveniente de fuentes celestes, los frentes de onda han recorrido previamente un largo camino óptico, tras lo que se reciben por los telescopios. A lo largo de todo ese trayecto, pueden sufrir perturbaciones, pero, sin lugar a duda, existe una pequeña sección del camino donde aparecen la mayoría de las aberraciones, correspondiente con parte turbulenta de la atmósfera terrestre.

La atmósfera consiste en una mezcla de partículas en suspensión, cuyo comportamiento se asemeja a la dinámica de fluidos, formándose remolinos de aire de diferentes temperaturas y densidades, lo que provoca que la luz procedente de la fuente estudiada atraviese medios con índices de refracción variables. La situación sería similar a la de colocar un conjunto de lentes una detrás de otra, dando lugar a una diferencia de camino óptico. En ámbitos como la astronomía esta diferencia de camino óptico supone la pérdida de información del objeto observado y es por ello por lo que es tan importante comprender el efecto de la atmósfera en la luz. La función fundamental de las técnicas de óptica adaptativa es corregir la mayor parte de las aberraciones sufridas por los rayos de luz, en un tiempo lo más cercano posible al tiempo real.

Para llegar a ese objetivo, se debe comenzar partiendo de uno de los conceptos más elementales de la Óptica como es el principio de Fermat, donde se establece que los rayos de luz en el espacio seguirán la trayectoria que minimice su camino óptico. Si se aplica este principio en un medio isotrópico donde el índice de refracción sea constante para todos los puntos del espacio, la solución que se obtendría sería la línea recta entre los puntos inicial y final. En consecuencia, cuando los fotones se desplazan en medios con índice de refracción constante, estos describen trayectorias rectilíneas.

Generalmente en óptica no se suele trabajar con rayos, aunque la aproximación por rayos puede resultar útil en ámbitos relacionados con la óptica geométrica (sistemas de lentes, espejos, beamsplitters, etc.). En el caso de la óptica adaptativa, se deben tener en cuenta más propiedades ondulatorias de la luz a parte de su trayectoria, principalmente su fase.

Cualquier radiación electromagnética es solución de la siguiente ecuación de ondas (en este caso aproximada al vacío):

$$\nabla^2 E - \frac{\mu_0 \epsilon_0 \delta^2 E}{\delta t^2} = 0 \quad ; \quad \nabla^2 B - \frac{\mu_0 \epsilon_0 \delta^2 B}{\delta t^2} = 0 \quad (1)$$

La radiación es, por tanto, una superposición de soluciones de esta ecuación de onda [4]. Teniendo en cuenta solamente en el campo eléctrico, la solución vendría dada por:

$$E(\mathbf{r}, t) = E_0 e^{i(\vec{k} \cdot \vec{r} - \omega t + \phi)} \quad (2)$$

Siendo  $\vec{k}$  el vector de propagación de la onda,  $\vec{r}$  el vector de posición,  $E_0$  la amplitud del campo, y  $\omega = 2\pi f$  siendo  $f$  la frecuencia de la onda y  $\phi$  los posibles desfases que pueda sufrir

la onda. Conocido el campo eléctrico, el campo magnético de la onda tendrá la misma fase y será perpendicular a  $\vec{E}$  y  $\vec{k}$ .

Las aberraciones que se produzcan en los frentes de onda recibidos por los telescopios terrestres implicarán modificaciones en la fase de la onda, por lo que el objetivo de la OA será cambiar la fase del campo eléctrico para reestablecerla lo más similar a la original.

Los sistemas de óptica adaptativa cuentan con diversos componentes, siendo algunos de los más representativos: sensores de frente de onda, espejos deformables y sistemas de control. Estos se detallan posteriormente en sus respectivos apartados. Dichos componentes pueden aparecer en combinación con otros elementos como pueden ser divisores del frente de onda, divisores de haz, sistemas de control auxiliares o estrellas guía láser.

#### 4.1 LA TRANSMISIÓN DE LA LUZ EN LA ATMÓSFERA

La principal fuente generadora de aberraciones sobre los frentes de onda recibidos de cuerpos celestes es la atmósfera, debido al fenómeno de la turbulencia atmosférica provocada, fundamentalmente, por las variaciones del índice de refracción a lo largo del camino recorrido por la radiación. La mezcla de remolinos de aire con temperaturas y densidades diferentes que forma la atmósfera genera un efecto similar al de colocar un conjunto de lentes una detrás de otra, dando lugar a una diferencia de camino óptico entre un instante y el siguiente. El objetivo de los sistemas de óptica adaptativa consiste en la minimización de estos efectos en los frentes de onda recibidos en el menor tiempo posible, tratando de hacerlo en tiempo real.

El flujo del viento y las variaciones de temperatura que ocurren en la atmósfera de forma natural provocan cambios que se comportan como vórtices o remolinos. Diferentes procesos mecánicos que se producen entre sus capas y la inestabilidad térmica de las masas de aire, intensificada por el calentamiento diurno debido a la energía procedente del Sol, generan turbulencias que pueden persistir durante varios kilómetros. Precisamente, las fluctuaciones de la temperatura son el factor dominante de la turbulencia, ya que el índice de refracción del aire ( $n$ ) es principalmente función de la densidad  $\rho$  [5], que relaciona presión y temperatura a través de la ley de los gases ideales.

$$P \propto \rho T \quad (3)$$

De donde se deriva:

$$\delta \ln(n - 1) = \delta \ln(P) - \delta \ln(T) \quad (4)$$

Para velocidades mucho menores que la velocidad del sonido las fluctuaciones en la presión son despreciables, y se concluye que:

$$\delta n \simeq (1 - n) \frac{\delta T}{T} \quad (5)$$

Por tanto, la acumulación de las fluctuaciones de temperatura provoca la aparición de fluctuaciones en la fase y en la amplitud de las ondas, causando inhomogeneidades significativas en la radiación que acaba llegando a la superficie. Determinar de forma precisa el efecto de la turbulencia atmosférica sobre la luz en cada momento es un problema muy complejo; en primer lugar, debido a la descripción matemática de la atmósfera, que no es sencilla y, en segundo lugar, por la gran variabilidad que presenta con el tiempo. La modelización de la atmósfera debe estar variándose constantemente.

Existen diferentes modelos matemáticos que permiten modelizar el comportamiento de la atmosfera, entre los que se encuentran los modelos analíticos [6] y los modelos estadísticos [7], [8]. Los modelos analíticos se corresponden con descripciones matemáticas basadas en dinámica de fluidos, más concretamente en las ecuaciones de Navier-Stokes aplicadas a la atmósfera. En cambio, los modelos más precisos se basan en análisis estadísticos, ya que la complejidad de la atmósfera real está más allá de las capacidades de predicción determinista o incluso, el propio análisis numérico. La descripción estadística de la turbulencia ha permitido describir los efectos promedio generados por la turbulencia, como pueden ser la desviación o dispersión del haz incidente, mediante una serie de teorías o leyes de escala.

A continuación, se expondrá detalladamente el modelo estadístico de Kolmogorov, ya que se trata del modelo más utilizado en el diseño de sistema de óptica adaptativa, así como en los simuladores de la turbulencia atmosférica.

#### 4.1.1 Modelo de Kolmogorov

Andrey N. Kolmogorov, matemático ruso del siglo XX. y Aleksandr M. Obújov, físico y matemático ruso respectivamente, desarrollaron su teoría de la turbulencia atmosférica en base a argumentos de escalas. El principio de este modelo es la suposición de que las turbulencias se generan en cierta escala exterior  $L_0$ , y progresan a escalas más y más pequeñas debido a que los vórtices más grandes transfieren su energía a los más pequeños. Es un efecto descrito como “cascada de energía” hasta llegar a una escala mínima representada con el tamaño  $l_0$ , a partir de esa escala mínima las fuerzas viscosas producen la disipación de la energía restante en forma de calor [2], [4], [9], [10].

Es importante matizar que el modelo está desarrollado para aplicarlo en un estado pseudo-estacionario que se denomina rango inercial. Como se ha mencionado, las capas turbulentas reciben energía de capas mayores a la vez que disipan energía a otras inferiores, con el desarrollo de la turbulencia el ritmo de obtención y disipación de energía es aproximadamente constante, disipando las capas tanta energía como reciben. Esta situación es la que se denomina rango inercial.

La modelización de la turbulencia atmosférica es necesaria ya que permite conocer aproximadamente las propiedades de la atmósfera en las que se va a realizar una observación y prever los efectos que se van a producir sobre los haces incidentes. Además, es particularmente importante cuando se trabaja con datos simulados, ya que cuanto más precisos sean los modelos más próximos a la realidad serán los datos producidos.

El modelo de Kolmogorov estudia el comportamiento de la turbulencia atmosférica en rangos inerciales, para tamaños de turbulencia entre  $L_0$  y  $l_0$ . La principal ventaja del modelo es que, a partir de parámetros como temperatura, velocidad del viento de las capas, densidad, etc. que son, a priori, sencillos de medir con los sistemas presentes en los telescopios, es capaz de modelar la turbulencia atmosférica.

Kolmogorov estudió la diferencia entre velocidades cuadráticas medias entre dos puntos separados por una distancia  $r$ . Se definió el tensor de estructura  $D_{ij}$  como [2], [5], [7]:

$$D_{ij} = \langle [v_i(\vec{r}_1 + \vec{r}) - v_i(\vec{r}_1)] [v_j(\vec{r}_1 + \vec{r}) - v_j(\vec{r}_1)] \rangle \quad (6)$$

Siendo  $v_i, v_j$  las diferentes componentes de la velocidad de los puntos. En general, esta expresión es bastante complicada de tratar, aunque ciertas simplificaciones acerca de la atmósfera permiten conseguir un tensor de estructura más sencillo. Considerando condiciones de rango inercial, es

decir, homogeneidad local (la velocidad solo depende de  $r$ ), isotropía local (la velocidad solo depende del módulo de  $r$ ) e incompresibilidad de las turbulencias ( $\nabla \cdot v = 0$ ) se simplifica la definición del tensor de estructura en función de la velocidad como sigue:

$$D_v = \langle [v(\vec{r}_1 + \vec{r}) - v_r(\vec{r}_1)]^2 \rangle = C_v^2(h)r^{\frac{2}{3}} \quad (7)$$

Para llegar al segundo resultado se consideró que la separación entre los puntos es pequeña, por lo que queda un tensor de estructura únicamente dependiente en  $r$ , donde  $C_v^2$  es la constante de estructura para la velocidad, siendo una forma de medir la energía contenida en la turbulencia. En realidad, el tensor de estructura de las turbulencias atmosféricas se define como una cantidad dependiente del índice de refracción, ya que la velocidad de las turbulencias atmosféricas causa variaciones de presión y temperatura que alteran de manera sustancial el valor de  $n$ .

La constante solamente es válida en condiciones de aplicación del modelo de Kolmogorov, siendo el valor  $r$  de distancia entre puntos es tal que  $l_0 < r < L_0$ , donde  $l_0$  es la escala más pequeña y  $L_0$  es la escala mayor. El vórtice más pequeño (en inglés, *smaller eddy*) es el tamaño más pequeño por debajo del cual los efectos viscosos son importantes y la energía se disipa en forma de calor. El tamaño de la escala pequeña está relacionado con la viscosidad ( $\eta$ ) y la energía ( $\varepsilon$ ) de la turbulencia a través de la fórmula siguiente [4]:

$$l_0 = 7,4 \cdot \left( \frac{\eta^3}{\varepsilon} \right)^{\frac{1}{4}} \quad (8)$$

El rango de valores comunes que suele tomar este parámetro  $l_0$  varía en función de la región de la atmósfera en la que se esté observando, desde milímetros en la superficie hasta algunos centímetros en regiones más externas como la troposfera. Por otro lado, la escala  $L_0$  sigue un comportamiento cerca de la superficie  $L_0 \approx 0,4h$ , donde  $h$  es la altura. En capas más externas de la atmósfera, la escala  $L_0$  oscila entre valores de entre 10 y 100 metros.

A partir de la constante de estructura de la velocidad, puede el tensor de estructura asociado al índice de refracción [7]:

$$D_n = \langle |n(\vec{r}_1 + \vec{r}) - n(\vec{r}_1)|^2 \rangle = C_n^2(h)r^{\frac{2}{3}} \quad (9)$$

Este tensor depende de  $C_n^2(h)$ , la constante de estructura del índice de refracción, también conocido como perfil de la turbulencia atmosférica y al igual que la anterior, da información de la fuerza o intensidad de la turbulencia a una altura  $h$ .

Las funciones de densidad espectral (en inglés, *power spectrum*) aparecen en el contexto de la Teoría Espectral de la señal, desarrollada por Fourier para el estudio de las diferentes componentes de frecuencia de una señal cualquiera. En ciertas situaciones es conveniente trabajar en el dominio espectral en una escala espacial [5], o en su defecto, por su recíproco el número de ondas  $\vec{k}$ . Para la turbulencia isotrópica de Kolmogorov se define la función de estructura como:

$$D_n = 8\pi \int \Phi_n(k) \left[ 1 - \frac{\sin(kr)}{kr} \right] k^2 dk \quad (10)$$



Dado que se pasa del dominio del tiempo al dominio de frecuencias, se obtiene el espectro de frecuencias que permite, con su aplicación, obtener la densidad de potencia espectral  $\Phi_n(k)$  cuya expresión es la siguiente:

$$\Phi_n(K) = \frac{1}{(2\pi)^3} \int B_n(r_1) e^{-ikr_1} dr_1^3 = \frac{5}{18\pi} C_n^2 K^{-3} \int_{l_0}^{L_0} \text{sen}(Kr_1) r_1^{-\frac{1}{3}} dr_1 \quad (11)$$

con  $\vec{K} = (K, \theta, \varphi)$  el vector de onda y considerando como límite  $l_0 \rightarrow 0$  y  $L_0 \rightarrow \infty$ , se obtiene el espectro de Kolmogorov [2]:

$$\Phi(k) = \frac{\Gamma\left(\frac{8}{3}\right) \text{sen}\left(\frac{\pi}{3}\right)}{4\pi^2} C_n^2 k^{-\frac{11}{3}} \approx 0.033 C_n^2 |K|^{-\frac{11}{3}} \quad (12)$$

Gracias a la función de densidad de potencia espectral se puede conocer en función de la longitud de onda y la posición de la turbulencia, la distribución de energía de la turbulencia.

Toda esta modelización se corresponde con el modelo de Kolmogorov. Como se ha comentado al inicio, el modelo se encuentra limitado para turbulencias de escala superiores a  $L_0$ . Para solucionar este problema, Theodore Von Kármán, físico e ingeniero húngaro, propuso tener en cuenta los efectos de las turbulencias mayores que la escala  $L_0$  en la densidad espectral de la radiación. Ello le permitió concluir una expresión más general para el poder espectral, definida como sigue [5]:

$$\Phi(k) = 0.033 C_n^2 (k^2 + k_0^2)^{-\frac{11}{6}} \quad (13)$$

Donde  $k_0 = 2\pi/L_0$ , con  $L_0$  el tamaño de la escala exterior. Es inmediato notar que cuando en el rango de escalas adecuado, el modelo de Von Kármán generalizado recupera las expresiones iniciales del modelo de Kolmogorov.

Una vez estudiada la turbulencia atmosférica el objetivo es estudiar cómo afecta sobre los rayos que se están propagando a través de ella. Cuando las ondas electromagnéticas recorren la turbulencia atmosférica reducen su grado de coherencia, ya que en el camino se van encontrando con diferentes índices de refracción, provocando que cada frente de onda recorra un camino óptico diferente. Fried determinó que existe una longitud característica, llamada longitud de coherencia, que es el máximo tamaño permitido para la apertura sin que se rompa la coherencia. Se dice que la luz es coherente cuando toda la luz recibida tiene la misma amplitud de onda y la misma fase, lo que es esperable cuando proviene de una misma fuente. Una apertura mayor que la longitud de coherencia hará que se produzca la dispersión de la luz. La longitud de coherencia para un camino en el que la constante de estructura depende de la altura respecto de la superficie terrestre se escribe como [2]:

$$r_0 = \left[ 0.423 k^2 \sec\beta \int_C C_n^2(h) dh \right] \quad (14)$$

Donde  $C$  es el camino seguido por la luz y  $\beta$  es un factor corrector que se corresponde con el ángulo cenital de la observación. La corrección se aplica para tener en cuenta los cambios en la longitud de la trayectoria de viaje de la línea de visión con el ángulo de observación. Esta expresión es válida cuando se trabaja con ondas planas, para ondas esféricas habría que considerar más factores. Sin embargo, los frentes de onda recibidos de fuentes celestes pueden

aproximarse por frentes de onda planos ya que, pese a que inicialmente se traten de frentes de ondas esféricos, las secciones de ese frente de onda que se reciben a una distancia tan lejana cumplen esa aproximación.

El parámetro de Fried  $r_0$  es una medida de la calidad de la transmisión óptica a través de la atmósfera debido a cambios aleatorios en el índice de refracción del aire que compone la atmósfera. En general se define  $r_0$  como el diámetro de un área circular sobre la cual el error cuadrático medio del frente de onda debido al paso a través de la atmósfera es igual a 1 radián.

El parámetro de Fried  $r_0$  tiene una gran importancia en la OA ya que es el valor que se utiliza como referencia a la hora de determinar de una forma rápida la intensidad de una turbulencia. Tanto a lo largo de los estudios realizados durante este trabajo como en situaciones de observación real, el valor de  $r_0$  es el que se utiliza para caracterizar la turbulencia. Su valor suele variar entre unos pocos *cm*, alrededor de 4 – 5 *cm* que se correspondería con un día de muy alta turbulencia, tormenta, etc. hasta los 20 *cm* aproximadamente que sería una situación nocturna ideal para la observación.

Desde un punto de vista matemático, el parámetro de Fried se presenta como un parámetro extremadamente útil para reescalar leyes relacionadas con fenómenos atmosféricos ya que incorpora la fuerza de la turbulencia ( $C_n$ ), la longitud de onda  $\lambda$  y el camino de propagación en un único parámetro [2]. Un ejemplo de ello es el espectro de la turbulencia en el dominio espacial. Con el parámetro de Fried medido en metros y la frecuencia espacial  $\xi$  en  $m^{-1}$  la expresión del espectro [11]–[13] es:

$$\Phi(\xi) = 0.023 \frac{\xi^{-\left(\frac{11}{3}\right)}}{r_0^{\frac{5}{3}}} \quad (15)$$

Este espectro puede ser empleado para describir los modos de aberración de la atmósfera turbulenta en términos de los coeficientes de Zernike [7], que se detallarán posteriormente. Además, la varianza total del frente de onda se puede obtener a partir de integrar tal que:

$$\sigma^2 = \int \Phi(\xi) d^2\xi \quad (16)$$

El tiempo en el que las aberraciones del frente de onda cambian de manera notable es otro de los parámetros claves. Este tiempo se puede calcular asumiendo que la turbulencia que atraviesa la apertura del telescopio es movida por viento a escalas de tiempo mucho menores que la variación de la turbulencia [13]. Esta suposición se conoce como hipótesis de Taylor, o de flujo congelado [14]. La hipótesis de Taylor da lugar a que el tiempo que delimita el tiempo de exposición, o tiempo de coherencia ( $\tau_0$ ), dado por:

$$\tau_0 = 0.314 \frac{r_0}{\bar{v}^{\frac{2}{3}}} \quad (17)$$

donde se denota con la barra y el subíndice el promedio de  $v$  ponderado en la turbulencia, siendo  $v$  la velocidad del viento de la capa de turbulencia dominante.

En el caso de que la velocidad del viento sea rápida, los remolinos de viento tardan menos en atravesar la apertura del telescopio, de manera que la frecuencia de dichos cambios es alta

[14]. Se establece entonces una frecuencia de referencia para determinar si la frecuencia de cada momento es alta o baja, a esta frecuencia se le denomina Frecuencia de Greenwood y viene determinada por:

$$f_G = 2.31\lambda^{-\frac{6}{5}} \left[ \sec(\beta) \int C_n^2(z) v^{\frac{5}{3}} dz \right]^{\frac{3}{5}} \quad (18)$$

Siendo  $\lambda$  la longitud de la onda que esté atravesando la turbulencia.

Otro aspecto importante a tener en cuenta son las posibles fluctuaciones en la fase que se pueden producir entre los diferentes caminos ópticos que puede recorrer la luz. La importancia de estas fluctuaciones aumenta en sistemas de óptica adaptativa que trabajan con un gran campo de visión, pues la corrección realizada no será óptima para todas las posibles direcciones de propagación. Este efecto se conoce como anisoplanatismo. Se define el ángulo isoplanático  $\theta_0$  para una altura  $h$  como:

$$\theta_0 = \left[ 2.914k^2 \sec^{\frac{8}{3}}(\beta) \int C_n^2(z) h^{\frac{5}{3}} dz \right]^{-\left(\frac{3}{5}\right)}$$

Este ángulo determina la región angular en los frentes de onda atraviesan la misma turbulencia. Puede expresarse en función del parámetro  $r_0$  con la misma notación que en la ecuación 18 como:

$$\theta_0 = 0.314 \frac{r_0}{h^{5/3}} \quad (19)$$

De este modo se ha obtenido un conjunto de valores que permiten, a partir de mediciones rápidas que se realizan en la superficie de la Tierra, caracterizar la atmósfera. Frecuentemente las condiciones de la atmósfera se representan con el valor de  $r_0$ , ya que todos los demás parámetros pueden expresarse en función de él. Por tanto, conocido el valor de  $r_0$ , la velocidad del viento en la superficie, y la temperatura en la zona cercana al objetivo del telescopio, puede modelizarse la atmósfera en la zona del objeto a observar, con el fin de utilizarla a continuación en la corrección de las imágenes.

Otro parámetro utilizado frecuentemente en óptica adaptativa es la razón de Strehl, que se define como la relación de la intensidad del pico real de la imagen de una fuente puntual, frente a la intensidad del pico de la imagen en un telescopio perfecto de misma apertura. Se expresa de la siguiente forma:

$$S = \exp(-\sigma^2) \quad (20)$$

Siendo  $\sigma^2$  es el cuadrado del error medio en la imagen. Su valor frecuentemente es utilizado para conocer la calidad de la reconstrucción realizada por el sistema de óptica adaptativa, cuanto más cercano sea el valor a la unidad, mejor será la reconstrucción.

Otro aspecto importante a tener en cuenta son las posibles fluctuaciones en la fase que se pueden producir entre los diferentes caminos ópticos que puede recorrer la luz. La importancia de estas fluctuaciones aumenta en sistemas de óptica adaptativa que trabajan con

grandes campos de visión, pues la corrección realizada no sería óptima para todas las posibles direcciones de propagación. Este efecto se conoce como anisoplanatismo. Se define el ángulo isoplanático

#### 4.1.2 Propagación de la luz en la atmósfera

Una vez que se han modelizado las propiedades de la atmósfera, es necesario conocer cómo van a afectar las fluctuaciones del índice de refracción en la propagación de la luz. No existe una teoría analítica que exprese la propagación de la luz a través de un medio turbulento, no obstante, existen aproximaciones que se pueden considerar válidas en la mayoría de las situaciones que se dan en los grandes telescopios terrestres.

Por simplicidad en las ecuaciones, solamente se analizará el campo eléctrico. Lo mismo sucederá en el campo magnético, perpendicular al campo eléctrico analizado y también al vector de ondas  $\vec{k}$ . El vector de posición  $\vec{r}$  puede descomponerse en dos componentes:  $r$ , que se corresponde con los vectores perpendiculares al plano de propagación de la onda y  $z$ , que se corresponde con la dirección de propagación de la radiación de la radiación. Por tanto,  $\vec{r} = (r, z)$

Asumiendo que la turbulencia es isotrópica, se puede llevar a cabo un tratamiento escalar de la propagación de la onda, ya que la atmósfera no afecta a la polarización de la luz [5]. Luego, la ecuación de ondas a resolver será:

$$\nabla^2 E(r, z, t) = \frac{N^2}{c^2} \delta_t^2 E(r, z, t) \quad (21)$$

Siendo  $E$  el vector campo eléctrico,  $N$  el índice de refracción del medio y  $c$  la velocidad de la luz. Como es conocido, la solución general de la ecuación es:

$$E(\vec{r}, t) = \psi(\vec{r}) e^{i(\vec{k} \cdot \vec{r} - \omega \cdot t)} \quad (22)$$

Siendo  $\vec{k}$  el vector de ondas de módulo  $\frac{2\pi}{\lambda}$  y  $\psi(\vec{r})$  la amplitud correspondiente con el módulo del campo eléctrico. Dadas las componentes del vector  $\vec{r}$ , se divide el operador laplaciano en componente transversal y longitudinal:

$$[\nabla^2 + 2i\vec{k}\nabla + \vec{k}^2(N^2 - 1)]\psi(\vec{r}) = 0 = \left[ \nabla_r^2 + \frac{\delta^2}{\delta z^2} + 2ik \frac{\delta}{\delta z} + k^2(N^2 - 1) \right] \psi(\vec{r}, z) \quad (23)$$

Esta ecuación puede simplificarse asumiendo que:

1.  $\psi(\vec{r}, z)$  cambia lentamente a medida que la onda se propaga. Los cambios más significativos se producen cuando  $L \gg \lambda$ , el segundo término de esta ecuación es mucho más pequeño que el tercero y es generalmente despreciable [2].
2. Dado que  $n = N - 1 \ll 1$  se tiene utilizando la aproximación paraxial que  $N^2 - 1 \approx 2n$ , de ella se obtiene la ecuación de ondas paraxial:

$$\left[ \nabla_r^2 + 2ik \frac{\delta}{\delta z} + 2k^2 n \right] \psi(\vec{r}, z) = 0 \quad (24)$$

La resolución de esta ecuación pasa por conocer el índice de refracción en todos los puntos. Sin embargo, en múltiples situaciones es suficiente conocer propiedades estadísticas del campo en la posición del observador [5], [14]. La propiedad estadística más simple es la función de coherencia de primer orden que se define como el valor promedio del campo  $\Psi$ . El valor

promedio del campo después de la propagación a través de la atmósfera hasta el observado, situado en  $z = 0$  viene dado por:

$$\langle \psi \rangle = \psi_0 \exp \left[ -2\pi^2 k^2 \int_0^\infty dz \int_0^\infty \phi_n(\vec{k}, z) \vec{k} d\vec{k} \right] \quad (25)$$

Donde  $\psi_0$  es el campo justo encima de la atmósfera,  $\kappa$  es el vector de ondas bidimensional (direcciones transversales al movimiento) y  $\Phi_n$  es la PSD definida en el modelo de Kolmogorov.

Hay que mencionar que el hecho de integrar  $z$  en el intervalo  $[0, h]$  o en el intervalo  $[0, \infty]$  no introduce ningún error [14]. Al tratarse de una exponencial negativa este resultado da cuenta de la reducción de la amplitud promedio de la parte coherente de la radiación debida a la difracción que provoca la fluctuación del índice de refracción [5]. La energía total se conserva pero la luz incoherente se distribuye sobre un ángulo mayor [14].

Volviendo sobre la resolución de la ecuación de ondas paraxial ( 24 ):

$$[\nabla_r^2 + 2k^2 n] \psi(\vec{r}, z) = -2ik \frac{\delta}{\delta z} \psi(\vec{r}, z) \quad (26)$$

Dividiendo por  $2k^2$ :

$$-\left[ \frac{1}{2k^2} \nabla_r^2 + n(\vec{r}, z) \right] \psi(\vec{r}, z) = \frac{i}{k} \frac{\delta}{\delta z} \psi(\vec{r}, z) \quad (27)$$

Se obtiene una ecuación formalmente idéntica a la ecuación de Schrödinger, donde la coordenada  $z$  juega el papel del tiempo,  $\lambda$  hace el papel de la constante de Planck y  $-n(r, z) = V(r, z)$ . Integrando en la variable  $z$  se obtiene una solución para  $\psi(\vec{r}, z)$ :

$$\psi(\vec{r}, z) = \psi(\vec{r}, 0) + \frac{i}{k} \int_0^z \left( \frac{1}{2k^2} \nabla_r^2 + n(\vec{r}, u) \right) \psi(\vec{r}, u) du \quad (28)$$

Esta expresión se corresponde claramente con una ecuación de Fredholm de segunda especie [15] con el término independiente una función dependiente de la variable  $r$ . El procedimiento de resolución de una ecuación de este tipo es complejo, ya que requiere de uso del formalismo del resolvente, una parte de teoría de operadores en ecuaciones integrales.

La solución de la ecuación se expresa como la serie de potencias siguiente [4]:

$$\psi = \psi_0 + \sum_{p=1}^{\infty} (ik)^p \int_0^z dz_0 \dots \int_0^{z_{p-2}} dz_{p-1} \prod_{i=1}^p \exp \left( i(z_i - z_{i-1}) \frac{\nabla_r^2}{2k} \right) n_{i-1} \psi_0 \quad (29)$$

Correspondiéndose  $\psi_0 = \psi(\vec{r}, 0)$ . En general, esta última ecuación puede escribirse como un desarrollo en serie de Fourier en función del índice de refracción. Esta serie permite analizar la propagación de la radiación a lo largo de la dirección  $z$ , así como su promedio y la desviación estándar de este último.

#### 4.1.3 Variaciones estadísticas del campo

Para obtener otras propiedades estadísticas como la varianza de las fluctuaciones en la amplitud o en la fase, es útil considerar el efecto en la amplitud y en la fase por separado [14].

Introduciendo  $\chi$  y  $\varphi$  que se corresponden con la amplitud logarítmica y las fluctuaciones de la fase respectivamente, en base a la definición de  $\Psi$ :

$$\psi = \psi_0 \exp(\chi + i\varphi) \quad (30)$$

Aplicando que  $\langle e^x \rangle = e^{\frac{(\alpha-x)^2}{2+\alpha}}$ , relación válida para cualquier variable aleatoria que siga una distribución gaussiana [2], se obtiene que la suma de las varianzas de la amplitud logarítmica ( $\sigma_x^2$ ) y de fluctuaciones de fase ( $\sigma_\varphi^2$ ) es:

$$\sigma_x^2 + \sigma_\varphi^2 = 4\pi^2 k^2 \int_0^\infty dz \int_0^\infty \phi_n(\kappa, z) \kappa d\kappa \quad (31)$$

La covarianza entre valores de campo medidos en las posiciones  $\vec{r}_1$  y  $\vec{r}_2$  en el plano transversal a la dirección del movimiento, viene dada por la función de coherencia de segundo orden ( $\Gamma_2$ ), también conocida como función de coherencia mutua [14]. Su expresión es la siguiente:

$$\Gamma_2(\vec{r}_1, \vec{r}_2) = \langle \psi_1(\vec{r}_1) \psi_2^*(\vec{r}_2) \rangle = I_0 \exp \left[ -4\pi^2 k^2 \int_0^\infty dz \int_0^\infty \phi_n(\kappa, z) [1 - J_0(\kappa r)] \kappa d\kappa \right] \quad (32)$$

Siendo  $r\chi = |\vec{r}_1 - \vec{r}_2|$ ,  $I_0 = |\psi_0|^2$  y  $J_0$  es la función de Bessel de primera especie. Sustituyendo la expresión (30) se obtiene que:

$$D_X(r) + D_\varphi(r) = 8\pi^2 k^2 \int_0^\infty dz \int_0^\infty \phi_n(\kappa, z) [1 - J_0(\kappa r)] \kappa d\kappa \quad (33)$$

Siendo  $D_Y$  la función de estructura de amplitud logarítmica o la función de estructura de fase según el caso. En el caso del modelo de Kolmogorov, este resultado es integrable, obteniéndose la constante de estructura de la radiación [7]:

$$D(r) \equiv D_X(r) + D_\varphi(r) = 2 \left[ \frac{24}{5} \Gamma\left(\frac{6}{5}\right) \right]^{\frac{5}{6}} \approx 6.884 \left( \frac{r}{r_0} \right)^{\frac{5}{3}} \quad (34)$$

Siendo  $\Gamma(x)$  la función Gamma. En la práctica la función de estructura de la amplitud logarítmica es mucho más pequeña que la función de estructura de fase [5], [14] y este resultado se reduce a la expresión:

$$D_\varphi \approx 6.884 \left( \frac{r}{r_0} \right)^{\frac{5}{3}} \quad (35)$$

Para separar los dos tipos de fluctuaciones es necesario obtener la función de coherencia de cuarto orden, que describe la correlación entre las fluctuaciones de la irradiancia (potencia por unidad de área) [14] y se define de la siguiente manera:

$$\Gamma_4 = \langle \psi_1(\vec{r}_1, z) \psi_1^*(\vec{r}_1, z) \psi_2(\vec{r}_2, z) \psi_2^*(\vec{r}_2, z) \rangle = \langle I_1(\vec{r}_1) I_2(\vec{r}_2) \rangle \quad (36)$$

Las soluciones analíticas encontradas hasta la fecha han sido para fluctuaciones de la amplitud logarítmica pequeñas en las que  $\sigma_X \leq 0.35$  [14]. En este límite se encuentra que:

$$\Gamma_4 = I_0^2 \exp \left[ 16\pi^2 k^2 \int_0^\infty dz \int_0^\infty \phi_n(\kappa, z) \cdot J_0(\kappa r) \cdot \sin^2 \left( \frac{z\kappa^2}{2k} \right) \kappa d\kappa \right] \quad (37)$$

Sustituyendo con (30) se obtiene la covarianza de  $\chi$ :

$$\begin{aligned} C_\chi(\vec{r}_1, \vec{r}_2) &= \langle |\chi(\vec{r}_1) - \langle \chi \rangle| |\chi(\vec{r}_2) - \langle \chi \rangle| \rangle \\ &= 4\pi^2 k^2 \int_0^\infty dz \int_0^\infty \phi_n(\kappa, z) \cdot J_0(\kappa r) \cdot \sin^2 \left( \frac{z\kappa^2}{2k} \right) \kappa d\kappa \end{aligned} \quad (38)$$

$$\sigma_\chi = 4\pi^2 k^2 \int_0^\infty dz \int_0^\infty \phi_n(\kappa, z) \cdot J_0(\kappa r) \cdot \sin^2 \left( \frac{z\kappa^2}{2k} \right) \kappa d\kappa \quad (39)$$

Dado que  $C_\chi(0) = \sigma_\chi^2$  se puede obtener:

$$\begin{aligned} \sigma_\chi^2 &= 4\pi^2 k^2 \int_0^\infty dz \int_0^\infty \phi_n(\kappa, z) \cdot J_0(\kappa r) \cdot \sin^2 \left( \frac{z\kappa^2}{2k} \right) \kappa d\kappa \\ \sigma_\varphi^2 &= 4\pi^2 k^2 \int_0^\infty dz \int_0^\infty \phi_n(\kappa, z) \cdot J_0(\kappa r) \cdot \cos^2 \left( \frac{z\kappa^2}{2k} \right) \kappa d\kappa \end{aligned} \quad (40)$$

Combinando la ecuación (33) con las expresiones recién obtenidas se concluye que:

$$\begin{aligned} D_\chi &= 8\pi^2 k^2 \int_0^\infty dz \int_0^\infty \phi_n(\kappa, z) \cdot [1 - J_0(\kappa r)] \cdot \sin^2 \left( \frac{z\kappa^2}{2k} \right) \kappa d\kappa \\ D_\varphi &= 8\pi^2 k^2 \int_0^\infty dz \int_0^\infty \phi_n(\kappa, z) \cdot [1 - J_0(\kappa r)] \cdot \cos^2 \left( \frac{z\kappa^2}{2k} \right) \kappa d\kappa \end{aligned} \quad (41)$$

Estos dos resultados muestran que la intensidad de las fluctuaciones aumenta con la distancia a la turbulencia, mientras que las fluctuaciones de fase son prácticamente independientes de la distancia teniendo en cuenta que la escala espacial ( $k^{-1}$ ) es mayor que la longitud de Fresnel  $\sqrt{\lambda z}$  [14]. Para grandes telescopios en los que se aplica la óptica adaptativa generalmente son válidas las aproximaciones a primer orden [14] tal que  $\sin \left( \frac{z\kappa^2}{2k} \right) \sim \frac{z\kappa^2}{2k}$  y  $\cos \left( \frac{z\kappa^2}{2k} \right) \sim 1$ .

## 4.2 COMPONENTES DE UN SISTEMA DE ÓPTICA ADAPTATIVA

A lo largo de la siguiente sección se van a introducir los componentes principales que integran un sistema de óptica adaptativa, mostrándose los modelos que se utilizan con mayor frecuencia de cada tipo de componente. El orden, la cantidad y el modelo de cada tipo de componente será diferente en función de cada sistema de óptica adaptativa.

Estos componentes será el encargado de llevar a cabo diferentes funciones, medir el frente de onda, aplicar una corrección, etc. Conocer los principios que rigen el funcionamiento de cada uno de ellos permitirá su máximo aprovechamiento cuando se intenten combinar con reconstructores basados en inteligencia artificial.

### 4.2.1 Sensor de frente de onda

El sensor del frente de onda (conocido como WFS por sus siglas en inglés *Wavefront Sensor*) es el elemento encargado de registrar los frentes de onda recibidos y extraer

información sobre las aberraciones que presentan una vez atravesada la turbulencia atmosférica.

Una de las propiedades fundamentales de un sistema de óptica adaptativa es que las correcciones deben realizarse rápidamente, tratando de corregir las imágenes lo más cerca al tiempo real. Por lo tanto, si todo el sistema tiene que funcionar prácticamente en tiempo real, en el diseño de cada uno de sus componentes prima tanto la velocidad como la precisión.

Los sensores de frente de onda que se utilizan frecuentemente en test ópticos en los que la fase de la onda recibida es medida y reconstruida con gran exactitud no son válidos en los sistemas de óptica adaptativa, ya que para obtener esos niveles de precisión el tiempo empleado son minutos, horas o incluso días.

El requisito fundamental de cualquier sensor de frente de onda en óptica adaptativa es la determinación de la fase de los rayos incidentes, independientemente de la intensidad de estos. Esto es debido a que un sistema de óptica adaptativa expuesto a la atmósfera recibirá luz con elevada variabilidad de intensidades, de ahí la necesidad de que sea capaz de identificar los rayos propios de la fuente que se está estudiando respecto a todos los demás, que en muchas ocasiones no serán los que presentan mayor intensidad.

El efecto de la fase en la propagación de la luz está altamente relacionado con el principio de Huygens-Fresnel, que se expresa a través de la integral siguiente [2]:

$$U(x, y) = -\frac{iA}{\lambda R} e^{-ikr} \int_S \frac{e^{i(\varphi+r)}}{r} dS \quad (42)$$

Siendo  $U(x, y)$  la amplitud compleja de la onda esférica  $\frac{A}{R} e^{ik(\varphi-R)}$  a una distancia  $r$  de la apertura. El patrón de intensidad observado está definido por el cuadrado de la amplitud  $U$ , es decir,  $I(x, y) = |U(x, y)|^2$ . El patrón de intensidad de una figura es lo que se conoce como la función PSF (por sus siglas en inglés; *Point Spread Function*).

El mayor interés del sensor de frente de onda reside en medir la fase del frente de onda recibido. Conocer la fase es un problema complejo, ya que al contrario de lo que pueda parecer, no se puede medir de manera directa sobre los rayos de luz recibidos por varios motivos:

1. La fase que presentan los fotones de los rayos de luz recibidos son fruto de la interacción con otros fotones presentes en su camino óptico, por lo que las fases medidas siempre se obtienen de forma relativa.
2. No existe la función inversa de la exponencial de la fase, ya que la función exponencial compleja no es inyectiva, por tanto, no existe una única solución para:

$$\varphi(x, y) = f(U(x, y), |U(x, y)|) \quad (43)$$

Siendo el campo  $U(x, y) = |U(x, y)| e^{-i\varphi(x, y)}$

Un primer intento de aproximación al problema de la determinación de la fase, a pesar de las evidentes dificultades que comporta, fue intentar extraer información directamente del patrón de intensidad. Autores como Robinson intentaron resolver este problema con condiciones extremadamente restrictivas sobre la medición de la intensidad. Propuso que la



fase de la radiación puede calcularse como la transformada de Hilbert [16] del logaritmo de la intensidad, es decir:

$$\phi(x) = -HT[\log |U(x)|] \quad (44)$$

No obstante, el resultado es válido solamente en campos de fase mínima. Este tipo de campos, que ocurren cuando por ejemplo la fase está concentrada en los bordes de una apertura, rara vez ocurren en los campos electromagnéticos que son objeto de estudio de la Óptica. De ahí que su rango de validez sea bastante limitado.

Existen también técnicas basadas en métodos estadísticos [2], como la que se basa en una computación iterativa de la función de máxima verosimilitud del frente. El método consiste en minimizar la función

$$\sum_i \frac{[I_{media} - I_i]^2}{\sigma^2} \quad (45)$$

Donde  $I_i$  son las intensidades del frente,  $I_{media}$  un parámetro que mide la media de la intensidad y  $\sigma$  es la varianza. Este método es bastante robusto (en términos de consistencia matemática) pero requiere de métodos no lineales para alcanzar una solución óptima, ya que no resulta sencillo minimizar este tipo de funciones. La determinación del frente se lleva a cabo utilizando funciones base en dos dimensiones basadas en los patrones de intensidad resultado de evaluar la ecuación anterior.

Otros métodos iterativos basados en algoritmos estadísticos es el método de Teague [2], que obtiene la fase del frente de onda aplicando la técnica de los momentos. Esta técnica se caracteriza por ser muy compleja y verse muy afectada por factores como el ruido, por lo que se trata de un método con una validez muy limitada.

Por lo tanto, pese a la existencia de métodos iterativos que reconstruyen el frente de onda de manera directa, sus aplicaciones son muy limitadas por diversos motivos, donde destacan especialmente su mal funcionamiento frente a la gran cantidad de ruido que se registra en las medidas y la dificultad para su implementación en tiempo real debido a la complejidad de las técnicas.

#### 4.2.1.1 Sensor de Shack-Hartmann

El sensor de Shack-Hartmann (SH o SHWFS por sus siglas en inglés, *Shack-Hartmann Wavefront Sensor*) es el sensor de frente de onda más utilizado en los sistemas de OA actualmente y uno de los primeros en ser desarrollados para los sistemas pioneros. Su método de medición de la fase sigue siendo válido hoy en día pese a tratarse de uno de los primeros desarrollados, si bien la tecnología utilizada en los sensores ha ido evolucionando a lo largo de los años. En un principio fueron desarrollados para observaciones nocturnas pero debido a gran rendimiento han sido optimizados para su aplicación también en óptica adaptativa solar.

El origen del sensor reside en un test para lentes o espejos llamado el test de Hartmann que consiste en colocar una máscara opaca con agujeros detrás de la lente sobre la que se hace el test [13]. En esta configuración, cada agujero de la máscara actúa como una apertura y, como la luz converge tras atravesar la lente, la imagen que se produce es un conjunto de puntos. Con la calibración apropiada, la posición de dichos puntos es una medida directa de la inclinación del frente de onda local en cada agujero y, por tanto, una descripción de la calidad de la lente [13], [17]. La variación introducida por Shack fue colocar lentes también en los agujeros de tal manera

que la eficiencia de captación de la luz de la máscara aumentaba y, con los puntos enfocados, se reducían los efectos de difracción debidos a los agujeros [13], [17].

Tomando como principio el test explicado anteriormente, el SHWFS consiste en una matriz de lentes a una distancia  $f$  de una matriz detectora de fotones (sensor CCD) tal que cuando un frente de onda plano incide sobre el conjunto de lentes, la luz coincide con los puntos centrales de la matriz CCD. Cuando el frente de onda está aberrado (tiene curvatura), la luz no se proyecta sobre los centros de la matriz CCD, sino sobre puntos arbitrarios cualesquiera. El desplazamiento de la imagen respecto del punto es proporcional a la pendiente del frente en cada lente.

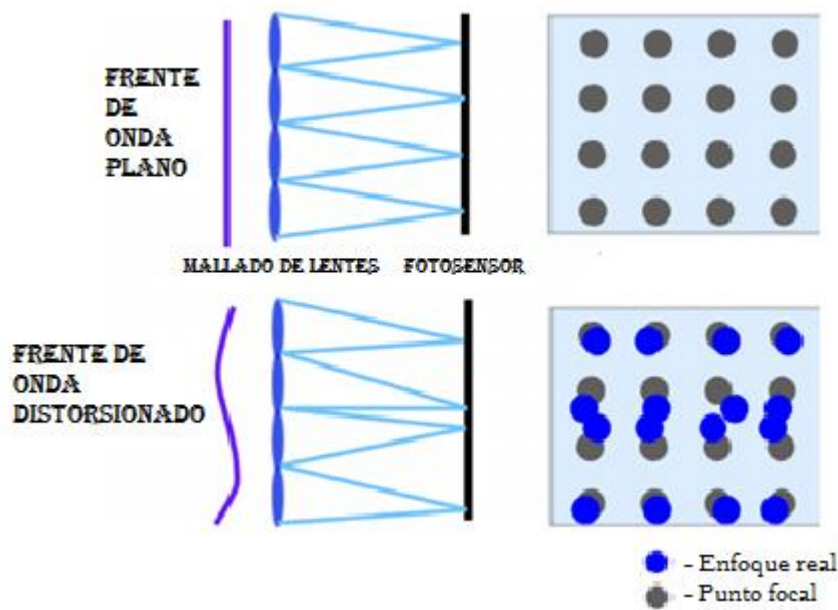


Figura 1: Sensor de Shack-Hartman explicado gráficamente. En la matriz de detección de luz, los puntos perfectamente distribuidos se corresponden con las imágenes del frente de onda al atravesar las lentes. Si el frente de onda está modificado por una turbulencia atmosférica, la proyección sobre el plano imagen de esta radiación no se distribuye de manera perfecta, sino que cada punto se sitúa a una distancia  $\Delta y$  del punto que se considera de equilibrio. Esta cantidad es clave para posteriormente estimar, tal y como se comenta a continuación, la fase global de la radiación incidente.

Hay dos elementos principales que caracterizan al SHWFS; el conjunto de lentes y los sensores CCD. Las lentes utilizadas en los SHWFS se caracterizan por ser elaboradas con gran precisión, tanto en la alineación que presentan en la matriz como en la calidad de los materiales utilizados, ya que deben introducir la menor aberración posible sobre la luz focalizada. Por otro lado, los detectores CCD son los encargados de medir la luz recibida, se trata de sistemas que basan su funcionamiento en el efecto fotoeléctrico [18], la carga liberada por estas matrices es directamente proporcional a la intensidad de la radiación incidente.

El ángulo de desplazamiento  $\theta_c$  del frente de onda plano aberrado respecto a los centros de la matriz CCD se puede obtener como [2], [19]:

$$\theta_c \approx \tan\theta_c = \frac{\Delta y}{f} \quad (46)$$

Siendo  $\Delta y$  el desplazamiento de la imagen respecto del centro de cada lente y  $f$  es la distancia focal de cada una de las lentes.

Como este ángulo de desplazamiento es proporcional a la pendiente del frente en cada lente, se tiene que:

$$\theta_c \sim \langle \nabla \varphi(x) \rangle \quad (47)$$

Es decir:

$$\theta_c \approx \frac{1}{A} \int_A \nabla \varphi(x) dx \quad (48)$$

A primer orden:

$$\varphi(x) = \varphi(0) + x \nabla \varphi(x) + O(|x|^2) \quad (49)$$

De esta forma se obtiene una reconstrucción de la fase a primer orden. En realidad, este proceso no es el que realiza el SHWFS, ya que se está teniendo en cuenta un único valor del ángulo de desplazamiento (podría ser el de una única subapertura o el valor medio entre todas ellas). Por tanto, se estaría perdiendo mucha información en comparación con tener en cuenta todas las medidas registradas por cada una de las subaperturas.

Para compensar el efecto de la atmósfera turbulenta en la luz, es necesario medir la variación del frente de onda local de manera precisa en cada subapertura y es por eso por lo que en óptica adaptativa se emplean múltiples variaciones de la técnica descrita previamente.

La posición del punto de luz focalizado por la lente en cada una de las subaperturas se denomina centroide [2], [20], [21]. El centroide de una distribución de intensidad es igual al desplazamiento del origen del plano de la imagen, que es proporcional a la inclinación del frente de onda y su valor se puede obtener de la siguiente forma:

$$\begin{aligned} \bar{x} &= \frac{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x I(x, y) dx dy}{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I(x, y) dx dy} \\ \bar{y} &= \frac{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y I(x, y) dx dy}{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I(x, y) dx dy} \end{aligned} \quad (50)$$

La ecuación 51 describe un modelo ideal en el que toda la superficie de la subapertura registra la luz. En realidad, la capacidad de senseo del sensor está limitada por el número de píxeles CCD y el espacio que existe entre cada uno de ellos, teniendo esto en cuenta, la ecuación se modificaría de la siguiente forma:

$$\begin{aligned} x_{c,k} &= \frac{\sum_{i,j} x_{i,j} I_{i,j}}{\sum_{i,j} I_{i,j}} \\ y_{c,k} &= \frac{\sum_{i,j} y_{i,j} I_{i,j}}{\sum_{i,j} I_{i,j}} \end{aligned} \quad (51)$$

Donde el sufijo  $c$  se refiere a centroide y la  $k$  representa el número de subapertura, realizando el sumatorio sobre todos los píxeles asignados a la lente  $k$  [13], [20]. En OA es frecuente el uso de estrellas guía de referencia (que pueden ser naturales o artificiales), situadas

en una zona cercana al objeto que se quiere observar. El uso de estas estrellas guía permite comparar los centroides recibidos del objeto a observar con los frentes de onda de una fuente completamente conocida para reconstruir la turbulencia atmosférica. Por tanto, los centroides de la fuente a estudiar serán comparados con los de la estrella guía [2] para obtener las pendientes de los frentes de onda de la siguiente forma:

$$\begin{pmatrix} \langle dw/dx \rangle \\ \langle dw/dy \rangle \end{pmatrix}_k = \begin{pmatrix} \beta_x \\ \beta_y \end{pmatrix}_k = \frac{1}{L_H} \begin{pmatrix} x_c - x_r \\ y_c - y_r \end{pmatrix} \quad (52)$$

Siendo  $L_H$  la distancia del plano focal.

El SHWFS ha sido el sensor escogido para el desarrollo de los trabajos que se han llevado a cabo a lo largo de esta tesis, puesto que al ser el más utilizado existe mayor número de simuladores y datos que se utilizaron para entrenar nuestros reconstructores. A pesar de ello, a continuación, se van a introducir el sensor de curvatura y el sensor piramidal que, pese a ser utilizados en un gran menor número de sistemas, también están presentes en muchos telescopios en la actualidad. Dada su menor relevancia solamente se expondrá una breve idea de su funcionamiento.

#### 4.2.1.2 Sensor de curvatura

Inventado por Roddier [17], [22] en el año 1988, su funcionamiento se basa en aprovechar que a medida que la luz se propaga desde la pupila al foco, las fluctuaciones de fase se traducen en fluctuaciones de la irradiancia. En consecuencia, medir las fluctuaciones de irradiancia en una imagen desenfocada se puede utilizar para estimar la curvatura del frente al entrar en la pupila y por tanto la fase. La relación de la irradiancia con la fase y la curvatura está dada por la ecuación siguiente:

$$\frac{I_1(r) - I_2(-r)}{I_1(r) + I_2(-r)} = \frac{f(f-s)}{s} \left[ \nabla^2 \varphi \left( \frac{f}{s} r \right) - \frac{\delta}{\delta n} \varphi \left( \frac{f}{s} r \right) \delta_c \right] = \frac{2f^2 c_\omega}{s} \quad (53)$$

Donde  $c_\omega = \frac{1}{r_{\text{omega}}}$  es la curvatura local en el punto  $r$ ,  $f$  representa la distancia focal del telescopio,  $s$  la separación de los sensores y  $\delta c$  es la delta de Dirac, que representa la intensidad de la derivada normal de la señal. La reconstrucción del frente de onda a partir de esta expresión es compleja, pero se basa en técnicas iterativas parecidas a las que se utilizan en la resolución de la ecuación de Poisson.

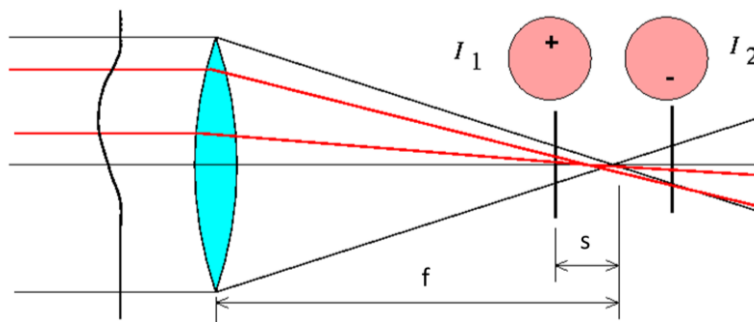


Figura 2: Sensor de curvatura simple consta de dos conjuntos de detectores situados a cada lado del plano focal, separados la misma distancia del plano focal  $l$ , la curvatura se estimará en función de las diferencias de intensidad registradas por ambos sensores.

### 4.2.1.3 Sensor piramidal

Se trata de uno de los sensores desarrollados más recientemente, en concreto fue ideado por Ragazzoni en 1996. Este dispositivo consiste en una pirámide de cuatro caras que se coloca a una distancia  $f$  igual a la distancia focal de la apertura al plano imagen. La pirámide divide el rayo que proviene de la apertura en 4 rayos diferentes. Posteriormente, la imagen se proyecta en un detector CCD [2].

En el caso de que se recibiese un frente de onda plano, en las cuatro pupilas detectores se observaría el mismo patrón de irradiancia. Cualquier desviación, por pequeña que sea, del frente de onda original, generará patrones de irradiancia diferentes en cada una de las pupilas de detección.

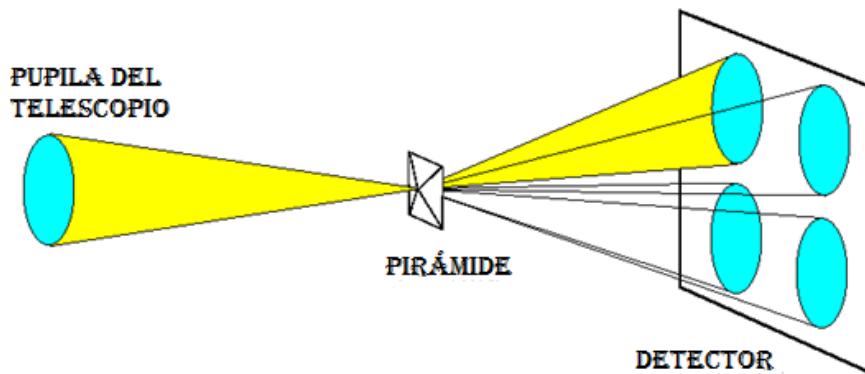


Figura 3: Representación esquemática del sensor piramidal. En el caso de un frente de onda plano sin aberrar, los cuatro detectores recibirían el mismo patrón de irradiancia.

Para medir la fase, el prisma se moverá en las direcciones  $X$  e  $Y$  y se recogerán las intensidades que se recojan en cada uno de los planos, como se indica en la figura. Habrá un desplazamiento máximo en cada eje denotado por  $\pm L$ .

El prisma se moverá siempre con velocidad constante, el desplazamiento del rayo en el plano focal, como ya se ha explicado en apartados anteriores [2], vendrá dado por:

$$\nabla_{desplazamiento} = f \frac{\nabla \varphi}{k} \quad (54)$$

Donde  $f$  es la distancia focal,  $\nabla_{\varphi}$  el promedio de la fase y  $k$  el número de ondas de la radiación que se está considerando. Si miramos la imagen del diagrama, se puede comprobar que cuando el prisma oscila en la dirección  $x$ , el tiempo que está el rayo en la pupila  $a$  es:

$$t_a = \frac{L + f \nabla_x \varphi / k}{2L} \quad (55)$$

Mientras que la fracción de tiempo que permanece en la pupila  $c$  es:

$$t_c = \frac{L - f \nabla_x \varphi / k}{2L} \quad (56)$$

En consecuencia, si se restan las intensidades se puede obtener una forma de calcular el gradiente de la fase del frente:

$$\frac{I_a - I_c}{I_a + I_c} = \frac{f \nabla_x \varphi}{kL} \quad (57)$$

La combinación del movimiento tanto en el eje  $X$  como en el eje  $Y$  permite deducir las expresiones de las componentes del gradiente de la fase:

$$\begin{aligned} \nabla_x \varphi(\vec{r}) &= \frac{kL}{f} \left( \frac{I_a(\vec{r}) - I_c(\vec{r}) + I_b(\vec{r}) - I_d(\vec{r})}{I_t} \right) \\ \nabla_x \varphi(\vec{r}) &= \frac{kL}{f} \left( \frac{I_b(\vec{r}) - I_a(\vec{r}) + I_d(\vec{r}) - I_c(\vec{r})}{I_t} \right) \end{aligned} \quad (58)$$

El sensor piramidal representa una gran ventaja respecto a otros tipos de sensores ya que, como se puede ver, se trata de un modelo muy sencillo para la medición de las variaciones de fase, lo que permite implantarlo con facilidad en tiempo real. Además, la resolución espacial del sensor no está limitada por una lente (como puede ser en el caso de los SH y otros tipos de sensores que utilizan lentes) sino que solamente dependerá del tamaño del píxel del detector.

#### 4.2.2 Sistemas correctores del frente de onda

Una vez conocidas las aberraciones presentes en el frente de onda es necesario disponer de sistemas que permitan corregirlas. Corregir la aberración presente en un frente de onda es tan simple como producirle otra aberración que sea la conjugada de la primera, para obtener el frente de onda original.

Algunos elementos que serían capaces de corregir las aberraciones podrían ser lentes o espejos. En el caso de la OA estas correcciones serán llevadas a cabo por espejos que son capaces de modificar sus alguna de sus propiedades en función de las aberraciones presentes. A continuación, se presentan los modelos de espejos más utilizados en óptica adaptativa.

##### 4.2.2.1 Espejos tilt

Denominados espejos de inclinación en castellano, se utilizan para corregir las desviaciones en la dirección de los rayos que se reciben. Al realizar una observación, los telescopios apuntan en la dirección en la que se encuentra el objeto a observar, no obstante, fruto de las aberraciones producidas por la turbulencia atmosférica, los rayos recibidos de dicha fuente provienen de direcciones diferentes a la que cabría esperar en ausencia de atmósfera. Para compensar estas diferencias se utilizan los espejos tilt [13], [23], [24]. La gran mayoría de estos espejos tienen asociados sensores de posición de sistemas de retroalimentación (conectados al ordenador de control) que les permiten ejecutar las órdenes de corrección que les son dadas.

Existen ciertos requerimientos para el buen funcionamiento de un espejo, como: ancho de banda (frecuencias que es capaz de corregir), la velocidad o la aceleración. En general se necesita que los espejos de inclinación estén en la posición requerida en menos de la mitad de tiempo que el tiempo característico de respuesta del ordenador de control.

Este tipo de espejos [13] están diseñados para eliminar ciertas aberraciones del frente de onda, además de la propia inclinación. La mayor cantidad de tilt (inclinación) oscila entre  $M = \pm 2,5\sigma_{tilt}$ , donde  $\sigma_{tilt}$  es la desviación estándar de la inclinación. Matemáticamente [12]:

$$\sigma_{tilt} = \sqrt{0.184 \left( \frac{D}{r_0} \right)^{\frac{5}{3}} \left( \frac{\lambda}{D} \right)^2} \quad (59)$$

Donde  $\lambda$  es la longitud de onda,  $D$  el diámetro de la apertura y  $r_0$  la longitud de coherencia de Fried.

En capítulos posteriores se explicará cómo la aberración del frente de onda puede ser expresada mediante una serie de polinomios denominados los polinomios de Zernike. Cada tipo de aberración se asociará a un polinomio diferente también llamado modo. La corrección más simple de un frente de onda es la variación de la dirección del haz o la inclinación del frente de onda, que son los elementos asociados a los polinomios de Zernike de orden más bajo [13], [17]. Los elementos encargados de corregir los modos de Zernike más bajos en un sistema de OA son los espejos tilt.

#### 4.2.2.2 Espejos segmentados

Pertenecen al grupo ya de los denominados espejos activos, que se caracterizan por tener la capacidad de modificar la forma de su superficie, permitiendo así correcciones de aberraciones de mayor orden [2].

En el caso de los espejos segmentados, su superficie está dividida en pequeñas unidades individuales de espejos con capacidad de corrección independiente cada una de ellas. Existen muchos tipos de partes individuales y tienen geometrías diferentes: cuadradas, hexagonales y hasta circulares se han utilizado. Los espejos segmentados se utilizan para corregir aberraciones estáticas en un sistema óptico o para compensar turbulencias atmosféricas. La ventaja de este tipo de dispositivos es que son altamente escalables, es decir, se pueden fabricar espejos muy versátiles y con más funcionalidades a medida que se añaden más y más unidades al espejo. Los espejos segmentados pueden tener dos tipos de configuraciones: pistón (no hay inclinación de la superficie metálica) o pistón-inclinación (se produce al mismo tiempo desplazamiento e inclinación de la superficie metálica del sistema).

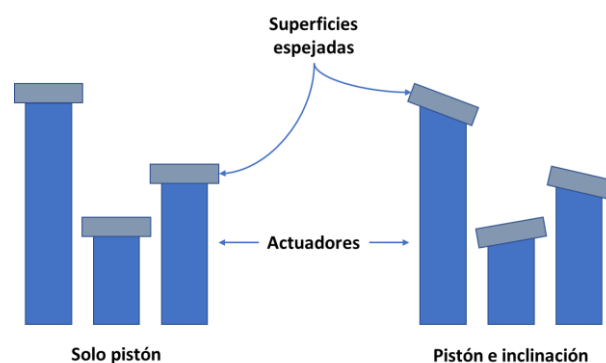


Figura 4: Representación esquemática de la diferencia entre espejos segmentados pistón y espejos segmentados pistón-inclinación.

Pese a la gran variedad de formas que puede tomar la superficie, cada unidad de espejo no puede colocarse en posiciones libres sin tener en cuenta al resto. Debe existir una continuidad a lo largo de la superficie del espejo ya que, de lo contrario, las intersecciones entre cada uno de los actuadores provocarían aberraciones mayores en el frente de onda.

La forma más común de las superficies espejadas es hexagonal. Los espejos normalmente tienen forma circular (al igual que la pupila del telescopio), maximizar el número de unidades individuales en un círculo da lugar a disposiciones hexagonales.

#### 4.2.2.3 Espejos Deformables (DM)

Los espejos deformables (normalmente conocidos como DM por sus siglas en inglés *Deformable Mirror*) se diferencian de los anteriores por ser espejos de superficie continua con un medio mecánico que permite deformar la superficie. Los dispositivos que realizan la deformación se denominan actuadores, que en general suelen ser continuos (espejos de membrana, espejos bimorfos...). También pueden tener actuadores discretos perpendiculares a la superficie, o pueden tener actuadores discretos en los bordes que imparten momentos de flexión. Debido a que la superficie del espejo parece doblarse de forma similar a una membrana de caucho estirada, los DM de superficie continua a menudo se denominan espejos de goma o de caucho [2], [11], [22].

Los primeros DM fueron desarrollados en la década de 1970 y son los denominados espejos de actuador discreto. El material que se utilizó para su diseño fue apilamientos de discos piezoeléctricos, de manera que, modificando el voltaje se modificaba la superficie del espejo. En aquel entonces el voltaje requerido para cada actuador era superior a  $1kV$ , lo que dificultaba enormemente su desarrollo.

Sin embargo, con el desarrollo de la tecnología de estos dispositivos se facilitó su implementación en situaciones reales. Algunos parámetros importantes que mejoran el rendimiento de los espejos son: miniaturización de actuadores, bajos voltajes, precisión en la posición y baja histéresis. En la actualidad, los materiales más comunes en la fabricación de estos dispositivos son: el titanato circonato de plomo y el niobato magnesio plomo [11]. Respecto de la capa externa, es necesario que esta tenga elevada reflectividad para que la radiación incidente no modifique el efecto de los actuadores sobre el espejo. Otras cuestiones importantes relativas a los espejos deformables es que su respuesta sea lineal con el voltaje para, de nuevo, evitar respuestas indeseadas en la modificación de la fase de la radiación.

Existen diferentes tipos de espejos deformables:

- Espejos de actuadores: Existen unidades piezoeléctricas a modo de pistones debajo de la capa que refleja la luz que modifican la superficie del DM.
- Espejos deformables bimorfos: Consisten en una placa de metal o vidrio unida a hoja de cerámica piezoeléctrica polarizada en la dirección normal a su superficie cubierta de electrodos independientes [2]. Al tratarse de un material piezoeléctrico, cuando se aplica una diferencia de potencial entre los electrodos frontales y traseros sus dimensiones cambian. El cambio en el radio local de curvatura [2], [25] viene dado por la siguiente expresión:

$$R = \frac{Vd_{13}}{t^2} \quad (60)$$

donde  $d_{13}$  es el coeficiente del tensor piezoeléctrico y  $t$  es el espesor de la hoja de cerámica más la placa de metal (o vidrio).

- Espejo deformable de membrana: Se caracterizan por tener una membrana reflectante entre un electrodo transparente y una serie de electrodos individuales en la parte posterior del espejo [2]. Cuando no existe diferencia de potencial la membrana es plana mientras que cuando  $V \neq 0$  la atracción electrostática entre electrodos deforma la membrana hasta conseguir la forma



deseada. La curvatura local de la superficie viene descrita por la ecuación de Poisson [11], [26].

$$\nabla^2 z(x, y) = -\frac{P(x, y)}{T} \quad (61)$$

donde T es la tensión de la membrana *fuerza/longitud* y P(x, y) es la presión en la posición (x, y).

#### 4.2.3 Sistemas reconstructores del frente de onda

Serán los encargados de, a partir de la información registrada por los sesores de frente de onda, calcular las correcciones necesarias. Se trata de la parte central de este trabajo puesto que el objetivo es desarrollar nuevos sistemas reconstructores basados en inteligencia artificial, en lugar de los típicamente usados cuyo funcionamiento depende de complejos algoritmos matemáticos.

La reconstrucción del frente de onda a partir de medidas de las pendientes es un problema que tiene, en general, dos maneras de ser resuelto. En este punto cabe recordar que las pendientes es la información medida por los sensores explicados en el apartado anterior. La primera es la reconstrucción por zonas en la que el frente de onda se estima directamente a partir de un conjunto discreto de medidas del gradiente de la fase y la segunda es la reconstrucción por modos [27]. En esta última, el frente de onda se expande en una base de funciones ortogonal y los coeficientes se estiman a partir de las medidas de las pendientes de la fase. La base de funciones que se emplea en astronomía para representar la fase ( $\varphi$ ) de la turbulencia atmosférica, donde las aperturas normalmente son circulares [27], son los polinomios de Zernike [2], [28].

$$\varphi = \sum_{k=0}^K a_k Z_k(x, y) \quad (62)$$

donde  $a_k$  son los coeficientes de los modos y  $Z_k(x, y)$  los polinomios de Zernike, que están formados por una parte angular y una parte radial. Se definen como:

$$\begin{aligned} Z_{nm\text{par}} &= \sqrt{n+1} R_n^m(r/R') \sqrt{2} \cdot \cos(m\theta) \\ Z_{nm\text{impar}} &= \sqrt{n+1} R_n^m(r/R') \sqrt{2} \cdot \sin(m\theta) \\ Z_{n0} &= \sqrt{n+1} R_n^0(r/R') \end{aligned} \quad (63)$$

Donde los polinomios radiales  $R_n^m$  se definen como:

$$R_n^{\pm m} \left( \frac{r}{R'} \right) = \sum_{s=0}^{\frac{n-m}{2}} (-1)^s \frac{(n-s)!}{s! \left( \frac{n+m}{2} - s \right)! \left( \frac{n-m}{2} - s \right)!} \left( \frac{r}{R'} \right)^{n-2s} \quad (64)$$

En óptica adaptativa los primeros modos de Zernike se asocian a correcciones particulares del frente de onda. Los tres primeros modos se conocen como pistón ( $Z_0^0$ ), inclinación y centelleo ( $Z_1^{\pm 1}$ ), y astigmatismo y desenfoque ( $Z_2^0$  y  $Z_2^{\pm 2}$ ) [29], [30]. Una representación esquemática de los modos de Zernike puede verse en la Figura 5.

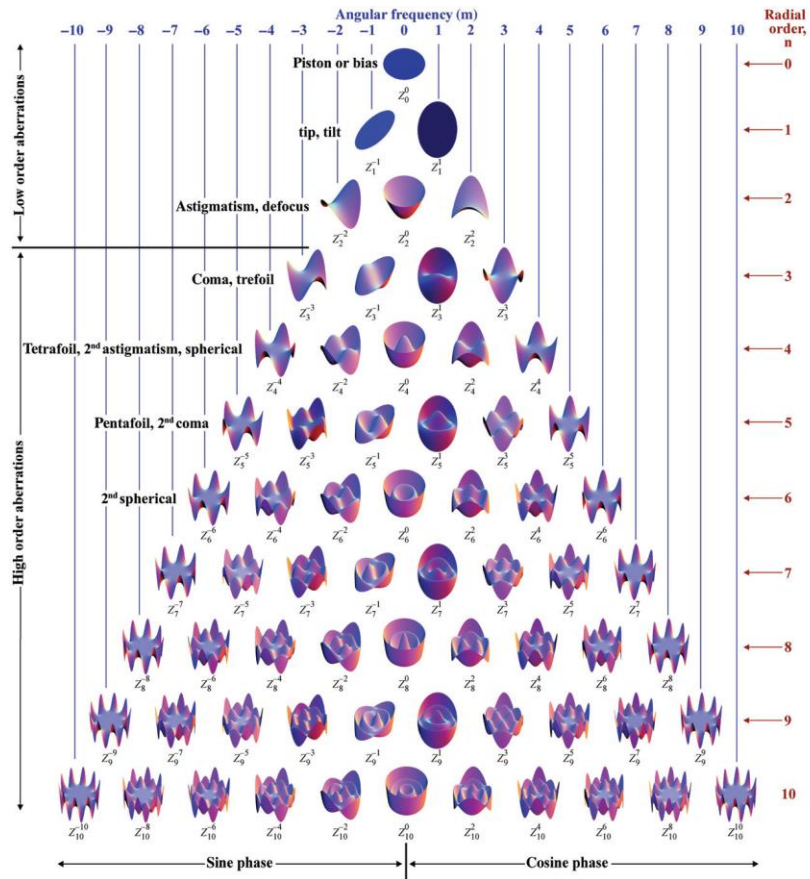


Figura 5: Modos de Zernike. Imagen extraída de [30].

Hay que mencionar que los polinomios de Zernike no son la única base empleada para la compensación de las turbulencias. De hecho, en términos de la matriz de covarianza, la base óptima son los polinomios de Karhunen-Loève (K-L) ya que en esta base dicha matriz es diagonal [27]. La definición de estos polinomios es la siguiente:

$$s_p^{\pm q}(r) = \sum_{n=m}^{\infty} C_n^m R_n^m(r) \quad (65)$$

siendo la suma sobre  $m, m + 2, m + 4, \dots$ . Como ambos polinomios radiales están normalizados [9] los coeficientes  $C_n^m$  han de satisfacer

$$\sum_{n=m}^{\infty} (C_n^m)^2 = 1 \quad (66)$$

Se ha demostrado que para una reconstrucción del frente de onda óptima que minimice el error total de reconstrucción y que maximice el cociente de Strehl, la reconstrucción de los modos con la base de funciones K-L sobre pupilas circulares con subaperturas cuadradas da lugar al error de reconstrucción más pequeño [27].

#### 4.2.3.1 Reconstrucción del frente de onda por mínimos cuadrados

Se trata del método de reconstrucción más común para obtener el frente de onda en función de la información medida por el SH. Se basa en la aplicación del método de mínimos cuadrados (LS por sus siglas en inglés, *Least-Squares*) para la obtención de los coeficientes de los polinomios de Zernike.

El sensor SH mide las pendientes de los frentes de onda en diferentes posiciones, ya que obtiene una medida para cada una de las subaperturas. Teniendo en cuenta todos esos datos, se obtendrá un sistema de ecuaciones lineales de la siguiente forma (teniendo en cuenta ( 62 )), siendo  $m$  la posición de la medida para cada una de las subaperturas:

$$\begin{aligned} \left. \frac{d\varphi}{dx} \right|_m &= \sum_{k=1}^K a_k \left. \frac{dZ_k(x, y)}{dx} \right|_m \\ \left. \frac{d\varphi}{dy} \right|_m &= \sum_{k=1}^K a_k \left. \frac{dZ_k(x, y)}{dy} \right|_m \end{aligned} \quad (67)$$

Teniendo en cuenta todos los datos recibidos por las subaperturas:

$$s = \begin{pmatrix} \left. \frac{d\varphi}{dx} \right|_1 \\ \left. \frac{d\varphi}{dx} \right|_2 \\ \vdots \\ \left. \frac{d\varphi}{dx} \right|_{M/2} \\ \left. \frac{d\varphi}{dy} \right|_1 \\ \vdots \\ \left. \frac{d\varphi}{dy} \right|_{M/2} \end{pmatrix} \quad a = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_k \end{pmatrix} \quad (68)$$

Se obtiene un sistema con dos vectores donde  $s$  son los valores de las pendientes evaluadas en los puntos  $m$  y  $a$  representa el valor de los diferentes coeficientes de los modos de Zernike. Se obtiene un sistema a resolver  $s = [B] \cdot a$ .

Obviamente la matriz  $B$  es conocido pues resulta de derivar los distintos polinomios de Zernike:

$$B = \begin{pmatrix} \left. \frac{dZ_k(x, y)_1}{dx} \right|_1 & \left. \frac{dZ_k(x, y)_2}{dx} \right|_1 & \dots & \left. \frac{dZ_k(x, y)_K}{dx} \right|_1 \\ \left. \frac{dZ_k(x, y)_1}{dx} \right|_2 & \left. \frac{dZ_k(x, y)_2}{dx} \right|_2 & \dots & \left. \frac{dZ_k(x, y)_K}{dx} \right|_2 \\ \vdots & \vdots & & \vdots \\ \left. \frac{dZ_k(x, y)_1}{dx} \right|_{M/2} & \left. \frac{dZ_k(x, y)_2}{dx} \right|_{M/2} & \dots & \left. \frac{dZ_k(x, y)_K}{dx} \right|_{M/2} \\ \left. \frac{dZ_k(x, y)_1}{dy} \right|_1 & \left. \frac{dZ_k(x, y)_2}{dy} \right|_1 & \dots & \left. \frac{dZ_k(x, y)_K}{dy} \right|_1 \\ \vdots & \vdots & & \vdots \\ \left. \frac{dZ_k(x, y)_1}{dy} \right|_{M/2} & \left. \frac{dZ_k(x, y)_2}{dy} \right|_{M/2} & \dots & \left. \frac{dZ_k(x, y)_K}{dy} \right|_{M/2} \end{pmatrix} \quad (69)$$

El número de filas de la matriz vendrá determinado por el número de subaperturas del sensor que se está utilizando. El número de columnas en cambio dependerá de la cantidad de modos que se quieran reconstruir, a mayor número de modos más precisa será la reconstrucción.



#### 4.2.4 Sistemas de control

Una vez explicados los dispositivos encargados de conocer cómo es el frente de onda recibido, el reconstructor y los elementos encargados de aplicar la corrección necesaria, es necesario hablar de los sistemas encargados de transmitir la información entre esos dos dispositivos, indicando a los dispositivos correctores la posición que deben tomar deseada.

Existen dos tipos de sistemas de control: open loop y closed loop.

- Sistemas open loop (en castellano sistemas de lazo abierto): Las medidas del frente de onda se utilizan para reconstruir la turbulencia atmosférica y aplicar una corrección utilizando los DM para cada instante. La luz se recibe en el sensor antes de ser corregida.
- Sistemas closed-loop (en castellano sistemas de lazo cerrado): en este tipo de sistemas el rayo de luz llega al sensor del frente de onda después de la corrección aplicada por el espejo deformable. Esto implica retroalimentación del sistema ya que la luz antes de ser medida es corregida con el cálculo realizado para el instante anterior. La mayor ventaja de este tipo de sistemas es que las modificaciones que deben introducirse al espejo son mucho menores, permitiendo implementarlas en tiempo real con mayor facilidad.

El sistema óptico tiene el objetivo de reconstruir el frente de onda que llega en tiempo real, es decir, las correcciones han de llevarse a cabo en el momento en el que el sistema hace la medición. Por tanto, se ha de buscar un equilibrio entre cómo llega la radiación al sistema y cuánto tiempo tarda este en responder. Sea  $\tau$  el tiempo característico de la radiación, es decir, el tiempo entre la llegada de un fotón y otro, se pueden dar tres casos:

$t_s > \tau$ . En este caso el sistema de control no llega a tiempo a corregir la radiación debido a que su respuesta tarda en producirse más que la llegada del siguiente fotón.

$t_s < \tau$ . En este caso el sistema se adelanta al siguiente fotón y corrige el frente de onda antes de que llegue la nueva radiación.

$t_s \approx \tau$ . En este caso, el tiempo de respuesta del sistema es similar al tiempo característico de los fotones. En consecuencia, se consigue que el ordenador de control envíe las señales de corrección a tiempo, lo que permite que el sistema de Óptica Adaptativa funcione correctamente.

Por tanto, el tiempo de respuesta debe ser muy similar al tiempo característico de la radiación que fluctuará debido a factores como la presión, la temperatura, etc.

La radiación que le llega al ordenador de control (una vez leída en el sensor del frente de onda) está dada por una ley sinusoidal que caracteriza a todos los movimientos de tipo ondulatorio. Es decir:

$$r(t) = A \sin(\omega t) \quad (74)$$

El sistema de control envía una señal que posteriormente el espejo deformable convertirá en deformación. La señal de salida puede expresarse como [13]:

$$c(t) = AM(\omega) \sin(\omega t + \phi(\omega)) \quad (75)$$

Donde  $A$  es la amplitud de entrada,  $M(\omega)$  es la relación entre las amplitudes de salida y de entrada para una frecuencia  $\omega$  y  $\phi(\omega)$  es el retardo de fase existente entre la radiación incidente y la radiación de salida, debido a que el sistema tarda una cantidad finita de tiempo en procesar toda la información.

Un mapeo de  $M(\omega)$  y  $\phi(\omega)$  permite ver la respuesta de la ganancia y la fase respecto de la frecuencia. En general, interesa conseguir ganancias superiores a  $0,707 \text{ dB}$ , ya que a partir de dichos niveles la respuesta del sistema se considera satisfactoria. El espectro de frecuencias para el que la respuesta es mayor que esos niveles se conoce como ancho de banda. Por tanto, la capacidad de corrección de un sistema de Óptica Adaptativa está limitada por su ancho de banda, que es clave a la hora de conseguir respuestas que sean capaces de hacer corregir los haces con los espejos deformables.


### 4.3 CONFIGURACIONES SISTEMAS DE ÓPTICA ADAPTATIVA

En función del número de los dispositivos mencionados anteriormente (sensores de frente de onda, espejos deformables, etc.) que se utilicen, su colocación y si el sistema funciona en lazo abierto o lazo cerrado existen diferentes configuraciones conocidas para OA. Cada una de las configuraciones presentará ventajas e inconvenientes respecto al resto en función del objetivo que pretenda la observación, en función del campo de visión que se desee registrar, el número de fuentes, etc.

Las configuraciones denominadas de lazo abierto se caracterizan porque el sensor de frente de onda es el primer instrumento que se encuentran los rayos de la luz al entrar en el sistema. Si por el contrario se tratase de una configuración de lazo cerrado, el primer objeto que se encontrarían los rayos sería un espejo deformable con la corrección calculada para el instante de tiempo justo anterior.

Las configuraciones que se muestran a continuación son las más comunes utilizadas típicamente en los sistemas de OA. Sin embargo, en algunos sistemas concretos, pueden existir variaciones en las mismas o presentar configuraciones únicas fruto de la combinación entre varias de las anteriores, en función del objeto que se pretenda observar.

#### 4.3.1 SCAO (*Single Conjugated Adaptive Optics*)

Se trata de la configuración en lazo cerrado más sencilla, ya que consiste en un espejo deformable asociado a un único sensor de frente de onda. Es un método que se aplica tanto en OA nocturna como solar que para bajos campos de visión ha presentado buenos resultados. Se caracteriza por tener tanto el sensor como el espejo en la dirección de una estrella guía situada en el eje de observación tal y como muestra la . No se encuentra el origen de la referencia..

Cuando se realizan observaciones en el eje de observación o muy cerca de este, el sistema obtiene buenas reconstrucciones con altos ratios de Strehl, sin embargo, su mayor limitación reside en el anisoplanatismo. Al no disponer de sensores que le permitan obtener información de la turbulencia en direcciones diferentes a la del eje de observación, si se quiere obtener una buena reconstrucción el campo de visión FOV (por sus siglas en inglés *Field of View*) está muy restringido a pequeñas distancias angulares del centro del telescopio.

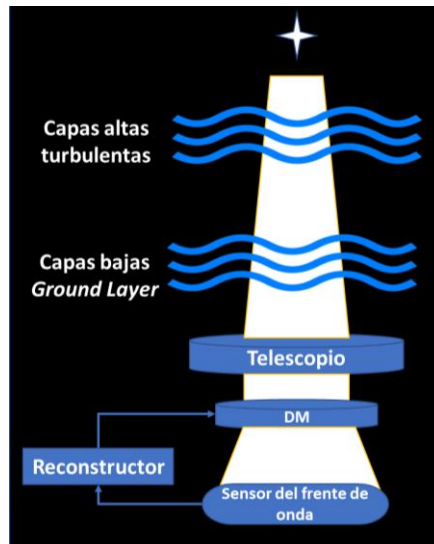


Figura 6: Representación esquemática de la configuración de OA SCAO, donde el espejo deformable se sitúa antes de la recepción del frente de onda por parte de los sensores ya que funciona en lazo cerrado.

#### 4.3.2 GLAO (Ground-Layer Adaptive Optics)

Esta configuración es similar a la anterior, pero en este caso lo que busca es corregir fundamentalmente la capa turbulenta más cercana a la superficie terrestre, denominada en inglés *Ground-Layer*. Sigue siendo una configuración en lazo cerrado, encontrándose el DM justo a continuación de la pupila del telescopio.

A diferencia del SCAO, el objetivo, además de corregir la capa más cercana es obtener un FOV mayor, por lo que se emplea más de un WFS enfocando diferentes direcciones para conocer la turbulencia atmosférica. Con este sistema se reduce el error de anisoplanatismo que aparecía muy acentuado en las configuraciones SCAO.

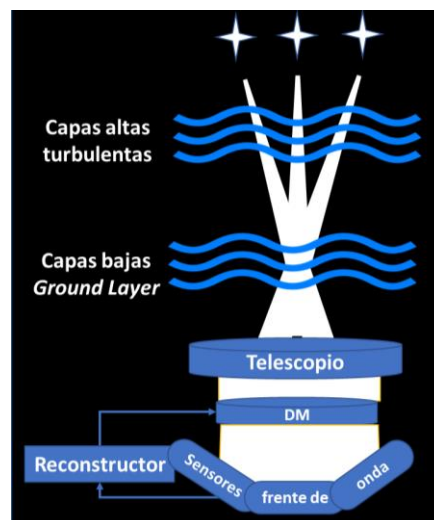


Figura 7: Representación esquemática de la configuración GLAO. El sistema dispondrá del mismo número de sensores de frente de onda como fuentes de referencia se consideren en la observación. En el caso mostrado en la imagen, los 3 WFS apuntan en la dirección de cada uno de los objetos celestes.

Al igual que en el caso anterior puede ser empleado tanto en OA nocturna como diurna, enfocando cada WFS a estrellas guía en la dirección de observación de cada uno de ellos o manchas solares en las citadas direcciones respectivamente.

#### 4.3.3 MCAO (*Multi-Conjugated Adaptive Optics*)

Se trata de una de las técnicas más recientes de óptica adaptativa que busca mejorar el rendimiento de las mostradas anteriormente, aunque también conlleva mayor complejidad. Funciona también en lazo cerrado y se caracteriza por tener varios WFS apuntando en diferentes direcciones en combinación con varios DM que corrigen diferentes turbulencias cada uno de ellos.

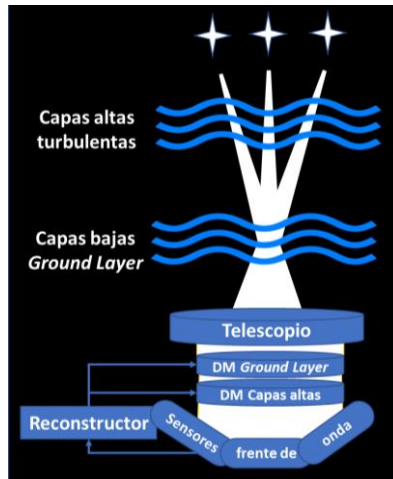


Figura 8: Representación esquemática de la configuración MCAO. El número de DM dependerá del sistema utilizado en cada caso, siendo mínimo de 2. El número de sensores de frente de onda también dependerá de la configuración elegida, no teniendo por qué ser igual al número de fuentes consideradas, aunque suele ser habitual que ambos términos coincidan como en este caso mostrado.

Su nombre de multi-conjugado se refiere a que los DM están conjugados para funcionar a diferentes alturas, es decir, cada uno corrige las turbulencias de una determinada capa. La mayor complejidad del sistema deriva de dividir la turbulencia atmosférica total en capas para cada uno de los DM. Además, la luz deberá reflejarse en todos ellos, por lo que la corrección de cada DM deberá realizarse teniendo en cuenta que las del resto de ellos.

El sistema permite una corrección más exacta para un FOV mayor que en los casos anteriores, tanto para OA nocturna como solar. Al igual que en GLAO, varios WFS apuntando en diferentes direcciones permitirán conocer la turbulencia para amplios campos de visión. La principal ventaja de este sistema es reducir el efecto del anisoplanatismo en las observaciones.

#### 4.3.4 MOAO (*Multi-Object Adaptive Optics*)

Se trata de una configuración diseñada también para corregir el problema del anisoplanatismo en los sistemas de óptica adaptativa, pero en este caso a diferencia de los anteriores trabajando en lazo abierto.

En lugar de corregir todo el campo de visión los sistemas MOAO están destinados a la corrección orientada a objetos, es decir, corregir una o varias zonas donde se encuentran las fuentes que se quieren observar. Estas zonas podrán tener un menor o mayor FOV en función de si se trata de una estrella, un cúmulo de estrellas o, en el caso de OA Solar, una o varias manchas solares.



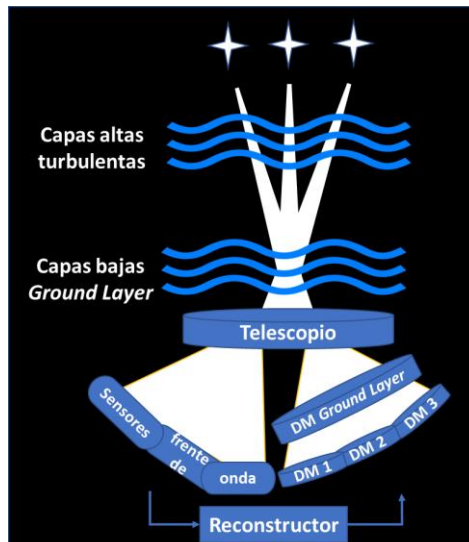


Figura 9: Representación esquemática de una posible configuración MOAO. Es importante resaltar que estas configuraciones trabajan en lazo abierto, por lo que a la salida de la pupila normalmente está situado un divisor de haz para que los diferentes WFS recojan el frente de onda aberrado sin haber atravesado los DM.

La configuración es flexible, de manera que el número de WFS o DM puede variar, pero frecuentemente contiene tantos WFS como estrellas guía se están considerando, apuntando a cada una de ellas, por lo que no tienen por qué encontrarse en el eje de observación. Por su parte, en la configuración más común existe un DM para cada objeto que se quiere observar, corrigiendo la turbulencia existente cada uno de ellos en las direcciones de esos objetos.

El motivo por el que estos sistemas se diseñan en lazo abierto es que no corrigen la turbulencia atmosférica que se encuentra en las direcciones de las estrellas guía. Todos los casos anteriores pretendían corregir toda la turbulencia atmosférica, y cuanto mayor fuese el FOV mejor. En este caso no, solamente interesa corregir la turbulencia en unas direcciones concretas que no tienen por qué ser las de los haces recibidos de las estrellas guía, por ello se necesita obtener en cada momento en los WFS la máxima información de las aberraciones posible.

#### 4.4 ÓPTICA ADAPTATIVA SOLAR

Una vez descrita la turbulencia atmosférica y los sistemas de óptica adaptativa tradicionales, es necesario mencionar en un apartado algunas de las características de la OA solar, diseñada para las observaciones diurnas, que ofrece una serie de dificultades y retos añadidos a los propios de la OA nocturna.

La estructura y componentes de un sistema de OA Solar son los mismos o, en algunas ocasiones, muy similares a los de OA nocturna con pequeñas modificaciones para su funcionamiento durante el día. Por ello toda la introducción sobre los sistemas de OA realizada en los apartados anteriores es perfectamente válida.

A lo largo de este apartado se describirán las principales diferencias y dificultades que implica realizar observaciones solares, así como las modificaciones realizadas a los sistemas de OA para adaptarlos a las observaciones solares.

##### 4.4.1 Cambios en la energía de la turbulencia

La OA solar implica realizar observaciones durante el día, cuando la luz del sol incide sobre la atmósfera terrestre dotándola de energía y calentando la superficie de la Tierra. Esto

implica dos factores en la turbulencia atmosférica; en primer lugar, la incidencia de los rayos del Sol dota de energía a las capas superiores, que la van cediendo al resto de capas a la vez que calienta el viento, provocando un mayor movimiento de la turbulencia. Por tanto, durante el día las capas superiores de la atmósfera serán más turbulentas y presentarán mayor fluctuación entre ellas que durante la noche.

Además, el calor de la superficie terrestre dará energía a *Ground-Layer*, como se había visto se trata de la capa que más aberraciones produce y aumentarle su energía dificultará aún más corregir esas aberraciones.

La consecuencia principal que implica el aumento de la turbulencia es la mayor variabilidad del parámetro  $r_0$ . En la Figura 10, se muestran datos reales medidos por Dunn Solar Telescope de la fluctuación de este parámetro con el tiempo.

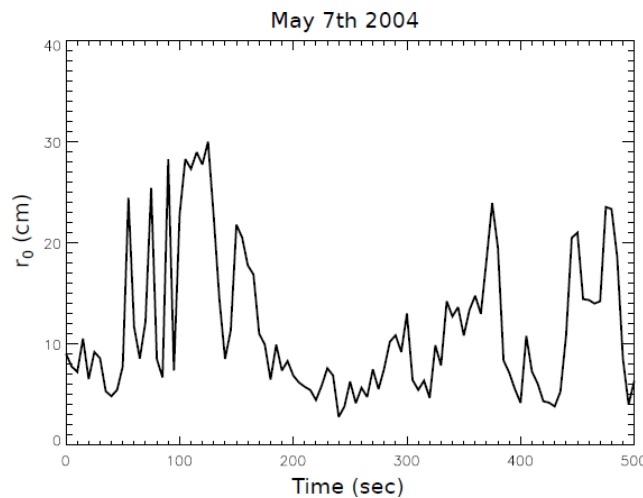


Figura 10: Fluctuación de la longitud de coherencia de Fried  $r_0$  en función del tiempo en una observación diurna realizada por el DST. Gráfica obtenida de [13].

#### 4.4.2 Cambios en la longitud del frente de onda

Es otra de las principales diferencias a la hora de trabajar con observaciones diurnas o nocturnas. En el caso de las nocturnas existe un gran abanico de longitudes de frente de onda que puedes ser empleadas, como el infrarrojo (IR) y el visible, siendo la primera normalmente la opción escogida.

En el caso de las observaciones diurnas, se tiene que trabajar siempre en las longitudes de onda del visible. Como se comentó en el apartado de la turbulencia atmosférica, existe una relación entre la longitud de onda  $\lambda$  y el parámetro  $r_0$ , cuando  $\lambda$  disminuye (como en el caso de pasar del IR al visible)  $r_0$  también, lo que implica que las condiciones para la observación son peores, presentando mayores aberraciones los frentes de onda recibidos.

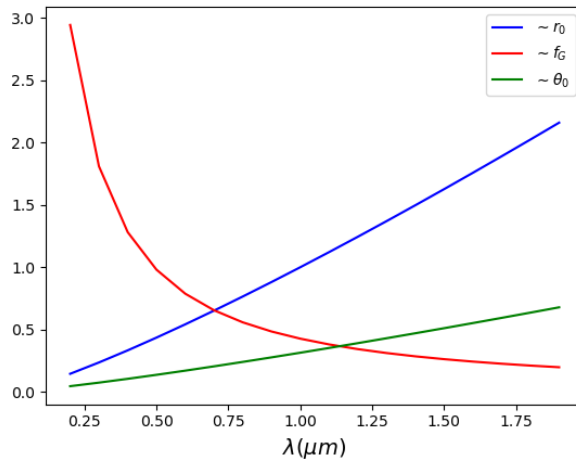


Figura 11: Parámetros  $r_0$ ,  $f_G$  y  $\theta_0$  en función de la longitud de onda. Se observan las ventajas que ofrece trabajar en rangos de longitudes de onda del infrarrojo como en el caso de observaciones nocturnas en comparación con realizar observaciones en el rango del visible como sucede en las observaciones diurnas.

Como se puede ver en la **¡Error! No se encuentra el origen de la referencia.**, la disminución del frente de onda implica también que la  $f_G$  aumente, significando que la turbulencia tiene mayor variación con el tiempo y que  $\theta_0$  disminuya, provocando que se afecte el fenómeno de anisoplanatismo.

#### 4.4.3 El sensor de frente de onda Shack-Hartmann Solar

La OA solar conlleva otra gran diferencia con la nocturna; la luz procedente de una estrella que se está observando es un punto luminoso sobre la matriz CCD del sensor de Shack-Hartmann, en otras palabras, un punto de luz desplazado del centro de cada una de las matrices de detección, la estrella se puede considerar como una fuente de luz puntual. Cuando se trata de observaciones solares, debido a su cercanía, el Sol satura completamente las imágenes y no puede ser ya considerado como una fuente puntual de radiación, pese a ser una estrella de tamaño mediano. Además, el uso de estrellas guía también queda limitado, ya que la radiación del Sol es demasiado intensa en todo el FOV de la observación como para utilizar una estrella guía.

Mientras que las imágenes obtenidas durante un periodo de observación nocturno permiten ver el centroide de manera clara, ya que la imagen recibida de cada fuente consiste en un punto de luz aberrado, cuando se trata del Sol la situación es completamente diferente [13]. El Sol satura todo el campo de visión, haciendo que las subaperturas del sensor de frente de onda Shack Hartmann no detecten la intensidad del haz de luz en un único conjunto de píxeles, sino que esta estará distribuida por toda la subapertura, como si de una mancha de luz se tratase. Una comparativa de imágenes obtenidas en periodo nocturno y diurno puede verse en la Figura 12.

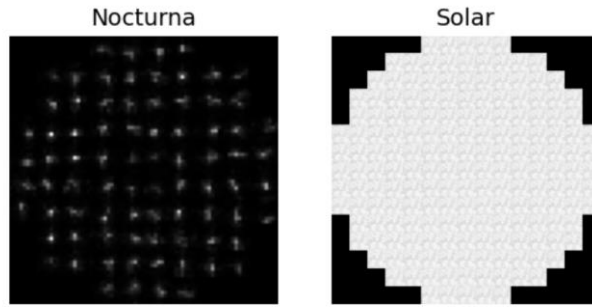


Figura 12: Comparativa de la radiación recibida por un SH en óptica adaptativa nocturna (izq.) y solar (dcha.). Las imágenes se han obtenido a partir de situaciones simuladas utilizando el simulador DASP [31] que se introducirá más adelante.

Los sensores SH utilizados en OA nocturna también son empleados en el caso de observaciones solares. Como en estas últimas no se puede calcular el centroide de cada subapertura, los SH emplean un algoritmo denominado algoritmo de correlaciones mediante el cual procesan la luz recibida con el objeto obtener unos valores medibles de un punto que represente el centro de gravedad de la luz recibida en cada subapertura, comparable a los centroides en el caso nocturno y que permita calcular las pendientes del frente de onda. El FOV en el caso de observaciones solares ha de ser suficientemente grande y contener un número considerable de gránulos solares para que el algoritmo empleado pueda trabajar de manera consistente [13].

El algoritmo calcula correlaciones cruzadas entre una imagen de referencia que se toma al comienzo de la observación, puede ser una de las subaperturas registradas y las imágenes recibidas en cada momento. El algoritmo calcula la correlación entre la imagen de referencia y la imagen que se recibe en cada momento de la siguiente forma:

$$CC(\vec{\Delta}_i) = \sum \sum I_M(\vec{x}) \times I_R(\vec{x} + \vec{\Delta}_i) \quad (76)$$

Donde  $I_M$  es la imagen obtenida de la subapertura,  $I_R$  es la imagen de referencia,  $\vec{\Delta}_i$  es el desplazamiento en píxeles entre ambas imágenes por tratarse de diferentes subaperturas y  $CC(\vec{\Delta}_i)$  son las correlaciones cruzadas para dicho desplazamiento. Esta cantidad se puede limitar a unos pocos píxeles asumiendo que las inclinaciones locales son pequeñas, consiguiendo reducir el número de sumas a realizar. Una de las alternativas es computar las correlaciones cruzadas haciendo uso de las transformadas de Fourier [8].

$$CC(\vec{\Delta}_i) = F^+ [F^- [I_S(x)] \times F^{-*} [I_R(x)]] \quad (77)$$

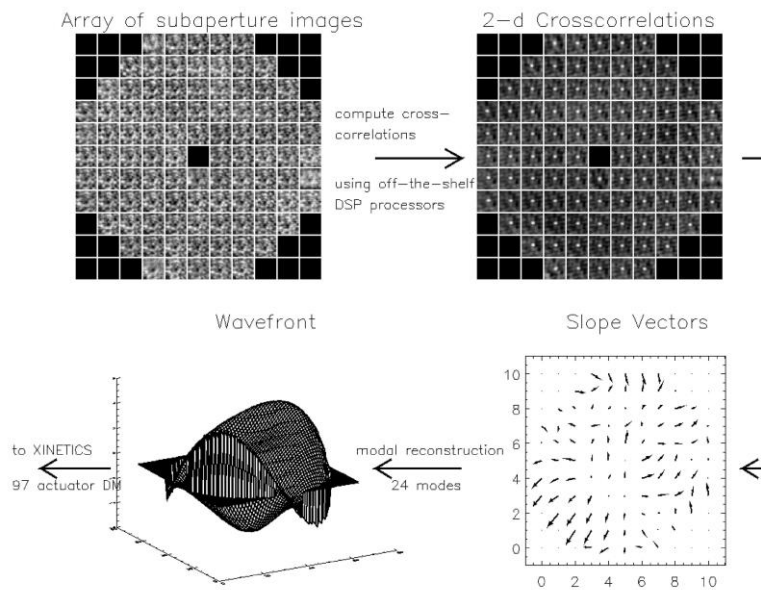


Figura 13: Determinación de las aberraciones de un frente de onda a través de las correlaciones cruzadas. Imagen extraída de [13].

Tal y como se mostró en la Figura 12 las imágenes obtenidas del Sol son similares a la mostrada en la parte superior izquierda de la Figura 13. Determinando el máximo de las correlaciones cruzadas, es posible determinar el desplazamiento entre la imagen en cuestión y la de referencia. Así, a través de dicho algoritmo, se pueden conseguir imágenes análogas a las obtenidas en periodo nocturno (similares a la figura superior derecha de la Figura 13) partiendo de imágenes solares. A partir de ellas se calcula el centroide en cada subapertura del sensor de frente de onda Shack Hartmann y se mide el gradiente local del frente de onda. Con la matriz de pendientes en cada subapertura (imagen inferior derecha de la Figura 13) se pueden estimar las aberraciones en los frentes de onda, reconstruyendo mediante los modos de Zernike (imagen inferior izquierda de la Figura 13). A partir de ellos ya se da por conocida la turbulencia y se calcula la corrección a aplicar por el DM. El principal desafío es computar este algoritmo en tiempo real entre imágenes obtenidas de las subaperturas y una imagen obtenida de una subapertura seleccionada al azar, que servirá como referencia.

Por tanto, la óptica adaptativa solar requiere de un proceso de cálculo adicional en comparación con las observaciones nocturnas. Dada la importancia de corregir las imágenes en tiempo real, este proceso debe llevar el menor tiempo posible empleando para ello los medios informáticos necesarios.

#### 4.4.4 Algunos sistemas de OA Solar en funcionamiento actualmente

##### 4.4.4.1 Telescopio Solar Sueco (SST)

El SST es un telescopio de vacío de 97 cm de apertura ubicado en La Palma, España. Es uno de los ejemplos de telescopios de óptica adaptativa solar que existen en España. Se caracteriza por contar con un único DM combinado con un único SHWFS. El DM consta de 37 elementos y se trata de un espejo bimorfo. El SHWFS por su parte consta de 37 subaperturas, cada una de ellas con forma hexagonal, de tamaño  $d = 13.8 \text{ cm}$  [13]. Esto da lugar a que el sistema corrija aproximadamente 35 modos de Zernike. Cada una de las subaperturas consta de una cámara CCD de 955 *fps*, por lo que el ancho de banda está limitado a 65 Hz.

Gracias a este telescopio, se han obtenido algunas de las imágenes de la superficie Solar con mayor calidad hasta la fecha [13]. En cambio, la corrección de 35 modos de Zernike se considera insuficiente en comparación con las que ya se están obteniendo por otros sistemas, siendo limitante para el telescopio ya que solo es capaz de realizar observaciones de calidad bajo muy buenas condiciones atmosféricas. Actualmente se está trabajando en actualizarlo con un DM bimorfo de 85 elementos que permitirá ampliar el abanico de condiciones de observación.

#### 4.4.4.2 Kiepenheuer institute (KAOS)

El sistema de Óptica Adaptativa de este instituto está localizado en el telescopio VTT, localizado en la isla de Tenerife desde 2002 [13]. El sistema es parecido al SST tanto en diseño como en implementación, consta de un sensor del frente de onda con 36 subaperturas dispuestas de forma hexagonal, con  $d = 10\text{cm}$  de tamaño de subapertura y un espejo deformable bimorfo de 35 elementos.

El campo de visión de la apertura (FOV) es de  $12'' \times 12''$  haciendo un muestreo de  $24 \times 24$  píxeles en cada subapertura. Se ha visto que incrementar el uso de píxeles en los detectores mejora el resultado del sistema, de ahí que se haya volcado parte del esfuerzo en estas líneas de investigación.

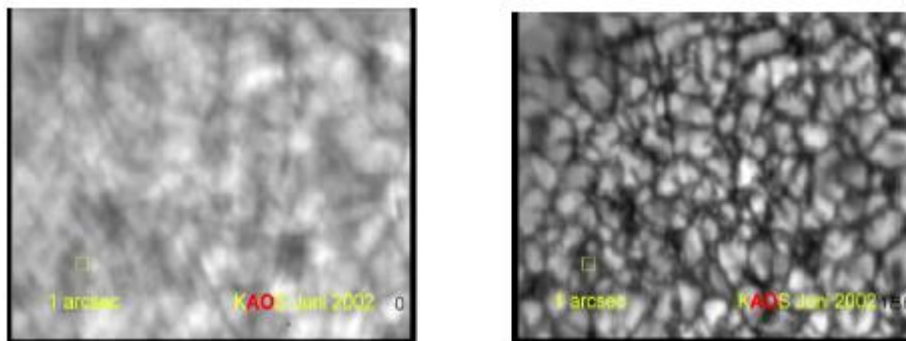


Figura 14: Ejemplo de granulación solar corregida por el sistema KAOS extraído de [13].

#### 4.4.4.3 McMath-Pierce

La peculiaridad de este sistema es que se ha desarrollado para trabajar específicamente en la zona infrarroja [13]. El sistema se ha desarrollado para trabajar en combinación con el telescopio McMath-Pierce situado en el Kitt Peak National Observatory en Arizona, Estados Unidos. Se trata de un telescopio solar de  $1.6\text{m}$  de diámetro construido a principios de la década de 1960.

Se basa en un sistema con 37 actuadores electrostáticos con espejos de membrana deformable en el telescopio McMath-Pierce.

#### 4.4.4.4 Otros sistemas actualmente en desarrollo

Actualmente existen varios sistemas de AO Solar siendo desarrollados en el observatorio Chida (Japón) y también se están intentando hacer ciertas mejoras al THEMIS, localizado en Tenerife [13]. El sistema THEMIS utiliza 66 subaperturas Shack-Hartmann trabajando a  $1000\text{ fps}$  y 59 micro-actuadores desarrollados por OKOTECH. El control modal se apoya en un procesador Intel Core Quad.

#### 4.5 SIMULADOR DE ÓPTICA ADAPTATIVA – PLATAFORMA DASP

La plataforma DASP (*Durham Adaptatives Simulation Platform*) [31], [32] consiste en un simulador de óptica adaptativa y turbulencia atmosférica desarrollado por la Universidad de Durham. El simulador DASP permite modelar diferentes configuraciones de sistemas de óptica adaptativa que pueden ser personalizados completamente, como el número de estrellas guía, el número de sensores de frente de onda, el número de espejos deformables... Con todo ello se pueden simular diferentes configuraciones de las mostradas anteriormente como SCAO, GLAO, MCAO, etc. tanto para el caso de óptica adaptativa nocturna como para óptica adaptativa solar.

La turbulencia atmosférica simulada se genera a partir del modelo estadístico de Von Karman implementada con simulaciones de Montecarlo. El haz incidente se perturba en diferentes capas definidas por el propio usuario, tanto en intensidad como en altura, velocidad del viento y dirección de este, etc.

La plataforma permite guardar datos de todos los procedimientos de la simulación, en el caso del SH permite guardar tanto las imágenes de la subapertura, las correlaciones (en el caso Solar) y las pendientes del frente de onda medidas por el sensor en cada una de las subaperturas. En el caso del DM permite obtener los coeficientes de Zernike (calculados mediante el procedimiento de LS) y los voltajes de los actuadores. En el caso de las capas turbulentas, si se generan más de una, permite obtener imágenes de cada una de las capas turbulentas y los valores de cómo desfasan el frente de onda en cada una de las regiones. Otro aspecto particularmente útil es que permite obtener los coeficientes de Zernike ideales que representan la fase simulada, se trata de los valores exactos para esa fase a diferencia de los otros obtenidos mediante reconstrucción.

Para cada una de las simulaciones es necesario generar un fichero de parámetros en dónde se indican los valores de todos los parámetros que se quieren determinar como el  $r_0$ ,  $L_0, \dots$ . A los parámetros no indicados se les asignará un valor por defecto. En este fichero es necesario también indicar qué valores de la simulación se van a guardar y el nombre del archivo en el que se guardarán.





## 5 FUNDAMENTOS DE INTELIGENCIA ARTIFICIAL.

---

Cada vez con mayor frecuencia se escucha el concepto de Inteligencia Artificial (IA) en el debate público. Términos como, algoritmo, datos son frecuentes en el vocabulario del día a día, aunque siguen siendo tratados como conceptos que no están al alcance de la comprensión de la mayoría de la población. El objetivo de este capítulo es realizar una breve introducción de estos conceptos, necesarios para la obtención de los resultados mostrados en el último capítulo.

Dar una definición concreta de la inteligencia artificial resulta ser una tarea compleja, ya el propio término de qué es inteligencia genera un debate en sí mismo, lo que implica que surjan múltiples definiciones de 'inteligencia artificial'. Una de las más aceptadas sería del estilo de la siguiente: Campo de estudio de la informática, y por tanto de la física computacional, que se encarga de investigar en el desarrollo de los ordenadores para que sean capaces de emular comportamientos inteligentes como el reconocimiento de voz o el análisis de patrones [33].

El origen de la Inteligencia Artificial se remonta a mediados de la década de 1980 y, como suele ser habitual, con una motivación bélica, pues el mundo se encontraba inmerso en la Guerra Fría. El objetivo que se pretendía por aquel entonces era obtener sistemas de control capaces de manejar otros instrumentos de forma inteligente a partir de ordenadores analógicos. El desarrollo de estos dispositivos, como se verá más adelante, hizo que tanto su estructura como la forma de operar guardase semejanza con estructuras de inteligencia biológicas. Por todo ello se les conoció como redes neuronales artificiales (ANN, por sus siglas en inglés, *Artificial Neural Network*).

Las redes neuronales son uno de los paradigmas de programación más eficientes y pragmáticos jamás inventados. En un enfoque de programación tradicional, los procesos se resuelven dando el usuario órdenes o pasos sobre cómo resolverlo. No es así en el caso de una red neuronal, donde los sistemas son capaces de predecir y aprender a partir de las situaciones físicas que se les presentan, obteniendo su propia solución al problema.

Hasta 2006 no se conocía como aplicar las redes neuronales salvo para sistemas de control muy básicos, y por ello no es a partir de finales de esa década cuando tanto los términos como sus aplicaciones comienzan a aparecer en el día a día. Lo que cambió en 2006 fue el descubrimiento de técnicas de aprendizaje en las llamadas redes neuronales profundas [34]. Estas técnicas conocidas como aprendizaje profundo (*Deep Learning* en inglés), han continuado desarrollándose aún más y, en la actualidad, logran un rendimiento sobresaliente en muchos problemas importantes como son la visión por computadora, el reconocimiento de voz y el procesamiento del lenguaje natural. Grandes compañías de diferentes rangos implementan estos sistemas en la actualidad como son Microsoft, Google, Amazon, Tesla, etc.

### 5.1 LAS REDES NEURONALES ARTIFICIALES (ANN)

Se conocen con la abreviación ANN a las redes neuronales artificiales [35], se trata de una familia de algoritmos capaces de dar una respuesta numérica a unos datos, y que son capaces de modificar esa respuesta modelizando sistemas físicos o matemáticas. La estructura y forma de procesar los datos se asemejan a la organización biológica del cerebro, de ahí su nombre y algunos de los nombres de sus componentes.

El elemento fundamental de una ANN es la neurona, es la unidad computacional de la red que aplica una función matemática a los datos que recibe como entrada (input) obteniendo

así una salida (output). Las neuronas, en función de la complejidad de la red, se agrupan en diferentes capas estando interconectadas entre sí, formando una red, de manera similar al caso biológico. La profundidad del aprendizaje aumenta a medida que aumenta el número de capas. Existen tres grupos de capas principales:

- Capa de entrada: Recibe los datos para los que se quiere obtener una respuesta a través de la ANN.
- Capas ocultas: Es el nombre que reciben todas las capas intermedias que hay entre la capa de entrada y la capa de salida. A mayor número de capas mayor complejidad de la ANN. En ellas se realizan la mayoría de los cálculos que darán lugar a la respuesta de la red.
- Capa de salida: Última capa de la ANN que recoge las salidas de las capas ocultas y proporciona un output al problema en función de los datos de entrada dados.

En la estructura típica de una ANN, las neuronas de cada capa están interconectadas con todas las neuronas de la capa anterior y de la siguiente. Cada una de ellas aplicará una función matemática a los datos que reciba mediante la cual obtendrá una salida que transmitirá a las neuronas de la capa siguiente. Existen diversas funciones que pueden ser utilizadas como funciones de activación, el uso de una u otra dependerá de cuál de ellas se adapte mejor al problema a modelar en función del estudio de los datos tanto de entrada como de salida (rango de valores, números solo positivos o negativos como entrada, etc.). Algunas de las funciones más utilizadas son: sigmoide, tangente hiperbólica, la unidad lineal rectificadora (ReLU por sus siglas en inglés) o la unidad lineal rectificadora paramétrica (PReLU por sus siglas en inglés) [36], [37].

Las salidas de cada neurona serán transmitidas como entradas de la capa siguiente, cada una de las conexiones entre neuronas está regulada por un peso, ese valor regula la relevancia de la información dada por esa conexión a la neurona siguiente en comparación con los datos recibidos del resto de neuronas. El valor del peso se mueve entre 0 y 1 y se va modificando a lo largo del entrenamiento de la red. La modificación de esos pesos es uno de los métodos de las ANN para ajustar las salidas, obteniendo el resultado con el menor error posible [37].

Además de esos pesos, otro de los parámetros importantes es el parámetro de sesgo, o *bias* en inglés, que también puede ser modificado durante el proceso de entrenamiento. Por tanto, el caso más simple de una neurona que reciba como input un único valor ( $x$ ) y obtenga un output ( $x'$ ) seguirá el siguiente proceso:

$$x \rightarrow x' = f(b + \omega x) \quad (78)$$

Siendo  $f$  la función de activación,  $b$  el parámetro bias y  $\omega$  el peso de la conexión correspondiente. El proceso de propagación hacia adelante (*feedforward*) explica cómo avanza la información en las redes neuronales. La neurona  $j$  –ésima de la capa  $k$  vendrá dada por [34]:

$$z_j^{(k)} = \sum_{i=1}^{n_{k-1}} \omega_{ij}^{(k)} a_i^{(k-1)} + b_j^{(k)} \text{ siendo } a_j^{(k)} = f(z_j^{(k)}) \quad (79)$$

Donde  $z_j^{(k)}$  es la suma de la neurona  $j$  en la capa  $k$ ,  $n_{k-1}$  es el número de neuronas de la capa  $(k - 1)$  –ésima,  $\omega_{ij}^{(k)}$  son los pesos asociados a la neurona  $j$  en la capa  $k$ ,  $b_j$  es el término de sesgo asociado a cada neurona.

Como se ha comentado durante el entrenamiento la ANN modifica los pesos y los parámetros de sesgo en busca de la mejor solución mediante la técnica conocida como la retropropagación del error [38]. Existen múltiples métodos para llevar a cabo este problema de optimización, como por ejemplo el método de Newton, el método del gradiente conjugado, o el más empleado, el método del descenso del gradiente. A continuación, se describirá el método más utilizado.

En el proceso de entrenamiento de la red neuronal, el proceso más importante es conocer cómo influye la información del error hacia atrás entre las diferentes capas de la red. El entrenamiento puede llevarse a cabo de dos formas diferentes:

- Aprendizaje supervisado: En este caso el proceso de entrenamiento se realiza con un conjunto de datos denominado datos de entrenamiento en los que se conoce tanto las entradas como las salidas. De esta forma los outputs de la red se pueden comparar con las salidas originales y en función de ello calcular el error.
- Aprendizaje no supervisado: Los datos de salida durante el proceso de aprendizaje son desconocidos y la red busca patrones y relaciones entre los valores de salida.

Durante el entrenamiento, la manera que tiene la red de validar los valores de los parámetros es a través de un subconjunto de los datos de entrenamiento llamado conjunto de validación. Una vez la red ha sido entrenada se le introducen como datos de entrada un conjunto de valores llamado conjunto de prueba para comprobar si los resultados que produce son los esperados. El proceso de entrenamiento de la red puede no ser exitoso por dos fenómenos no deseados; el sobre-entrenamiento, más comúnmente conocido con el término inglés *overfitting*, que se produce cuando la red entiende que todos los detalles del conjunto de entrenamiento, incluido el posible ruido, son detalles para aprender y esto le impide generalizar. La red aportará un error muy bajo cuando se prueba utilizando datos de entrenamiento, pero al utilizar nuevos datos desconocidos las salidas no serán tan buenas como se esperaba. El otro fenómeno es el inverso, el infra-entrenamiento, también más conocido por el término inglés *underfitting*. En este caso lo que sucede es que la ANN es incapaz de modelizar los datos de entrenamiento ni generalizarlos de manera correcta, comúnmente debido a no tener suficientes datos de entrenamiento para poder obtener los patrones o relaciones necesarias.

Suponiendo el caso de un aprendizaje supervisado (es el más común y el que se utilizará en los estudios posteriores), cuando las salidas de las últimas capas llegan a la capa  $L$  (se corresponde con la última capa de la red neuronal), se han de comparar estos resultados con los valores reales que debería obtener la red  $y$ . Es decir, hay que hacer una evaluación de cómo son los resultados reales frente a los esperados. Para ello haremos uso de la función error o de coste, que puede ser de varias formas: error cuadrático medio, entropía cruzada... Se define la función error cuadrático medio como sigue [34], [39]:

$$C = \frac{1}{2} \|y - a^L\|_2^2 = \frac{1}{2} \sum_j (y_j - a_j^L)^2 \quad (80)$$

El mismo razonamiento se aplicaría para cualquiera de las funciones de coste  $C$ . Lo que se busca es encontrar una colección de pesos y bias  $\omega_{ij}^{(k)}, b_j^{(k)}$  para cualquier capa  $k$  y neurona  $j$  de tal manera que la función de coste  $C$  sea mínima. Se busca resolver:

$$\min_{\forall i, j \in [1, \dots, n_k], \forall k \in [1, \dots, L]} C(\omega_{ij}^{(k)}, b_j^{(k)}) \quad (81)$$

### 5.1.1.1 Método del descenso del gradiente

Los problemas de minimización de funciones de coste son clave en el proceso de aprendizaje de las redes. En función de la morfología de la red este problema podrá ser más o menos complejo, ya que dependerá del número de pesos y parámetros de sesgo a modificar. El método del descenso del gradiente puede ser utilizado para la resolución de complejos problemas de minimización [34], [39].

Siendo  $C$  la función que se quiere minimizar, esta podrá depender de multitud de variables ( 81 ), para exponer un ejemplo sencillo se considerará una función de coste que dependa únicamente de dos variables  $C(v_1, v_2)$ . El objetivo es determinar en qué punto  $(v_1, v_2)$  la función  $C$  alcanza un mínimo. Si la función  $C$  tiene una expresión analítica sencilla, el procedimiento del cálculo del mínimo es sencillo:

1. Se calcula el gradiente de la función de coste y se iguala 0. Es decir:

$$\begin{aligned} \frac{\delta C(v_1, v_2)}{\delta v_1} &= 0 \\ \frac{\delta C(v_1, v_2)}{\delta v_2} &= 0 \end{aligned} \quad (82)$$

2. Una vez conocidos esos puntos denotados por  $\hat{v}_i = (v_{1i}, v_{2i})$  se construye la matriz Hessiana de la función de coste:

$$H(C(v_1, v_2)) = \left( \frac{\delta^2 C(v_1, v_2)}{\delta v_i \delta v_j} \right)_{1 \leq i, j \leq 2} \quad (83)$$

3. Se evalúan los menores principales de la hessiana en cada extremo y si  $\forall i$  se cumple que  $|H_i| > 0$ , entonces  $C(v_1, v_2)$  alcanza un mínimo relativo en dicho punto.

El problema así visto con solamente dos variables parece sencillo, pero con la complejidad de una red neuronal y todos los parámetros a optimizar resulta muy difícil obtener una solución analítica al problema.

La **¡Error! No se encuentra el origen de la referencia.** representa el funcionamiento del algoritmo de descenso del gradiente. En primer lugar, se considera un punto inicial sobre esa superficie denominado  $v_0$  y se calcula el valor de  $\nabla C$  en él. Con ello se habrá obtenido un vector que indica cómo crece la función error en las diferentes direcciones. El siguiente paso consiste en calcular el punto  $v_0 - \nabla C$ , de manera que se encontrará más cerca del mínimo, y repetir el

paso de calcular  $\nabla C$  en él. Repitiendo de forma iterativa este procedimiento cada vez se estará más cerca del mínimo de la función de coste [34], [39], [40].

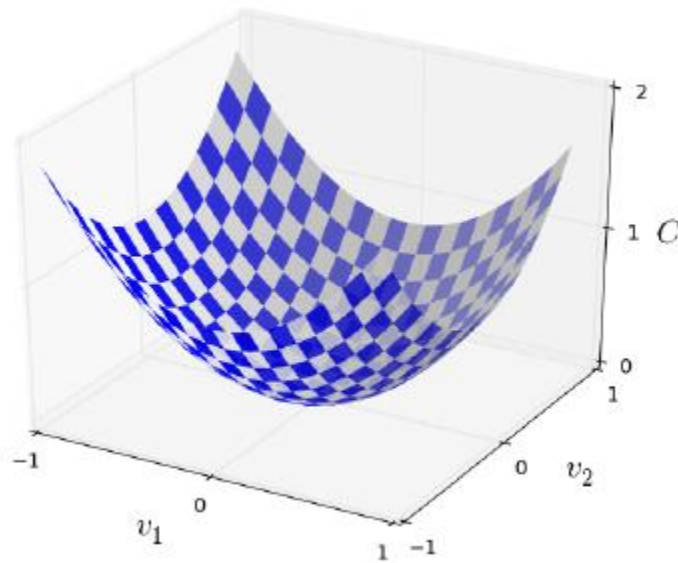


Figura 15: Representación del funcionamiento del algoritmo del gradiente. Considerándose un punto inicial  $v_0$  en la superficie y se calcula  $\nabla C$  en dicho punto, se obtiene un vector que indica cómo de rápido crece la función  $C$  en cada una de las direcciones respecto de las que se deriva. Equivalentemente,  $-\nabla C$  indica cómo decrece la función de coste en cada una de las direcciones. Si ahora se le resta al punto  $v_0$  la cantidad  $\nabla C$ , se habrá descendido una cantidad  $v_0 - \nabla C$ , estando más cerca del mínimo. Este proceso se debe de repetir de forma iterativa hasta llegar a un entorno del mínimo de la función de tal manera que se consigue calcular dicho punto sin necesidad de hacer cálculos complejos [34], [39], [40].

Aunque el uso de este algoritmo está bastante extendido en la literatura [34], [39], es común encontrar otros métodos de optimización de parámetros más rápidos y con velocidades de convergencia algo mejores, como el método del descenso del gradiente estocástico y el método Adam [41].

#### 5.1.1.2 Algoritmo de retropropagación (backpropagation)

El algoritmo de *backpropagation* consiste en entender cómo el cambio de pesos y bias afecta al comportamiento de la función de coste y es la técnica que con mayor frecuencia se aplica en IA para modificar los valores de los pesos y bias durante el proceso de entrenamiento.

Para explicar este algoritmo, en primer lugar, se debe introducir  $\delta_j^l$  [7], el cuál designa el error en la predicción de la neurona  $j$  –ésima en la capa  $l$  –ésima:

$$\delta_j^l = \frac{\delta C}{\delta z_j^l} \quad (84)$$

Siendo  $C$  la función de coste al igual que en los apartados anteriores. Se entiende también que el término  $\delta^l$  engloba el error total cometido en la capa  $l$  –ésima. Para modificar los parámetros de la red neuronal es necesario relacionar ese error con  $\frac{\delta C}{\delta \omega_{jk}^l}$  y  $\frac{\delta C}{\delta b_j^l}$ , es decir, con los términos que se deben optimizar.

En el caso de la capa externa (que viene definida por  $l = L$ ) ese error se puede calcular de la forma [34]:

$$\delta_j^L = \frac{\delta C}{\delta z_j^L} \quad (85)$$

Que teniendo en cuenta la regla de la cadena se puede calcular como:

$$\delta_j^L = \sum_{k=1}^{n_L} \frac{\delta C}{\delta a_k^L} \cdot \frac{\delta a_k^L}{\delta z_j^L} \quad (86)$$

Siendo como en apartados anteriores  $a_k^L$  el resultado de aplicar la función de activación en la neurona  $k$  – ésima de la capa  $L$ . Por tanto, en ese sumatorio, ese valor únicamente será distinto de 0 cuando  $j = k$ , en término matemáticos  $\frac{\delta a_k^L}{\delta z_j^L} = \delta_{jk}$  siendo el resultado la Delta de Didac. Por tanto, se obtiene que:

$$\delta_j^L = \frac{\delta C}{\delta a_j^L} \cdot \frac{\delta a_j^L}{\delta z_j^L} \quad (87)$$

Pero teniendo en cuenta que  $a_j^{(k)} = f(z_j^{(k)})$ , su derivada parcial será por tanto la derivada de la función de activación, de manera que [34]:

$$\delta_j^L = \frac{\delta C}{\delta a_j^L} f'(z_j^{(k)}), \forall j \in \{1, \dots, n_L\} \quad (88)$$

Que escrita en forma vectorial:

$$\delta^L = \nabla_a C \odot f'(z^L) \quad (89)$$

Siendo  $\odot$  el producto de Hadamard [42].

Hasta aquí, por tanto, se ha obtenido el error para el caso de las neuronas de la última capa, si se calculase para cualquier otra capa siguiendo el mismo procedimiento que antes:

$$\delta_j^l = \frac{\delta C}{\delta z_j^l} \sum_{k=1}^{n_{l+1}} \frac{\delta C}{\delta z_k^{l+1}} \cdot \frac{\delta z_k^{l+1}}{\delta z_j^l} = \sum_{k=1}^{n_{l+1}} \frac{\delta z_k^{l+1}}{\delta z_j^l} \cdot \delta_k^{l+1} \quad (90)$$

Evaluando el primer término para cualquier capa de la red neuronal, hay que tener en cuenta que:

$$z_k^{l+1} = \sum_{j=1}^{n_l} \omega_{kj}^{l+1} a_j^l + b_k^{l+1} = \sum_{j=1}^{n_l} \omega_{kj}^{l+1} f(z_j^l) + b_k^{l+1} \quad (91)$$

Por tanto, la derivada parcial:

$$\frac{\delta z_k^{l+1}}{\delta z_j^l} = \omega_{kj}^{l+1} f'(z_j^l) \quad (92)$$

Volviendo entonces sobre el cálculo del error de la neurona  $j$  – ésima:

$$\delta_j^l = \sum_{k=1}^{n_{l+1}} \omega_{kj}^{l+1} \delta_k^{l+1} f'(z_j^l) \quad (93)$$

O expresándolo en términos del producto de Hadamard se obtendrá que:

$$\delta^l = (\omega^{l+1} \cdot \delta^{l+1}) \odot f'(z^l) \quad (94)$$

Una vez obtenido este término hay que relacionar las ecuaciones que miden variaciones en la función de coste en relación con los parámetros de sesgo y los pesos, teniendo en cuenta que:

$$a_j^l = f(z_j^l) = f\left(\sum_{k=1}^{n_{l-1}} \omega_{kj}^l a_k^{l-1} + b_j^l\right) \quad (95)$$

Así que para relacionar la función de coste con los pesos habría que aplicar al igual que antes la regla de la cadena:

$$\frac{\delta C}{\delta \omega_{kj}^l} = \frac{\delta C}{\delta a_j^l} \cdot \frac{\delta a_j^l}{\delta z_j^l} \cdot \frac{\delta z_j^l}{\delta \omega_{kj}^l} \quad (96)$$

Un factor del producto ya es conocido de los desarrollos anteriores  $\delta_j^l$ , el otro puede obtenerse derivando  $z_j^l$  para un peso determinado:

$$\frac{\delta C}{\delta \omega_{kj}^l} = \delta_j^l a_k^{l-1} \quad (97)$$

Y habría que realizar lo mismo relacionando la función de coste con el parámetro de sesgo:

$$\frac{\delta C}{\delta b_j^l} = \frac{\delta C}{\delta z_j^l} \cdot \frac{\delta z_j^l}{\delta b_j^l} \quad (98)$$

Del cuál se obtiene el siguiente resultado [34]:

$$\frac{\delta C}{\delta b_j^l} = \delta_j^l \quad (99)$$

Obtenido este resultado, ya se tienen todas las ecuaciones fundamentales del algoritmo de *backpropagation*, que se resumen en 4 [34]:

$$\delta^L = \nabla_a C \odot f'(z^L) \quad (100)$$

$$\delta^l = (\omega^{l+1} \cdot \delta^{l+1}) \odot f'(z^l) \quad (101)$$

$$\frac{\delta C}{\delta \omega_{kj}^l} = \delta_j^l a_k^{l-1} \quad (102)$$

$$\frac{\delta C}{\delta b_j^l} = \delta_j^l \quad (103)$$

Estas cuatro ecuaciones gobiernan el funcionamiento del algoritmo en el aspecto analítico y obtienen las variaciones de la función de coste en función de los pesos y parámetros de sesgo, que se van a ir actualizando durante el proceso de entrenamiento. Este proceso de actualización debe realizarse de forma iterada para encontrar la combinación de valores que minimizar la función de coste.

El algoritmo de *backpropagation* seguirá entonces los siguientes pasos, gobernados por las ecuaciones anteriores:

1. **Entrada:** se activa la primera capa  $a^1$  con el conjunto de datos de entrada que se pasan a la red.
2. **Feedforward o propagación hacia delante:** para cada  $l = 2, 3, \dots, L$  se calcula  $z^l = \omega^l a^{l-1} + b^l$  siendo  $a^l = f(z^l)$ .
3. **Error de salida:** se calcula el error para la última  $\delta^L = \nabla_a C \odot f'(z^L)$ .
4. **Backpropagation o propagación hacia atrás:** para cada  $l = L - 1, L - 2, \dots$  se calcula  $\delta^l = (\omega^{l+1} \cdot \delta^{l+1}) \odot f'(z^l)$ , obviamente para realizar este cálculo se debe ir desde las capas más cercanas a las capas de salida hasta llegar a la capa de entrada.
5. **Salida:** se busca el mínimo de la función de coste actualizando los valores de los pesos y los parámetros de sesgo.

Este proceso es aplicable a todos los tipos de red que existen, en este caso se el modelo se ha explicado concretamente para una red MLP (*Multi-Layer Perceptron*) ya que como se verá más adelante, se trata del modelo con la morfología más sencilla. No obstante, el algoritmo puede adecuarse a la particularidad de cada red, tipo de red, morfología, función de coste escogida, etc.

En todo modelo de red neuronal existe una serie de parámetros que se pueden ajustar con el objetivo de obtener el mejor entrenamiento posible. No hay una combinación de estos parámetros que garantice un mejor funcionamiento, sino que dependerá del problema concreto que se trate de resolver. Por ello, cuando se exponen los resultados obtenidos por una ANN es importante conocer el ajuste de estos parámetros, al igual que también se determina el modelo de red neuronal utilizada o la morfología de esta.

- **Batch size:** Cuando se entrena una red neuronal lo más común no es emplear el conjunto de entrenamiento al completo antes de que se retropropague el error, sino que se define el tamaño del lote (*batch size* en inglés) como un subconjunto del conjunto de datos de entrenamiento que procesa la red neuronal antes de que se retropropague el error. Su valor puede ir desde la unidad (retropropagar el error tras cada muestra) hasta el tamaño total de muestras del set de datos de entrenamiento.
- **Epoch:** Las épocas (*epochs* en inglés) es el parámetro que determina el número de veces que la red recibe el conjunto de entrenamiento completo durante el proceso de aprendizaje.
- **Loss:** Es el nombre que comúnmente recibe la función de coste. Como se ha visto, en cualquier modelo ha de existir una función que cuantifique como de grande o de pequeña es la diferencia entre los resultados predichos por el



modelo y los resultados conocidos. En las redes neuronales este parámetro se conoce como pérdida (*loss* en inglés). Un ejemplo de esta función es el error cuadrático medio.

- **Metrics:** La métrica es la función que calcula la tasa precisión de todas las predicciones llevadas a cabo por la red neuronal, de manera que la métrica proporciona información sobre como ajusta el modelo a los datos reales. En este sentido, la pérdida previamente definida se considera como un tipo particular de métrica.
- **Learning rate:** En el algoritmo de optimización, es el parámetro que indica el tamaño de paso en cada iteración en la búsqueda del mínimo de la función de coste.
- **Momentum:** Se trata de otro término del algoritmo de optimización, al calcular el siguiente paso en la búsqueda del mínimo de la función de coste, establece un combinación lineal entre el paso de la actual iteración y el incremento anterior, aplica los pasos una “inercia” que permite al algoritmo no detenerse en posibles mínimo locales.

## 5.2 EL PERCEPTRON MULTICAPA (*MULTI-LAYER PERCEPTRON*)

Se trata del modelo más sencillo de red neuronal artificial, denominado comúnmente MLP por sus siglas en inglés. El diseño se inspiró en el de una red biológica y es el que se suele tomar como ejemplo para explicar el funcionamiento de las redes neuronales, así como los algoritmos de *backpropagation*, etc. debido a su sencillez.

El perceptrón simple se trata de uno de los primeros elementos clave para la IA; desarrollado por Frank Rosenblatt [34], [39], consistió en un sistema con una neurona al que se le otorgaban unos datos y ofrecía otros como salida. Su simplicidad limita bastante el margen de actuación, ya que no se pueden conseguir grandes resultados con una única unidad. Es por ello por lo que Rosenblatt, para conseguir sistemas con una mejor capacidad de clasificación y de interpretación de datos, consideró agrupar los perceptrones en capas, tal y como ocurre con las neuronas biológicas en nuestro cerebro. Aparecía por primera vez un sistema inteligente artificial basado en los mismos principios que gobernaban el comportamiento de nuestro sistema nervioso. Comenzó la época de la Inteligencia Artificial.

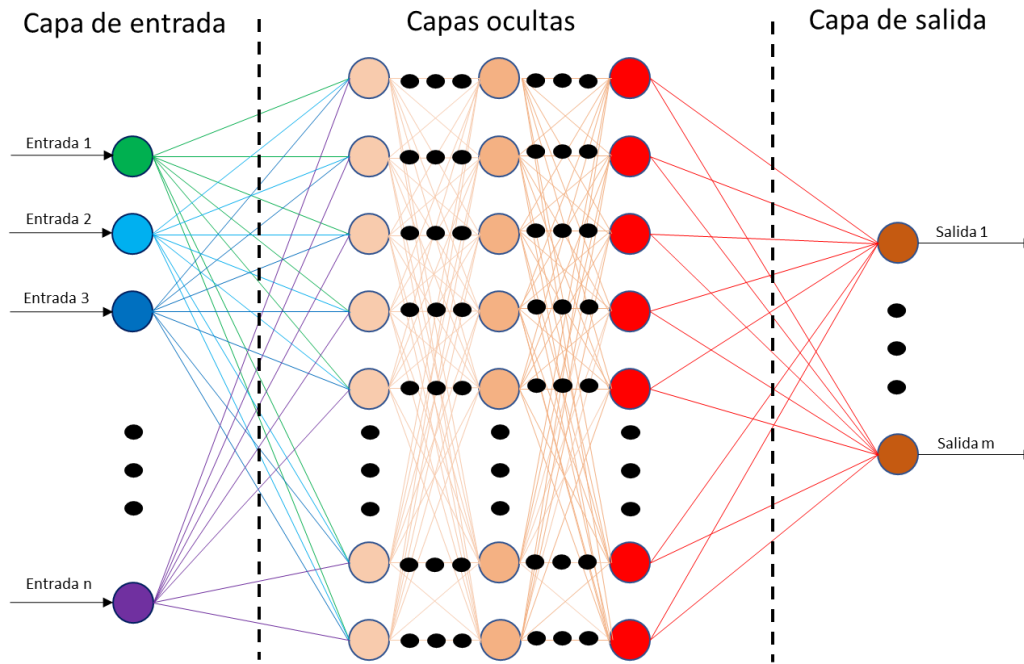


Figura 16: Estructura simplificada de una red neuronal artificial. Consiste en una capa de entrada (inputs), una o varias capas ocultas (donde se hacen el grueso de las operaciones) y la capa de salida (outputs).

La forma más sencilla para entender el funcionamiento de las MLP es ver con un modelo sencillo la forma en la que se obtienen los resultados, para compararlo después con los algoritmos expresados previamente. Suponiendo el modelo de red neuronal de la **¡Error! No se encuentra el origen de la referencia.**, se debe recordar la notación empleada. Con  $z_j^l$  se denota la entrada de la neurona  $j$  en la capa  $l$  y con  $\omega_{kj}^l$  se denomina el peso que va desde la neurona  $k$  de su capa  $l$  a otra neurona  $j$ .

Partiendo de los parámetros de entrada  $x_1$  y  $x_2$ , las conexiones en este caso irían de izquierda a derecha:

$$\begin{aligned} a_1^1 &= f(b_1^1 + \omega_{11}^1 x_1 + \omega_{21}^1 x_2) \\ a_2^1 &= f(b_2^1 + \omega_{12}^1 x_1 + \omega_{22}^1 x_2) \end{aligned} \quad (104)$$

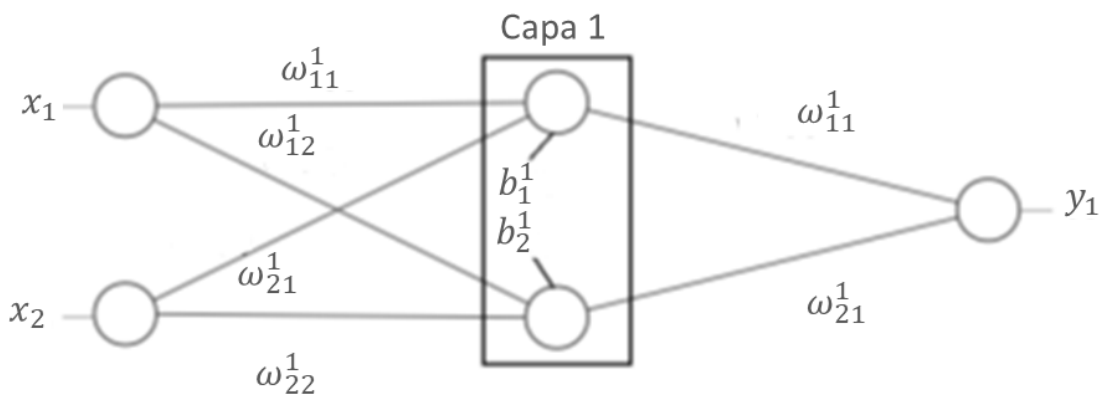


Figura 17: Representación esquemática de una MLP sencilla consistente en dos neuronas de entrada, una única capa oculta con dos neuronas y una neurona de salida. Es un modelo útil para comprender los algoritmos detrás de estos modelos.

Luego, el resultado final que se obtiene de esta red neuronal será:

$$y_1 = f\left(b_1^2 + \omega_{11}^2 f(b_1^1 + \omega_{11}^1 x_1 + \omega_{12}^1 x_2) + \omega_{21}^2 f(b_2^1 + \omega_{12}^1 x_1 + \omega_{22}^1 x_2)\right) \quad (105)$$

Obviamente estas expresiones resultarán más complejas a medida que se utilicen redes neuronales de mayores dimensiones, más neuronas por capa y mayor número de capas. Los datos de entrada de este tipo de ANN consisten en un vector de datos

### 5.3 REDES NEURONALES CONVOLUCIONALES (CONVOLUTIONAL NEURAL NETWORKS, CNN)

Se trata de un modelo de ANN posterior a las conocidas como MLP, de hecho, se trata de una variación de estas últimas que se caracterizan por su buen funcionamiento en su aplicación en datos con forma de matriz bidimensional como son las imágenes. El utilizar matrices o imágenes como datos de entrada se trataba de problemas con estructuras de datos demasiado complejas para poder aplicar en ellos un perceptrón multicapa, por lo que el desarrollo de nuevos modelos fue necesario.

En el caso de querer utilizar una imagen en una MLP, la solución consistiría en convertir toda la matriz de la imagen en un vector plano, lo que conllevaría elevar el coste computacional enormemente y, sobre todo, perder completamente la estructura espacial de la imagen [43].

Al resolver problemas de clasificación con estructura espacial como pueden ser las imágenes, se busca encontrar características diferenciales entre las entradas dadas, que permitan a la red juzgar si estas pertenecen a una u otra clase o categoría en particular. Todo ello implica que la estructura de las CNN sea más compleja que la propia de las MLP, aunque no por ello incomprensible.

En general, las redes convolucionales se caracterizan por tener imágenes como inputs, teniendo como salidas valores escalares o magnitudes vectoriales.

La estructura general de una red neuronal convolucional se describe a continuación como sigue [34], [43]:

1. **Input o datos de entrada:** Similar a lo que sucede en las MLP, es la primera capa en la que se reciben los datos de entrada, que en este caso van a consistir en imágenes bidimensionales cuyos datos (píxeles, que recogen la intensidad de las imágenes que se estudian) serán posteriormente tratados.
2. **Capas de convolución:** Es la parte siguiente de una red neuronal y realiza las operaciones de convolución que se verán más adelante. De forma resumida, estas operaciones extraen información local sobre las diferentes regiones de la imagen, como si se tratase de un filtro, es el principio fundamental de las CNN ya que este tipo de sistemas inteligentes buscan extraer rasgos o características en función de dónde se localicen en las imágenes, de ahí la importancia de considerar campos locales para obtener buenos resultados.
3. **Capas de podado o pooling:** con estas operaciones, las redes buscan extraer la información más característica de una determinada región de la imagen, al

mismo tiempo que se va reduciendo el tamaño de los datos de entrada, buscando quedarse únicamente con las características principales.

4. **Capa *flatten*:** Se trata de la última capa donde a partir de las características principales extraídas por las dos capas anteriores, se clasifican para obtener la salida final.

Llegado este punto hay que mencionar que las imágenes no son realmente matrices bidimensionales, a pesar de que se puedan utilizar en las CNN como datos de entrada. En un caso general, una imagen se considera una matriz de dimensión  $N_a \times N_l \times N_c$  siendo  $N_a$  el número de píxeles de altura,  $N_l$  el número de píxeles de ancho y  $N_c$  el número de canales. En imágenes en las que solamente haya un color (blanco y negro) habrá un único canal y sí que entonces sería una matriz bidimensional, pero en un caso general, el número de canales hace referencia al número de colores de la imagen (normalmente suele consistir en tres, RGB).

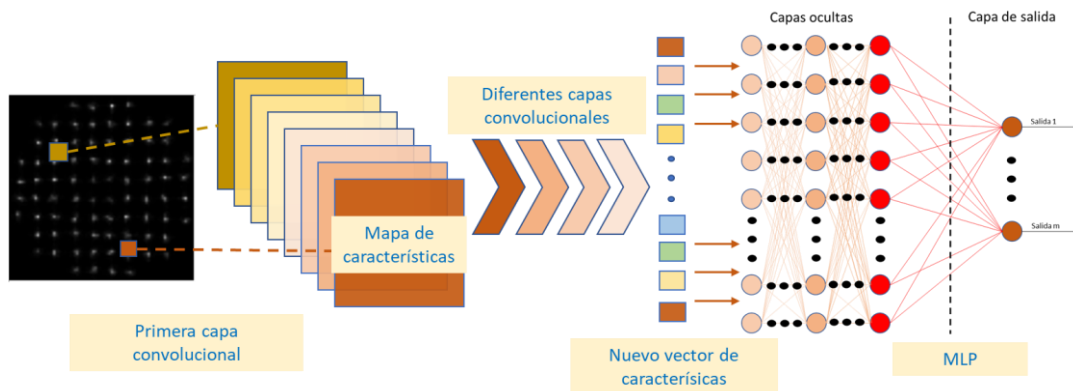


Figura 18: esquema de una red neuronal convolucional. En este caso como entrada de la red se muestra una imagen de las subaperturas de un SH en una observación nocturna. A partir de esta entrada se extraen datos a través de los mapas de características generados por las capas de convolución y a través de un proceso de podado, filtran información a un MLP o capa totalmente conectada, que se encarga de obtener los datos de salida que se buscan.

En la Figura 18 se muestra un esquema simplificado del esquema de una red neuronal convolucional. Las redes convolucionales, al igual que los perceptrones multicapa, tienen un proceso de propagación hacia delante y un algoritmo de retropropagación que consigue recalcular los pesos de la red para conseguir predicciones cada vez más exactas (a través de la minimización del error). Al tratarse de métodos muy similares al caso de las MLP no serán explicados con detalle nuevamente, mientras que sí se comentarán los procesos que llevan a cabo las capas convolucionales y las capas de podado.

La particularidad de las CNN frente a las MLP reside en el uso de capas convolucionales. Una capa convolucional se caracteriza respecto a otra por el número de *kernels* (núcleos) y por el *stride* (que significa el paso con el que se aplica cada uno de los núcleos). Los núcleos es lo que convoluciona sobre la matriz de entrada. Cada núcleo realiza convoluciones discretas por lo que la operación de convolución se ve resumida a realizar el sumatorio de los productos elemento a elemento entre la sección de la muestra sobre la que se aplica y el *kernel*.

Puesto que cada uno de los núcleos se aplica de forma local, se debe definir el tamaño de cada *kernel* cuando se realiza una red neuronal que, por normal general, deberá ser menor que el tamaño del tensor sobre el que se aplica. Al final cada uno de esos núcleos va a actuar como un filtro sobre regiones locales de la entrada, extrayendo las principales características. Una vez aplicado sobre una región local, el *kernel* se desplaza a lo largo de la muestra para volver

a aplicarse y repetir el proceso sobre toda la muestra, el tamaño del desplazamiento que se aplica entre la región donde se ha aplicado un *kernel* y la siguiente es lo que se denomina *stride*.

Otro parámetro interesante en el uso de una red neuronal convolucional es si se utiliza padeo (*padding*) o no. Los *kernels* se van aplicando a lo largo de todo el tensor sobre regiones locales y van avanzando con un tamaño de paso determinado. Si no se modifica la matriz, los píxeles situados en las esquinas de esta sufrirán un menor número de veces la aplicación de los filtros que aquellos situados en el centro, provocando quizás que las características situadas en estas regiones tuvieran menor relevancia que las de las regiones centrales. Este efecto es el que se pretende evitar con el padeo, que añade ceros en lugares externos de la matriz para que los *kernels* se apliquen en igual número de veces sobre todos los píxeles del tensor.

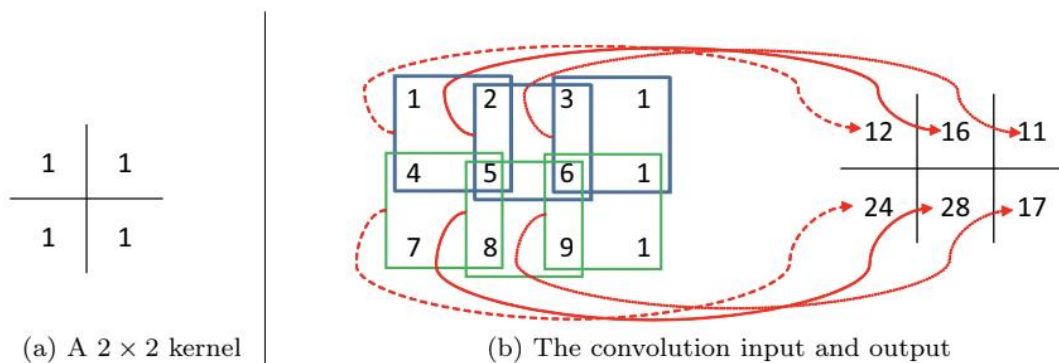


Figura 19: Ejemplo ilustrado de operación de convolución extraída [44]. En este caso se aplica un kernel de tamaño  $2 \times 2$  sobre una matriz de tamaño  $3 \times 4$ . Como se aprecia en este caso no se utiliza padeo y, como consecuencia, la matriz de salida tiene un tamaño menor que el tensor de entrada. Además, el núcleo no se aplica el mismo número de veces sobre todos los píxeles de la imagen, ya que sobre los dos centrales se llega a aplicar hasta en 3 ocasiones mientras que sobre los más esquinados solamente 1.

Cuando los *kernels* se aplican con paso 1, la salida de la capa convolucional resultará una nueva matriz con un mapa de características con dimensión igual a la matriz de entrada, siempre y cuando se aplique el padeo. En caso de no aplicarlo no se podrá mantener el tamaño de la muestra original, como se puede ver en la Figura 19. Se obtendrán tantos mapas de características como el número de núcleos aplicados [43], [44].

Un método frecuentemente aplicado para disminuir el tamaño de las salidas de cada capa convolucional es aumentar el tamaño de los *strides*. Así, con un tamaño de paso de 2 píxeles, el tamaño de las salidas se verá reducido a la mitad. Es importante reducir el tamaño de los mapas de características ya que al final de la red existe una capa MLP que necesita como entrada un vector y, cuanto mayor sea el tamaño del vector, mayor tiempo de computación necesitará.

Al igual que en el caso de las MLP es común aplicar una función de activación a la salida de cada uno de los mapas de características antes de la conexión con las capas siguientes. Las funciones de activación más utilizadas son las mismas que en el modelo anterior, ReLU, PReLU, tangente hiperbólica, etc.

La otra capa característica de una CNN es la capa de podado (*pooling*). La finalidad de estas capas es reducir la dimensionalidad espacial de los mapas de características, reduciendo el número de parámetros y el coste computacional de la red. Este proceso se lleva a cabo modificando agrupaciones de píxeles por el resultado de aplicar una operación sobre ellos.

Existen múltiples criterios en cuanto a la aplicación de este tipo de funciones (o capas) a pesar de que los más empleados son el de coger el valor máximo de esa región, llamado 'Max-Pooling' o el de coger el valor de la media aritmética, conocido como 'Average-Pooling'. Otras funciones de *pooling* populares son la norma rectangular de los píxeles vecinos o la media ponderada en base a la distancia al píxel central [43], [44]. En cualquier caso, el pooling sirve para hacer que la representación sea aproximadamente invariante bajo pequeñas traslaciones de la imagen de entrada en el sentido de que si se traslada la entrada una pequeña cantidad de píxeles, la mayoría de las salidas de la capa de pooling no cambian [44].

A continuación, se muestra un ejemplo completo del resultado de aplicar las capas convolucionales y las capas de podado sobre una matriz, por simplicidad, de un único canal:

1	1	0	1	0	1
0	1	0	0	1	1
0	1	1	1	1	0
1	0	1	1	0	1
1	0	1	1	0	0
1	1	0	0	0	1

1	0	1
0	1	0
0	0	1

Figura 20: Matriz de entrada de la capa convolucional situada en la izquierda de la imagen y filtro que se va a aplicar a la derecha.

Por simplicidad, en este caso tampoco se aplicará el padeo y el paso será 1, por lo que el filtro de tamaño  $3 \times 3$  se aplicará sobre cada submatriz  $3 \times 3$  del tensor original. El resultado en el mapa de características resultará de multiplicar elemento a elemento la zona local escogida con el filtro y sumando el resultado. El resultado de la posición (1,1) del mapa de características será:

$$1 \cdot 1 + 0 \cdot 1 + 0 \cdot 1 + 0 \cdot 0 + 1 \cdot 1 + 0 \cdot 0 + 0 \cdot 0 + 1 \cdot 0 + 1 \cdot 1 = 3$$

El resultado de aplicarlo sobre todo el tensor de entrada:

3	3	1	3
2	2	2	3
2	4	3	1
2	2	2	3

Figura 21: Resultado de la aplicación del filtro sobre el tensor mostrados en la **¡Error! No se encuentra el origen de la referencia..**

Si se hubiera aplicado padeo, la matriz hubiera sido modificada obteniendo un mapa de características de la misma dimensión que el tensor de origen, como se muestra en la figura B.

0	0	0	0	0	0	0	0
0	1	1	0	1	0	1	0
0	0	1	0	0	1	1	0
0	0	1	1	1	1	0	0
0	1	0	1	1	0	1	0
0	1	0	1	1	0	0	0
0	1	1	0	0	0	1	0
0	0	0	0	0	0	0	0

2	1	0	2	1	1
2	3	3	1	3	1
1	2	3	2	3	1
2	2	4	3	1	2
2	2	2	2	3	0
1	3	1	1	1	1

Figura 22: En la matriz de la izquierda se muestra el tensor de la **figura 20** resultado aumentar su dimensionalidad mediante padeo. A la derecha se observa el resultado del mapa de características, que en este caso sí que presenta la dimensionalidad del tensor inicial.

Si a continuación de la capa de convolución se aplicase una capa de podado sobre el mapa de características de la 56 ¡Error! No se encuentra el origen de la referencia. se reduciría su dimensionalidad, como se muestra en la ¡Error! No se encuentra el origen de la referencia..

Figura 23: Resultado de aplicar max-pooling (izquierda) o average-pooling (derecha) sobre el mapa de características mostrado en la figura 22. En este caso la dimensión del podado es  $2 \times 2$  por lo que la dimensionalidad del tensor fue reducido a la mitad.

Este conjunto de capas se podría anidar varias veces, reduciendo el tamaño de la imagen de entrada y aumentando el número total de mapas de características. Después de realizar todo el procesamiento convolucional, los mapas se conectan como entradas a la capa de entrada de la sección MLP de la topología de la red, para finalmente devolver una respuesta como salida. Se obtiene por tanto una salida numérica de la ANN a partir de una entrada en forma de tensor.

A día de hoy las CNN son el tipo de red neuronal con mayor número de aplicaciones ya que se utilizan en áreas muy diversas como el diseño de coches autónomos, diagnóstico por imagen en medicina, procesamiento de textos, reconocimientos de voz...

#### 5.4 REDES NEURONALES COMPLETAMENTE CONVOLUTIVAS (FULLY-CONVOLUTIONAL NEURAL NETWORKS, FCN)

Las redes completamente convolucionales pueden entenderse como un avance al desarrollo de las redes convolucionales. Poder utilizar matrices bidimensionales como entrada de las ANN permitió la aplicación de la IA a multitud de problemas para los que las MLP no ofrecían respuestas con la suficiente precisión. A pesar de ello, seguía existiendo aún una importante limitación, las CNN podían utilizarse como sistemas de regresión o clasificadores, pues la salida necesaria consistía en un número o un vector de longitud fija, pero no eran capaces de generar una respuesta de la misma dimensión que la entrada, es decir, una nueva matriz bidimensional o una imagen como salida.

No fue hasta el 2015 con la aparición de las FCN [45], [46] que se empezaron a obtener tensores como salidas de las ANN. El concepto de las FCN es sencillo una vez visto el funcionamiento de las CNN. Están formadas por un primer bloque convolucional que consiste en las mismas capas de convolución y podado que las de las redes CNN, pero al final, en lugar de tener un bloque completamente conectado, tienen un bloque de-convolucional. La finalidad de este nuevo bloque es, a partir de los mapas de características extraídos por el bloque convolucional, obtener un nuevo tensor de una dimensión dada como salida de la red.

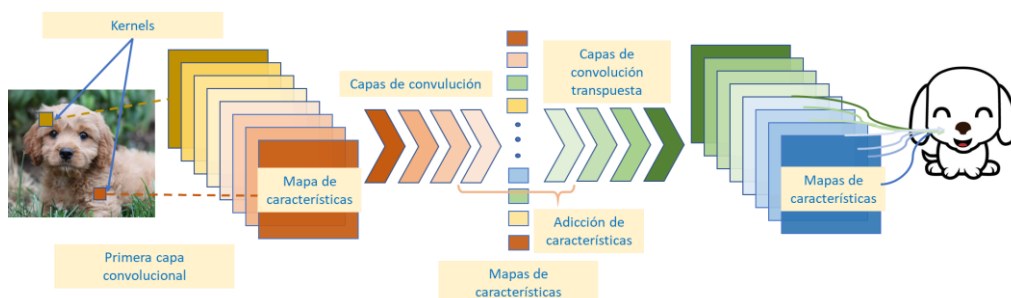


Figura 24: Esquema de una red completamente convolucional donde se pueden apreciar los dos bloques, el bloque convolucional y el deconvolucional que obtiene una nueva imagen como salida a partir de los mapas de características.

Las capas tradicionales del bloque deconvolucional son convolucionales transpuestas y capas de podado transpuestas. Su funcionamiento es el inverso a sus homólogos del bloque convolucional, desde un mapa de características pequeño van aumentando el tamaño del tensor recuperando la categoría y la zona local a la que pertenece cada píxel a partir de las características abstractas.

La forma en el que se aumenta el tamaño del tensor es la inversa que, en el caso del bloque convolucional, puede hacerse a través de las capas de podado transpuesto o modificando los *strides* de la capa convolucional transpuesta. Si en el caso de la CNN un tamaño de paso 2 reducía la dimensión del tensor a la mitad, en este caso lo aumentará al doble.

## 5.5 FRAMEWORKS DE REDES NEURONALES ARTIFICIALES

Existen diferentes frameworks para la implementación, modelización, entrenamiento y aplicación de redes neuronales artificiales. Con el desarrollo de estas tecnologías diversas compañías tecnológicas como Google, Amazon, Microsoft... desarrollan marcos en los que se simplifica la programación de este tipo de sistemas que permiten ejecuciones rápidas y, normalmente, la posibilidad de completarlos y ampliarlos mediante la importación de diferentes módulos externos a los que se puede acceder de forma sencilla.

En este trabajo en concreto se va a introducir *Tensorflow*, marco con soporte multi-GPU que se utilizó en la programación de las redes neuronales utilizadas a lo largo del trabajo. Desarrollado por Google en lenguaje Python, es uno de los *frameworks* más sencillos a la hora de trabajar con redes neuronales artificiales.

En combinación con Keras [47] consiste en uno de los métodos más utilizados en la actualidad. Keras consiste en una librería de código abierto que simplifica la creación de modelos de inteligencia artificial. Con la utilización de esta librería, el usuario debe indicar la topología de la red, así como los parámetros principales del modelo (*learning rate*, *batch size*, funciones de activación de las diferentes capas...) y la propia librería en combinación con Tensorflow crea y enlaza las neuronas, sin necesidad de código por parte del programador. De esta forma se consiguen códigos fácilmente interpretables y modificables incluso por usuarios distintos al creador de este, permitiendo así también de forma sencilla la colaboración entre diferentes usuarios.



## 6 OBJETIVOS

---

El objetivo principal de la tesis es la elaboración de un sistema reconstructor para óptica adaptativa solar basado en redes neuronales artificiales. Para ello, se pretende utilizar como punto de partida los buenos resultados obtenidos previamente en la aplicación de técnicas de inteligencia artificial como sistemas reconstructores para óptica adaptativa nocturna. Como ejemplo del buen funcionamiento de estos sistemas se podría mencionar el reconstructor CARMEN desarrollado por el grupo MOMA (MOdelización Matemática Aplicada) de la Universidad de Oviedo tanto en entornos simulados como en datos de telescopio [48], [49].

Este objetivo principal permite una gran cantidad de líneas de investigación diferentes, teniendo en cuenta tanto las técnicas de inteligencia artificial que se podrían testear como los diferentes sistemas de óptica adaptativa o variaciones en sus configuraciones sobre las que estas técnicas podrían ser aplicadas. Por todo ello, para el desarrollo de la tesis ha sido necesario ir marcando objetivos secundarios menos ambiciosos, de manera que se pueda ir comprobando la viabilidad de las técnicas para el problema a resolver, así como cuáles de estas técnicas resultan más efectivas o bajo qué circunstancias o configuraciones se obtienen los mejores resultados.

El primer objetivo que se marcó entonces fue realizar una comparación entre diferentes modelos de ANN para determinar cuál de ellos podría resultar más ventajoso a la hora de realizar las reconstrucciones para óptica adaptativa solar, y continuar las investigaciones. Este estudio se realizó a partir de datos simulados de observaciones nocturnas.

Una vez escogida la red a desarrollar para realizar las investigaciones, el segundo objetivo fue determinar qué datos de entrada de los medidos por los sensores SH resultan más favorables para las reconstrucciones con redes neuronales. Los resultados de esta investigación son los que dieron lugar a la segunda publicación de la tesis.

Por último, determinado ya el tipo de red a desarrollar y los datos de entrada a utilizar, se realizó una prueba para determinar si había alguna configuración de SH determinada, en cuanto a número de subaperturas y número de píxeles por subapertura, que favoreciera la reconstrucción con redes neuronales artificiales. Además, también se estudió la posibilidad de implementar redes neuronales con sensores de frente de onda más simples, favoreciendo así los intereses económicos frente a los sensores que se utilizan en la actualidad. De este tercer objetivo surgen los resultados mostrados en el último artículo publicado.

Todos estos objetivos son, en realidad, el camino a seguir para el desarrollo del objetivo principal, conseguir reconstructores con redes neuronales artificiales para óptica adaptativa solar y estudiar su viabilidad. De la aplicación de las redes para cada uno de los casos surgirán limitaciones que pueden ser entendidas como futuras líneas de investigación ampliatorias al trabajo realizado.



## 7 DISCUSIÓN DE LOS RESULTADOS PUBLICADOS

### 7.1 COMPROBACIÓN Y ELECCIÓN DE DIFERENTES MODELOS DE ANN PARA RECONSTRUCCIONES DE ÓPTICA ADAPTATIVA NOCTURNA.

En este primer estudio se comprobó el funcionamiento de diferentes modelos de redes neuronales artificiales bajo el mismo problema de óptica adaptativa. Con ello se pretende concluir qué modelos de ANN se ajustan mejor al problema de la OA para su posterior desarrollo en observaciones solares, cumpliendo así el primero de los objetivos en la realización de la tesis. A lo largo de este subcapítulo se discutirán los resultados correspondientes a la publicación: “*Overview and Choice of Artificial Intelligence Approaches for Night-Time Adaptive Optics Reconstruction*” [50]

En este caso concreto se realizaron reconstrucciones utilizando redes MLP, CNN y FCN. Tanto para el entrenamiento como validación de las redes fue necesario generar diferentes tipos de datos de entrada, puesto que las MLP solamente admiten entradas en forma vectorial mientras que tanto MLP y FCN admiten imágenes. Lo mismo ocurre con las salidas, siendo en forma vectorial las de la MLP o CNN y en formato imagen las de la FCN. Por ello los sets de datos se generaron guardando las imágenes de la luz recibida por el SH, las pendientes de cada subapertura medidas por el SH, los primeros 153 coeficientes de Zernike de la fase turbulenta, así como también la imagen de la fase turbulenta. Estos cuatro datos se guardaron para cada una de las simulaciones realizadas.

Tipo de red	MLP	CNN	FCN
Entrada	Pendientes SH	Imagen luz recibida SH	Imagen luz recibida SH
Salida	Coef. Zernike	Coef. Zernike	Imagen fase turbulenta

Tabla 1: Datos de entrada y de salida de cada una de las redes neuronales utilizadas.

Al tratarse del primer estudio se decidió realizar la prueba con observaciones nocturnas ya que, debido a los estudios previos realizados por el grupo, era más sencillo obtener datos simulados fiables de óptica adaptativa nocturna que de observaciones solares. Además, las salidas que debían obtener las ANN eran las mismas que en el caso diurno, la reconstrucción de la fase turbulenta, por lo que la prueba serviría para posteriores estudios en OA solar.

Para el estudio se generaron 5 set de datos diferentes. El mayor de todos ellos fue el de entrenamiento, que consistió en 800000 simulaciones diferentes, cada una de ellas con los 4 tipos de datos diferentes. La turbulencia atmosférica simulada consistió en una única capa a diferentes alturas, desde 0m hasta 20km en pasos de 500m. El valor del coeficiente  $r_0$  también se varió desde 7 hasta 17cm en pasos de 0.5 cm. Para cada una de estas combinaciones altura y  $r_0$  se generaron 100 simulaciones diferentes. Se generó de igual forma un set de datos para validación durante el proceso de entrenamiento que consistió en 4000 simulaciones generadas de igual forma que en el caso anterior.

De igual forma se generaron 3 set de datos para testear el funcionamiento de las redes, sobre los que medir la calidad de las reconstrucciones en situaciones turbulentas diferentes a las del entrenamiento. Cada uno consistía en 4000 iteraciones teniendo todos ellos el valor del parámetro  $r_0$  fijado, comprobando también así si la intensidad de la turbulencia condicionaba el funcionamiento de los diferentes modelos de ANN. Cada uno de los conjuntos de datos tenían un valor de  $r_0$  de 8, 10 y 12 cm respectivamente, que se identifican con turbulencia muy alta, turbulencia alta y turbulencia baja (situación de observación normal), respectivamente. En

cuanto a la altura de la capa turbulenta en los conjuntos de datos de test, se varió entre 0 y 20 km al igual que en los casos de entrenamiento, en pasos en este caso de 500m.

Como se ha mencionado, todos los experimentos fueron realizados a partir de datos obtenidos de simulaciones. Tanto las simulaciones como el procesamiento de los datos se realizaron en un ordenador provisto de Ubuntu LTS 14.04.3 con una CPU Intel Xeon E5-1650 v3 @ 3.50 GHz, una memoria DDR4 de 128 GB, tarjeta gráfica Nvidia GeForce GTX TitanX y disco duro SSD. El lenguaje de programación utilizado fue Python, siendo Keras el framework de redes neuronales escogido. Para la obtención de los conjuntos de datos, se utilizó el simulador DASP mencionado en la introducción, en su modalidad de óptica adaptativa nocturna.

Tanto en este primer estudio como en todos los que siguieron a continuación, la calidad de las reconstrucciones realizadas se midió en términos de error cuadrático medio de frente de onda, habitualmente conocido como RMSE WFE por sus siglas en inglés. El valor del RMSE WFE de una fase turbulenta se expresa en términos de unidades de longitud de frente de onda y se puede calcular como sigue:

$$\text{RMSE WFE} = \frac{2\pi}{\lambda} * \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N}} \quad (106)$$

Donde  $x_i$  se corresponde con cada píxel de la fase turbulenta,  $\bar{x}$  el valor medio de todos los píxeles,  $N$  el número total de píxeles y  $\lambda$  la longitud de onda de la radiación del frente de onda incidente, que en el caso concreto de este estudio  $\lambda = 640nm$ .

Para determinar la semejanza de una reconstrucción con la original conocida esta última, se recurre al RMSE WFE residual, en ese caso:

$$\text{residual WFE} = \frac{2\pi}{\lambda} * \sqrt{\frac{\sum_{i=1}^N (x_i - y_i)^2}{N}} \quad (107)$$

En este caso  $x_i$  son los pixels de la fase original e  $y_i$  de la reconstruida. En el caso de una reconstrucción perfecta el valor del error residual debería ser 0 y a medida que su valor aumente más diferirán las fases turbulentas entre ellas.

Tanto en el caso de la MLP como de la CNN hubo que realizar un tratamiento posterior de las salidas con el fin de determinar el valor del residual WFE, puesto que las salidas de ambos consistían en un vector con los primeros 153 coeficientes de Zernike. Utilizando una de las funciones del módulo *AOTools* [51] se puede, a partir de los coeficientes, obtener una imagen de la turbulencia atmosférica representada por esos valores. Como es de esperar, cuanto mayor sea el número de coeficientes, mayor precisión en la reconstrucción. Esas nuevas imágenes son las que se compararon con las originales para la obtención de los valores de residual RMSE WFE.

Llegado a este punto podría darse el caso de considerar que la comparación entre los diferentes modelos no sería justa, ya que en el caso de la FCN su salida se está comparando con una fase mucho más precisa que en el caso de los otros dos modelos de redes (que se comparan con fases reconstruidas a partir de coeficientes de Zernike). Por ello se añadió al estudio un cuarto caso a comparar, en el que se entrenaba una FCN con las salidas de la fase aproximada. Se utilizó el mismo conjunto de datos que para los otros casos, considerando la fase reconstruida a partir de los coeficientes con el módulo *AOTools* como salida idónea de la red.

### 7.1.1 Resultados

Los resultados obtenidos de los 4 ensayos se resumen en la tabla 2. La topología de las redes neuronales finalmente elegida para cada uno de los ensayos resultó ser la que menor error obtenía tras la realización de diversas pruebas, las cuatro topologías se pueden ver en [50].

Modelo de red	MLP		CNN		FCN		FCN (aproximación de fase)	
$r_0$ (cm)	WFE residual (nm)	WFE residual relativo	WFE residual (nm)	WFE residual relativo	WFE residual (nm)	WFE residual relativo	WFE residual (nm)	WFE residual relativo
8	100.84	28	86.58	24	253.63	19	146.68	37
10	90.65	30	104.91	35	191.50	18	177.23	44
12	84.54	33	112.05	44	160.94	17	203.72	52

Tabla 2: Resumen de los resultados obtenidos por los 4 modelos de redes neuronales probados a lo largo del estudio. Las topologías de cada una de las redes escogidas así como los parámetros de entrenamiento de cada una de ellas se encuentran publicados en [50].

Además de comprobar la calidad de las reconstrucciones, también se analizó el tiempo computacional necesario para cada uno de los modelos, pues en OA tan importante es obtener una buena reconstrucción como realizarlas lo más próximo a tiempo real. Todas las redes fueron testadas en el mismo equipo y bajo las mismas condiciones, permitiendo así realizar una comparación justa. El valor mostrado en la tabla 3 se calcula como el valor medio sobre las 4000 muestras de test.

$r_0$ (cm)	Tiempo de computación por muestra (ns)			
	MLP	CNN	FCN	FCN (aprox. fase)
8	230	575	675	675
10	238	600	700	670
12	243	575	678	675

Tabla 3: Comparación de los tiempos medios por muestra empleados por cada uno de los modelos para realizar la reconstrucción.

En las figuras de la 25 a la 28 se muestran ejemplos de las reconstrucciones alcanzadas por las redes neuronales, mostrándose siempre la fase turbulenta original (o la reconstruida a partir de los coeficientes originales, según el caso) y a la derecha la reconstrucción realizada por la ANN. Los ejemplos mostrados son los mismos que se adjuntaron al artículo publicado [50], siendo todos ellos escogidos de forma aleatoria sin priorizar la calidad de la reconstrucción.

Comprobar las reconstrucciones de forma visual, a parte del error en términos de WFE, es necesario. En situaciones en las que las ANN puedan haber sobreentrenado puede darse el caso que presente una fase media como salida sobre todas las fases con las que fueron entrenadas, el WFE residual tenga valores aceptables, pero cuando se realice una comprobación visual las fases originales y las predichas no tengan mucha semejanza.

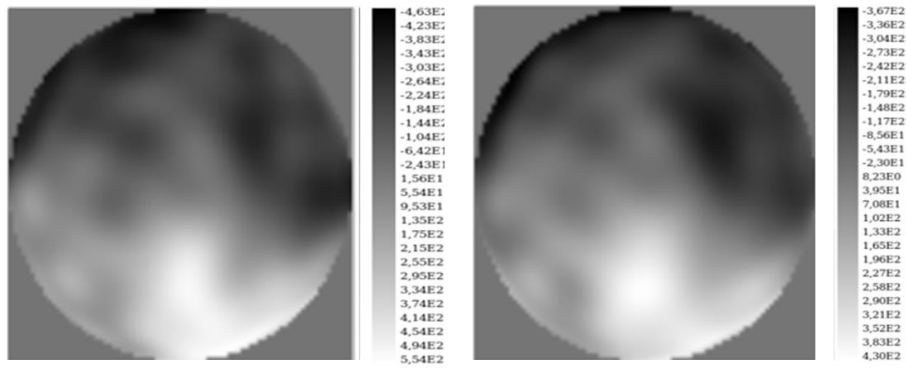


Figura 25: Ejemplo de la reconstrucción realizada por la red MLP en una muestra del test con valor  $r_0 = 12$  cm. La imagen de la izquierda se construye a partir de los coeficientes de Zernike originales y la de la derecha a partir de los predichos por la red. Sin tener en cuenta los valores de error en términos numéricos, a simple vista los perfiles de turbulencia atmosférica presentan grandes similitudes en cuanto a su forma, apareciendo las zonas de mayor y menos desfase en las mismas regiones de la imagen.

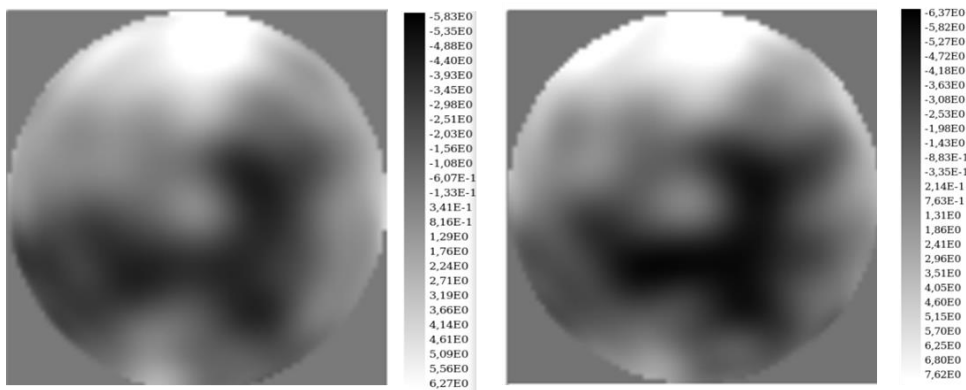


Figura 26: Ejemplo de reconstrucción realizada por la red CNN en una muestra del test con valor de  $r_0 = 8$  cm, se trata del test con mayor intensidad de turbulencia. La imagen de la izquierda se construye a partir de los coeficientes de Zernike originales y la de la derecha a partir de los predichos por la red. Al igual que en el caso anterior, a simple vista ambas fases presentan una gran similitud.

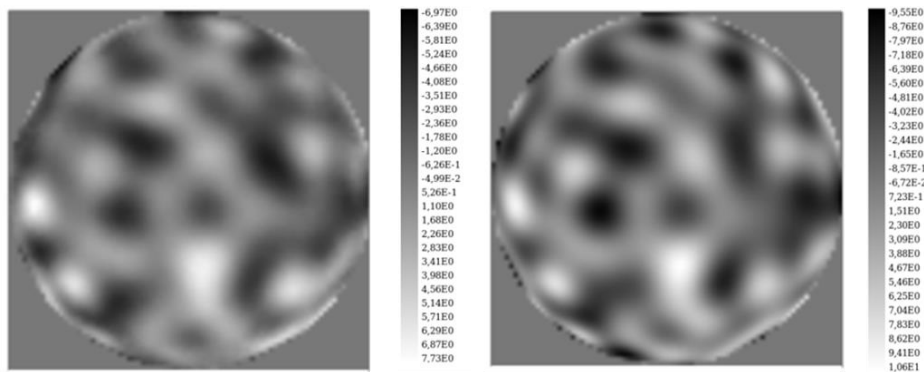


Figura 27: Ejemplo de reconstrucción realizada por la red FCN de la aproximación de la fase en una muestra del test con valor de  $r_0 = 10$  cm. La imagen de la izquierda se construye a partir de los coeficientes de Zernike originales y la de la derecha a partir de los predichos por la red. En este caso como cabría esperar a la vista de los resultados, es en el que ambas fases menos similitudes presentan. Pese a ello, la predicha sigue guardando bastante semejanza con la original.

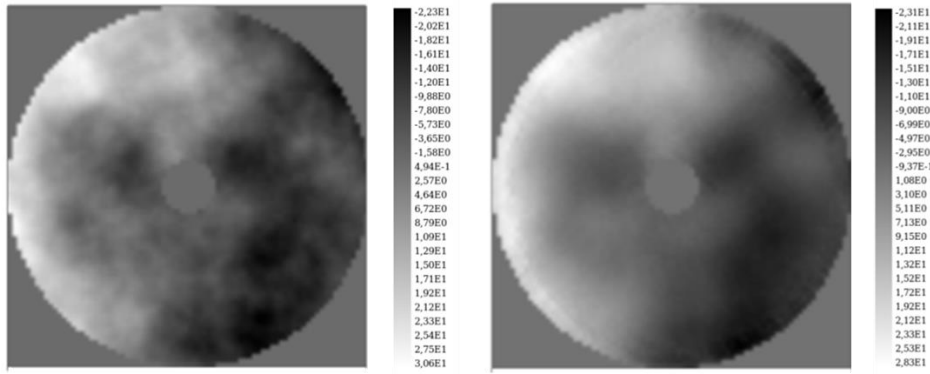


Figura 28: Por último, se muestra un ejemplo de reconstrucción realizada por la red FCN de la fase original en una muestra del test con valor de  $r_0 = 8$  cm. La imagen de la izquierda es la fase original y la de la derecha la predicha. A diferencia del resto de situaciones se ve muy claro como la imagen situada a la izquierda tiene un nivel de detalle mucho más alto que todas las demás, es debido a que se trata de la fase original y no de una aproximación como en los casos anteriores. Pese a ello la red FCN aproxima de forma correcta la fase original, tanto en una comparación visual como a la vista de los valores de WFE residual obtenidos.

### 7.1.2 Discusión de resultados

Empezando por el tiempo de computación necesario, no cabe duda de que la red que ha mostrado mayor rendimiento ha sido la MLP, siendo más del doble de rápida que su más inmediato perseguidor, la CNN. Se trata de un resultado esperable, quizás no con tanta diferencia pero hay que recordar que la MLP es el modelo que se enfrenta al problema más sencillo, puesto que de una entrada con forma de vector (con las pendientes ya medidas por el SH), aporta una salida que consiste en 153 valores. Todas las demás reciben la imagen de la luz recibida por el SH y, a partir de ella, tienen que interpretar las pendientes, las aberraciones, etc. Además, que la cantidad de información recibida en cada iteración es mucho mayor, en concreto 160 valores en el caso de la MLP por  $80 \times 80 = 6400$  datos de cada uno de los píxeles en el caso del resto. Teniendo en cuenta entonces estos parámetros, el ahorro de tiempo computacional de la MLP no parece ya tan significativo, aunque sigue siendo importante un ahorro de más del doble.

El otro término por analizar se corresponde con la calidad de las reconstrucciones realizadas y el error cometido por cada uno de los modelos estudiados. Los resultados mostrados en la tabla 2 deben ser analizados teniendo en cuenta que no es lo mismo reconstruir toda la información de la turbulencia que una aproximación de la misma basada en los primeros 153 coeficientes de Zernike.

A la vista de los resultados, cuando se trata de la aproximación de la fase, la MLP resulta ser el modelo que obtiene los mejores resultados. Los errores cometidos para todos los valores de  $r_0$  son los más bajos, no superando el error relativo en ningún caso el 33%, siendo del 44% en la CNN y del 52% en el caso de la FCN. Y lo más importante de todo, se trata del modelo más consistente ya que independientemente de la turbulencia, sus valores de WFE residual relativo se mantienen (aumenta un 5% en el caso de menor turbulencia). Solo hay un caso con  $r_0 = 8$  cm en el que la CNN mejora a la MLP en un 4% (24% vs 28%). Este resultado puede ser esperable ya que se produce en la situación de mayor intensidad de turbulencia. Es de esperar que la CNN que tiene como entrada mucha más información, en situaciones extremas sea capaz de caracterizar mejor el problema. Por otro lado, la FCN demostró ser el peor modelo de red posible cuando se pretenden reconstruir aproximaciones de la fase turbulenta, pues solamente logró obtener un 37% de error en el mejor de los casos (el más turbulento). Además, es la que

demonstró ser también la menos consistente ya que con la disminución de la intensidad de turbulencia el error aumentó a valores superiores al 50%.

En cambio, del estudio se puede extraer un resultado muy positivo en relación con las FCN cuando se utilizan para reconstruir la fase completa original (no una aproximación). Bajo esta situación las FCN obtuvieron los menores valores de error de todo el estudio, de tan solo 19% en el peor de los casos. Además, demostraron la mayor consistencia con la variación de intensidad de turbulencia de todos los modelos (solamente un 2% de variación). Otro aspecto que se podría destacar es que obtuvieron mejores resultados en condiciones normales de observación, en lugar de situaciones de máxima turbulencia como los otros caos. Sin embargo, este último resultado no es muy relevante ya que, dada la no variación con la intensidad, es prácticamente el mismo valor que en el resto de las situaciones.

Por todo ello, a raíz de los resultados obtenidos en este primer estudio se decidió continuar con el desarrollo del reconstructor para OA solar, priorizando las pruebas con redes del tipo FCN a la vista de los buenos resultados obtenidos. Dada la gran variación que existe entre dos situaciones de OA diferentes (cambios de configuración, de características de los sensores, etc.) no se puede descartar definitivamente el uso de redes MLP o CNN, pero si es cierto que, a partir de aquí, fueron utilizadas en situaciones donde las FCN no respondían de la forma esperada.

## 7.2 RECONSTRUCCIÓN DE TURBULENCIA ATMOSFÉRICA UTILIZANDO REDES COMPLETAMENTE CONVOLUCIONALES FCN. COMPARACIÓN DE LAS RECONSTRUCCIONES EN FUNCIÓN DE LOS PARÁMETROS DE ENTRADA ESCOGIDOS.

Se trata del segundo estudio publicado, que surge como continuación del anterior. Una vez escogido qué tipo de red ANN se va a utilizar, comprobar su funcionamiento como reconstructor en observaciones solares y estudiar qué tipo de entrada de los valores medidos por el SH optimiza el funcionamiento de la red, cumpliendo así el segundo objetivo en la realización de la tesis. A lo largo de este subcapítulo se mostrarán y discutirán de forma resumida los resultados correspondientes a la publicación: *“Fully Convolutional Approaches for Numerical Approximation of Turbulent Phases in Solar Adaptive Optics”* [52].

Una vez escogidas en el estudio anterior las FCN como los modelos de IA adecuados para la realización de reconstrucciones, se pretendía confirmar durante este estudio que se trataba de las redes adecuadas para reconstrucciones de OA Solar.

En segundo lugar, en la introducción se mencionó que los sensores SH en OA Solar realizan un preprocesamiento de los datos recibidos, calculando correlaciones cruzadas entre las diferentes subaperturas [53]. A lo largo del estudio se determina si esas correlaciones cruzadas pueden ser una buena entrada de información para la red neuronal (al igual que lo es buena para los reconstructores clásicos, como el método de mínimos cuadrados) o si, por el contrario, la imagen de la luz recibida por el SH sin preprocesar optimiza el funcionamiento de los modelos de IA.

Los sets de datos simulados consistieron en un primer conjunto de datos de entrenamiento, formado por 80000 muestras en las que se guardaban la imagen de la luz en el SH, la imagen de las correlaciones cruzadas calculadas por el SH (una imagen que guarda muchas similitudes con la luz que recibe un SH en OA nocturna) y la fase turbulenta original que en



ambos casos es la salida que se pretende obtener por la red. Para la generación de las 80000 muestras se programaron dos capas turbulentas, una situada siempre a  $0m$  de altitud con el 10% de todo el peso de la turbulencia y, una segunda capa, con altitud variante desde  $1m$  hasta  $20000m$  en pasos de  $100m$ . El valor de  $r_0$  también se varió desde  $8cm$  hasta  $16cm$  en pasos de  $1cm$ . Para cada una de las combinaciones posibles se generaron 50 muestras (consecutivas en el tiempo). Para la validación realizada durante el proceso de entramiento se generó un nuevo conjunto de datos más pequeño, que consistió en 3000 muestras donde tanto la altitud de la segunda capa como el parámetro  $r_0$  varió dentro de los mismos límites, pero con tamaños de paso mayores.

Para la comprobación de los resultados obtenidos por las redes neuronales en este caso se generaron un mayor número de conjuntos de datos, en concreto 12. A parte de comprobar cómo afectaba la intensidad de la turbulencia a los resultados se analizó también si la altitud de esa segunda capa turbulenta influía en ellos o no. Para ello, se realizaron conjuntos de 600 muestras con valores de  $r_0$ : 7, 9, 11, 13, 15, 17cm y con alturas de la segunda capa (la primera siempre a  $0m$ ) entre dos rangos: entre 0 y  $3000m$  simulando las capas turbulentas de baja altitud o *Ground-Layer* y entre  $9000$  y  $12000m$  simulando las capas turbulentas altas o *High-altitude Layer*.

Con el objetivo de realizar una comparación justa, para todos los casos se utilizó la misma topología de red neuronal FCN, detallada en [52]. Como se trató de uno de los primeros estudios con simulaciones de observaciones solares, se utilizó una de las configuraciones más sencillas, la configuración SCAO [13]. Tanto las simulaciones como el entrenamiento de las redes neuronales se realizaron en un sistema provisto de Ubuntu LTS 14.04.3 con una CPU Intel Xeon E5-1650 v3 @ 3.50 GHz, una memoria DDR4 de 128 GB, tarjeta gráfica Nvidia GeForce GTX TitanX y disco duro SSD. El lenguaje de programación utilizado fue Python, siendo Keras el framework de redes neuronales escogido. Para la obtención de los conjuntos de datos, se utilizó el simulador DASP mencionado en la introducción, en configuración SCAO Solar.

### 7.2.1 Resultados

El análisis de las reconstrucciones se realizó nuevamente utilizando el WFE residual al igual que en el experimento anterior. Los resultados obtenidos se muestran en la tabla 4, realizando una distinción entre los valores para baja capa turbulenta y alta capa turbulenta. No obstante, también es necesario realizar comprobaciones visuales entre las fases obtenidas y las originales, por lo que las figuras 29 y 30 contienen algunas muestras a modo de ejemplo.

Capa turbulenta	$r_0$ (cm)	WFE residual (rad)		WFE res. relativo (%)		Tiempo (ms)	
		SHWFS	CC	SHWFS	CC	SHWFS	CC
Capas turbulentas bajas (0m – 3000m)	7	2.55	2.85	32	36	4.53	4.33
	9	1.72	1.99	30	35	4.45	4.38
	11	1.51	1.78	28	33	4.48	4.33
	13	1.28	1.50	27	32	4.47	4.35
	15	1.11	1.31	29	34	4.41	4.38
	17	1.05	1.20	29	34	4.50	4.37
Capas turbulentas altas (9000m – 12000m)	7	1.69	1.78	18	19	4.50	4.37
	9	1.25	1.30	15	16	4.52	4.33
	11	0.95	1.02	14	15	4.47	4.33
	13	0.80	0.86	16	17	4.53	4.32
	15	0.72	0.79	15	17	4.41	4.32
	17	0.70	0.74	15	16	4.57	4.35

Tabla 4: Resultados obtenidos por la red FCN en función de la intensidad de la turbulencia por el parámetro  $r_0$  y la altitud de la principal capa turbulenta. El término SHWFS hace referencia a que la entrada de la red consistió en la imagen de la luz recibida por el sensor SH mientras que CC hace referencia a utilizar como entrada la imagen de las correlaciones cruzadas calculadas por el SH. Finalmente, las dos últimas columnas indican el tiempo medio empleado por la FCN para la reconstrucción de una muestra de cada conjunto de datos.

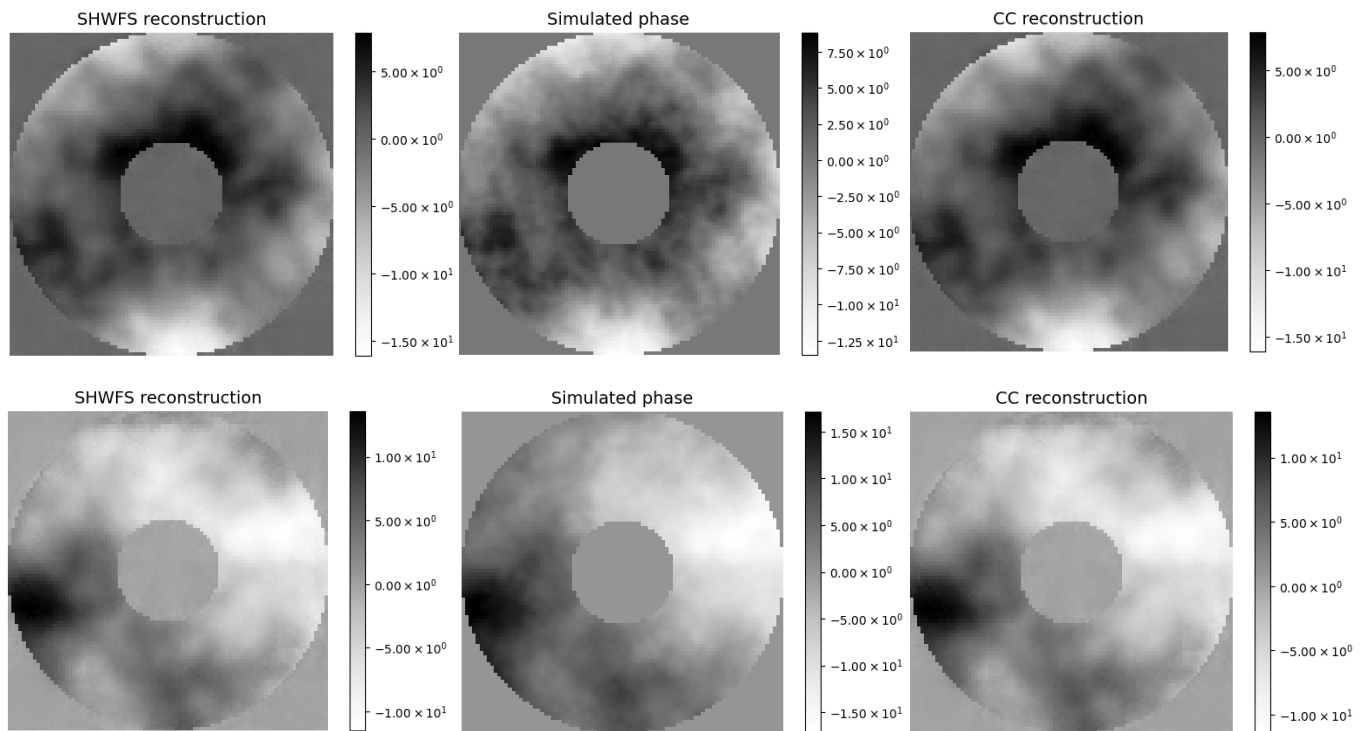


Figura 29: Comparación de fases reconstruidas para situaciones de capa turbulenta baja con  $r_0$  7cm y 9 cm para cada una de las filas respectivamente. En la columna del medio se muestra siempre la fase original simulada, estando en la columna de la izquierda la reconstruida a través de las imágenes del SH y la de la derecha a través de las imágenes de las correlaciones.

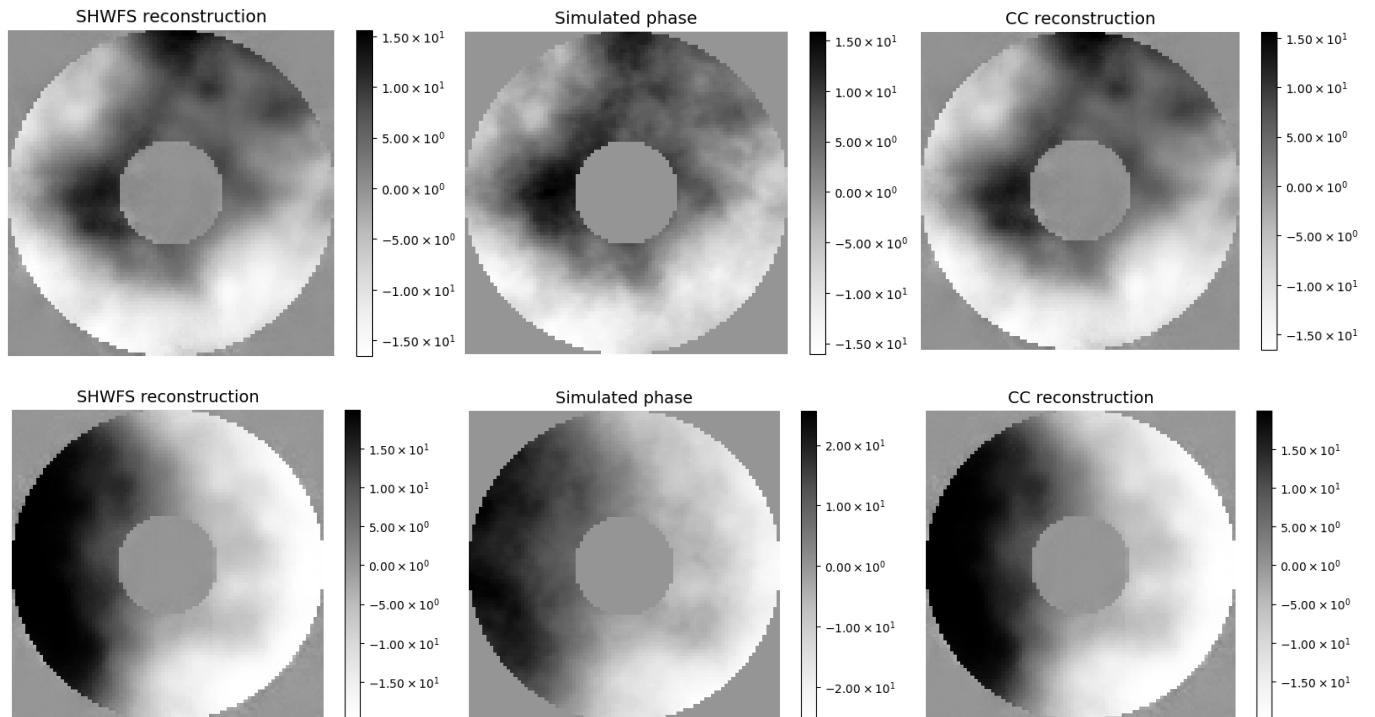


Figura 30: Comparación de fases reconstruidas para situaciones de alta capa turbulenta con  $r_0$  7cm y 9 cm para cada una de las filas respectivamente. En la columna del medio se muestra siempre la fase original simulada, estando en la columna de la izquierda la reconstruida a través de las imágenes del SH y la de la derecha a través de las imágenes de las correlaciones.

### 7.2.2 Discusión de resultados

A la vista de los resultados obtenidos se comprobó que las redes FCN mantenían los valores de error obtenidos en las reconstrucciones de óptica nocturna realizadas en el anterior estudio cuando se enfrentaban a problemas de óptica solar.

No obstante, era necesario comprobar si las reconstrucciones fueran suficientemente precisas en comparación con los métodos de reconstrucción más comunes empleados en óptica adaptativa solar. Antes de realizar este estudio tan detallado en función de valores de  $r_0$ , de altitudes, etc., se comparó con ayuda de unos sets de datos más sencillos el error de reconstrucción que aportaban las redes FCN en un problema de SCAO Solar frente al error que se obtendría realizando esa misma reconstrucción con el método LS. Los resultados fueron presentados en el artículo de congreso [54], pero los valores de la reconstrucción con LS (mostrados en la tabla 5) pueden volver a ser utilizados aquí para mostrar las ventajas de utilizar las FCN.

$r_0$ (cm)	8	10	12
WFE residual LS (rad)	2.60	2.09	1.91

Tabla 5: Valores de WFE residual en reconstrucciones de SCAO Solar aplicando el método de mínimos cuadrados (LS) extraídos de [54].

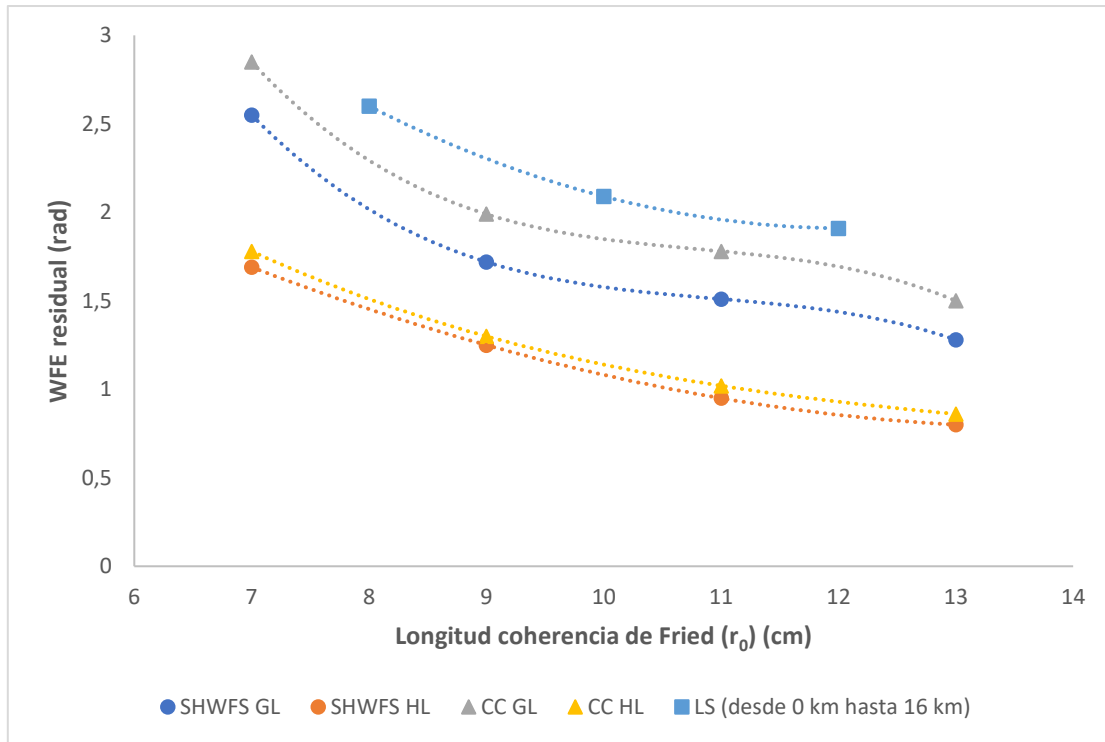


Figura 31: Comparación del error WFE residual obtenido por las FCN frente al obtenido por el método de mínimos cuadrados (LS). Las siglas SHWFS y CC hacen referencia a la entrada utilizada en la FCN, siendo la imagen de la luz en las subaperturas del SH o las correlaciones calculadas por el SH respectivamente. Los términos GL y HL hacen referencia a "Ground-Layer" y "High-altitude Layer", es decir, capas turbulentas cerca de la superficie (de 0 a 3000m) o capas turbulentas altas (de 9000 a 12000m) respectivamente.

Como las simulaciones no se realizaron bajo las mismas condiciones: diferentes valores del parámetro  $r_0$  y sin distinción por la altura de la turbulencia; fue necesario realizar una aproximación que se puede ver en la figura 31 con ajuste polinómico de orden 3 para poder comparar los datos. Se puede ver como los valores obtenidos por las redes FCN bajo diferentes condiciones son inferiores a los obtenidos por el método tradicional. De esta forma, se cumpliría el primer objetivo de esta investigación, mostrar que las redes neuronales son una buena alternativa a los métodos tradicionales, obteniendo reconstrucciones más precisas.

De los resultados obtenidos, analizando por ejemplo los valores para  $r_0 = 8$  cm, para el caso de capas altas turbulentas las FCN obtienen un 44.3% menos WFE residual que el método LS en el caso de utilizar como entrada las imágenes de SH o un 42.6% en el caso de las correlaciones. Para turbulencias cercanas a la superficie terrestre este valor disminuye hasta un 21.8% menos WFE residual con imágenes de SH como entrada o un 11.9% en el caso de las correlaciones. Es importante recordar que los datos de reconstrucciones con LS están calculados para alturas entre 0 y 16 km, por lo que la mejoría esperable con FCN sería un valor intermedio entre las dos situaciones expuestas.

En cuanto al segundo objetivo, comprobar qué tipo de entrada favorecía las reconstrucciones, se observa también claramente en la figura 31 que, bajo cualquier circunstancia de capa turbulenta, la reconstrucción utilizando como entrada las imágenes de SH mejora a la obtenida con las correlaciones, viendo los resultados de la tabla 4: desde un 5% hasta situaciones donde se mejora un 15% el WFE residual.

Entre los resultados mostrados en [52] también se encuentra la figura 32 en la que se hace una comparación de la similitud de las fases reconstruidas con las fases originales en función de la situación estudiada.

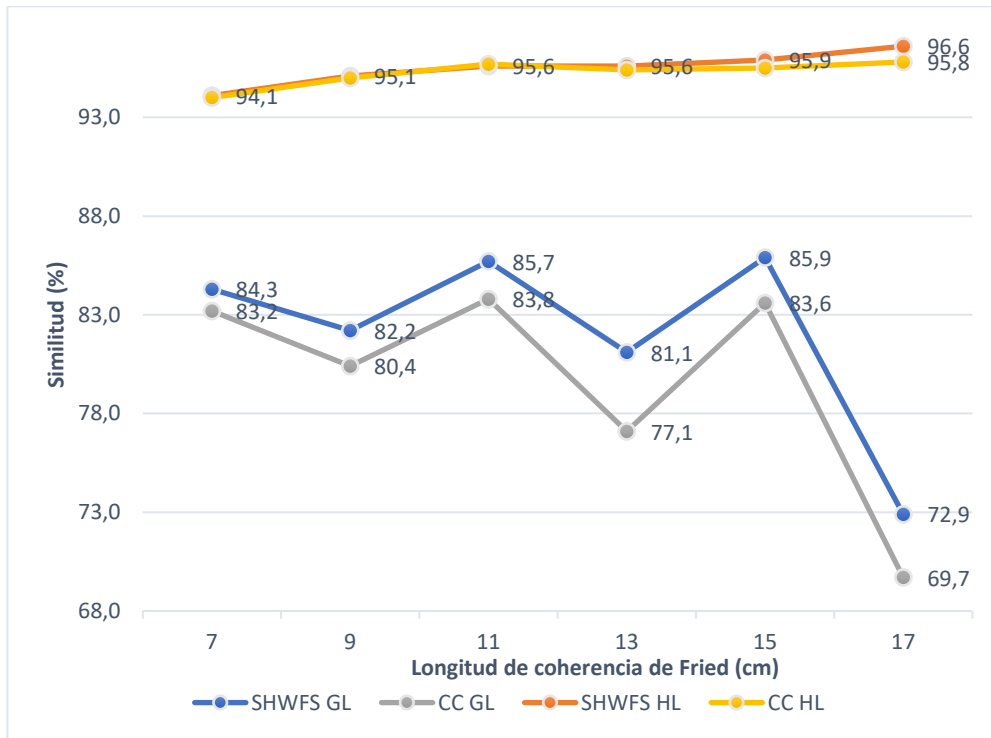


Figura 32: Comparación de la similitud obtenida bajo las diferentes configuraciones de alta (HL) o baja (GL) turbulencia y utilizando como entrada las imágenes del SH (SHWFS) o de las correlaciones (CC).

Las Figura 29 y Figura 30 confirman lo que se observa en las gráficas y las tablas. En ellas, las imágenes que se encuentran en la columna de la izquierda presentan mayor grado de detalle en comparación con la fase original mientras que, las de la derecha, reconstruidas a partir de la imagen de correlaciones, pese a seguir siendo similares a la fase original, las formas son más difuminadas.

Todos estos resultados concuerdan con lo que cabría esperar en un principio cuando se diseñó el experimento. Es de esperar que cuanto más información tenga la entrada que se le aporta a la FCN, mejor y más precisa sea la reconstrucción que se lleve a cabo. Las correlaciones son imágenes que surgen del tratamiento que realiza el SH a la información recibida, que la simplifica para darle un aspecto similar a los puntos de luz que se registraban en observaciones nocturnas. Obviamente, la imagen completa de la sección de la superficie solar que se esté observando va a contener más información de las aberraciones sufridas que si se reduce a un punto de luz. Sin embargo, sigue siendo necesario realizar esta comprobación. Pese a lo dicho anteriormente, las imágenes de la superficie solar contienen, aparte de información de las aberraciones sufridas, mucha otra información propia del Sol y variaciones debidas a otros fenómenos que pueden generar confusión o hacer que los filtros de las FCN extraigan características de otros aspectos en lugar de las debidas a la turbulencia atmosférica. Gracias a esta prueba se puede confirmar que las ANN son capaces de discernir entre todos esos datos y extraer los interesantes para el problema de la OA.

Por último, es necesario también realizar un comentario a propósito de los tiempos de computación necesarios expuesto en la tabla 4. Se aprecia una pequeña disminución del tiempo

cuando se trabaja con imágenes de las correlaciones como entrada, pero no se trata de algo relevante ya que en ningún caso esta diferencia es superior al 5% del tiempo total. Además, aquí habría que tener en cuenta el tiempo necesitado por el SH para el cálculo de las correlaciones que no sería necesario en el primer caso. En una aproximación para las imágenes de SH, el tiempo necesario de computación es alrededor de los 4.5ms, prácticamente el doble de lo esperado para un sistema de OA, donde suelen considerarse tiempos óptimos por debajo de los 2ms [55] para la implementación real. A este último dato hay que sumarle otros tiempos necesarios en métodos como el preprocesamiento de datos, traducción de imágenes de SH a imágenes de correlación, traducir salidas (generalmente coeficientes de Zernike) a fase, etc. En el caso de utilizar las imágenes del SH como entrada, la mayoría de esos procesos van incluidos en los 4.5ms que necesita la red para obtener la fase real por lo que, pese a ese alto valor, el tiempo total de toda la reconstrucción disminuiría el que necesitan los métodos actuales. Este último aspecto sería otra característica a favor de escoger las imágenes de SH como entradas óptimas para las FCN.

### 7.3 ESTUDIO DE DIFERENTES CONFIGURACIONES DE SENSORES SHACK-HARTMANN PARA RECONSTRUCCIONES DE TURBULENCIA ATMOSFÉRICA EN OA SOLAR.

En este subcapítulo se muestran y discuten los resultados del tercer y último estudio publicado durante el desarrollo de la tesis. Una vez conocidos el modelo de red a utilizar y las entradas más apropiadas para su funcionamiento, el tercer objetivo que se pretendió fue determinar las configuraciones de los sensores Shack-Hartmann y de las observaciones más apropiadas para el trabajo con redes FCN, pretendiendo sentar las bases para el desarrollo de sensores de frente de onda futuros diseñados para trabajar en combinación con sistemas de IA. A lo largo de este subcapítulo se discutirán los resultados más destacados correspondientes a la publicación: *“Comparative study of Shack-Hartmann configurations for atmospheric turbulence reconstructions in solar adaptive optics”* [56].

En los sistemas de OA solares actuales, los sensores SH se caracterizan por tener un gran número de subaperturas, en concreto, alrededor de  $15 \times 15$  subaperturas, donde en cada una de ellas se focaliza la imagen recibida por el telescopio. Como se ha visto con anterioridad, de cada una de las subaperturas se va a extraer el valor de la pendiente, a partir del cálculo de las correlaciones cruzadas. Por lo tanto, a mayor número de subaperturas, mayor número de valores de la pendiente del frente de onda en las diferentes zonas de la pupila y, por lo tanto, mayor precisión de la reconstrucción de la turbulencia atmosférica realizada por los métodos reconstructores tradicionales. En contraposición, el número de subaperturas de los sensores SH está limitado, en primer lugar, por el tamaño de la pupila y, en segundo lugar, por el aspecto económico. Aumentar el número de subaperturas encarece enormemente los sensores, pues se precisa desarrollar un mallado de lentes más preciso, con más lentes de menor tamaño para el mismo diámetro de pupila, al igual que el espacio que existirá para cada una de las subaperturas en el propio sensor será menor.

En el estudio mencionado anteriormente se comprobó como las FCN eran capaces de extraer más información de la turbulencia de cada una de las subaperturas del SH que en el caso de utilizar correlaciones. Por lo que una de las líneas de esta investigación consistirá en comprobar si el número de subaperturas afecta de igual forma utilizando reconstructores basados en IA que en los casos con algoritmos matemáticos como el método de LS y, en caso negativo, determinar qué configuración podría ser más beneficiosa.

Siendo el anterior punto mencionado el más importante del estudio, se analizaron dos factores más. En primer lugar, los sensores SH tienen sensores CCD en cada una de las subaperturas que registran la información, los reconstructores tradicionales no necesitan un gran número de píxeles por subapertura para el cálculo de las correlaciones, pero, en el caso de las FCN, como extraen más datos de la luz de cada subapertura, se quiere comprobar si tener un mayor número de píxeles podría implicar aumentar también la precisión al reconstruir. En segundo lugar, siendo un sistema SCAO Solar el escogido para la realización del estudio, se estudió si el campo de visión (FOV por sus siglas en inglés) con el que se realice la observación afecta a calidad de la reconstrucción realizada con las FCN.

Para este estudio fue necesario realizar nuevos conjuntos de datos de simulación, en los que se guardaban tanto las imágenes recibidas por los SH como la imagen de la fase de la turbulencia atmosférica que se pretendía reconstruir. Para cada una de las pruebas (influencia del número de subaperturas, del número de píxeles por subapertura y del FOV) se generó un conjunto de datos de entrenamiento, uno de validación y tres para realizar los tests. Las configuraciones que fueron probadas a lo largo del experimento se muestran en la tabla 6. Todos ellos en configuración SCAO Solar con el simulador DASP [32], con una apertura de 1.5m. Los conjuntos de entrenamiento consistieron en 80000 imágenes para cada una de las pruebas. Se simularon dos capas turbulentas, una siempre a 0m de altitud con el 10% del peso de toda la turbulencia y la otra restante variando desde 0m hasta 20000m en pasos de 200m. Se concedió el 90% de peso a la segunda capa con el fin de entrenar las redes a turbulencias en diferentes alturas. El valor del parámetro  $r_0$  fue variado desde 8cm hasta 16cm en pasos de 1cm. Para cada una de las situaciones se generaron 100 muestras. Al igual que en estudios anteriores, para testear el funcionamiento de las redes, se generaron tres conjuntos de datos de test para cada prueba con el valor del parámetro  $r_0$  fijado, con valores de 8, 10 y 12 cm. Permitiendo de esta forma comprobar el comportamiento en función de la intensidad de la turbulencia.

Debido a la gran diferencia de tamaño existente entre todos los datos de entrada que se probaron en el estudio, fue necesario utilizar cuatro modelos diferentes de red FCN. De todas formas, se pretendió que la topología de los modelos fuese lo más similar posible entre todas ellas. Las diferentes configuraciones de redes utilizadas se describen en [56].

Subaperturas del SH	15 × 15	10 × 10	7 × 7	3 × 3		
Píxeles por subapertura	20 × 20	28 × 28	42 × 42	24 × 24	48 × 48	96 × 96
Número total de píxeles del SH	300 × 300	280 × 280	294 × 294	72 × 72	144 × 144	288 × 288
Número de píxeles de la imagen de fase	75 × 75	70 × 70	77 × 77	18 × 19	36 × 36	69 × 69
Modelo de red neuronal utilizado	Modelo 1		Modelo 2		Modelo 3	Modelo 4

Tabla 6: Descripción de todas las pruebas realizadas. Los bordes del mismo color se corresponden con las simulaciones realizadas para cada una de las pruebas, indicándose en azul los conjuntos de datos utilizados para comparar el número de subaperturas y en naranja los utilizados para comparar cómo afecta el número de píxeles. Cada una de las combinaciones se realizó en dos ocasiones, con cada uno de los parámetros del FOV escogidos.

Al igual que en los experimentos anteriores, tanto las simulaciones como el entrenamiento de las redes neuronales se realizó en un ordenador provisto de Ubuntu LTS 14.04.3 con una CPU Intel Xeon E5-1650 v3 @ 3.50 GHz, una memoria DDR4 de 128 GB, tarjeta gráfica Nvidia GeForce GTX TitanX y disco duro SSD. El lenguaje de programación utilizado fue Python, siendo Keras el framework de redes neuronales escogido.

### 7.3.1 Resultados

Como se realizaron tres pruebas diferentes, en este caso se dividirán los resultados en tres apartados diferentes, permitiendo analizar cada una de las pruebas por separado.

#### 7.3.1.1 Influencia del número de subaperturas

Como se puede ver en la tabla 6, se probaron cuatro configuraciones de SH diferentes, partiendo de las 15 × 15 subaperturas como la configuración con mayor número y que se corresponde con el SH del telescopio Gregor [57], hasta las 3 × 3 subaperturas, lo que implica reducir 5 veces el número de subaperturas. Un resumen de los resultados obtenidos con los valores más importantes se puede ver en la tabla 7.

En la figura 33 se muestran un ejemplo de las reconstrucciones obtenidas, permitiendo así también realizar una comparación visual entre las fases reconstruidas y las originales.

$r_0$ (cm)	Número de subaperturas	WFE Residual (nm)			WFE Residual Relativo (%)		
		8	10	12	8	10	12
15 × 15	15 × 15	304.79	235.56	206.91	49	46	47
	10 × 10	249.88	206.11	180.65	47	47	48
	7 × 7	204.52	155.18	101.07	31	28	21
	3 × 3	202.93	146.43	93.9	30	26	19

Tabla 7: Resultados obtenidos de error de frente de onda en función del número de subaperturas del SH.

Para este estudio al igual que en los anteriores, me midió el tiempo de computación medio necesario para cada caso. El resultado obtenido fue prácticamente el mismo en todos ellos, siendo siempre inferior a los 4.6 ms, siendo el menor con 7 × 7 subaperturas y  $r_0$  10 cm, que se obtuvo un tiempo de 4.43 ms. El más largo fue el necesitado por el caso de 10 × 10 subaperturas con  $r_0$  10 cm, que fue 4.54 ms por muestra.



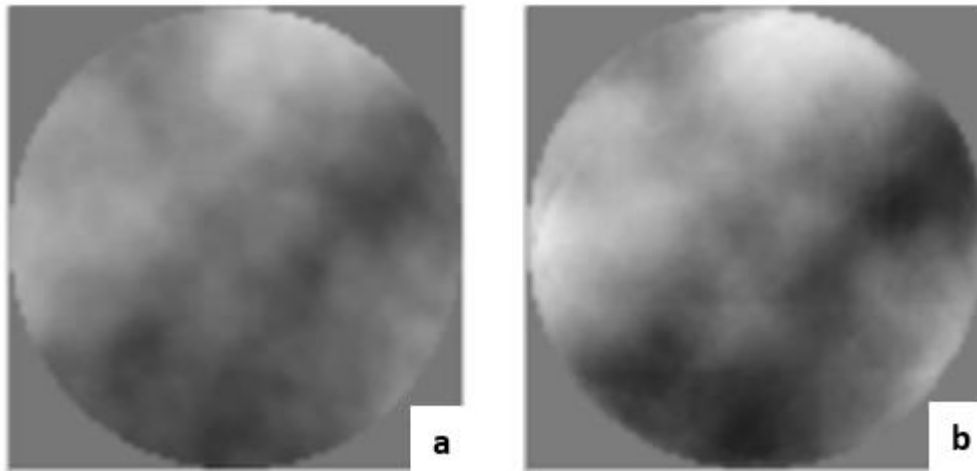


Figura 33: Ejemplo de la reconstrucción realizada por la FCN para un SH de  $7 \times 7$  subaperturas para una situación de  $r_0 = 8$  cm. La imagen a representa la fase original siendo la b la reconstruida con métodos de IA.

### 7.3.1.2 Influencia del número de píxeles por subapertura

Para este estudio se fijó el número de subaperturas en una configuración  $3 \times 3$ , escogida gracias a los resultados del apartado anterior, modificándose el número de píxeles por subapertura. Para el caso de mayor número, se utilizaron los conjuntos de datos simulados para la prueba anterior y, dado que se trataba de un número ya de píxeles por subapertura muy alto, se realizaron dos pruebas más con valores inferiores, en concreto  $48 \times 48$  y  $24 \times 24$ . Dado la gran diferencia de tamaño entre las imágenes de entrada que se le aportaban a la red (ver tabla 6), fue necesario utilizar para cada uno de los casos un modelo de red neuronal diferente. No obstante, los modelos se diseñaron tratando de mantener el mayor número de similitudes entre ellos.

Un resumen de los resultados obtenidos de esta comparación se muestra en la tabla 8. También se analizó el tiempo computacional empleado por los 3 modelos de redes neuronales. El modelo 3 de FCN obtuvo un tiempo medio de  $4.41$  ms,  $4.39$  ms y  $4.43$  ms para valores de  $r_0$  de 8, 10 y 12 cm respectivamente. En el caso del modelo 4, aplicado al caso de  $24 \times 24$  píxeles por subapertura, se obtuvieron tiempos medios de  $4.18$ ,  $4.22$  y  $4.23$  ms respectivamente.

$r_0$ (cm)		WFE Residual (nm)			WFE Residual Relativo (%)		
		8	10	12	8	10	12
Píxeles por subapertura	$112 \times 112$	202.93	146.43	93.9	30	26	19
	$48 \times 48$	179.85	126.53	105.05	28	24	23
	$24 \times 24$	213.27	124.94	101.07	35	25	23

Tabla 8: Resultados obtenidos de error de frente de onda obtenidos para configuraciones de SH de  $3 \times 3$  subaperturas con diferente número de píxeles por subapertura.

### 7.3.1.3 Influencia del campo de visión (FOV) de la observación

Todos los resultados mostrados anteriormente se realizaron con un FOV de 6 arcosegundos. Para este apartado lo que se hizo fue repetir todas las pruebas de los dos apartados anteriores, pero con 10 arcosegundos de campo de visión. Los resultados se muestran de forma resumida en la Tabla 9. Con la obtención de todas estas nuevas reconstrucciones se pretendió, a parte de la comparación del FOV, confirmar los resultados hallados en las pruebas anteriores.

$r_0$ (cm)	Píxeles / subapertura	WFE Residual (nm)			WFE Residual Relativo (%)		
		8	10	12	8	10	12
$15 \times 15$	$20 \times 20$	217.25	167.12	147.22	37	34	35
$10 \times 10$	$28 \times 28$	257.04	190.20	165.53	39	34	35
$7 \times 7$	$42 \times 42$	222.82	175.08	150.41	35	33	33
$3 \times 3$	$96 \times 96$	154.38	112.21	94.70	27	23	23
	$48 \times 48$	164.73	118.57	98.68	26	22	22
	$24 \times 24$	139.26	77.99	59.68	21	14	13

Tabla 9: Resultados obtenidos de error de frente de onda para diferentes configuraciones de SH, todas ellas realizadas sobre simulaciones con un campo de visión de 10 arcosegundos.

### 7.3.2 Discusión de resultados

En primer lugar, antes de analizar los resultados es conveniente mencionar los valores  $r_0$  sobre los que se están realizando las pruebas, ya que el máximo valor se corresponde con 12 cm, un valor donde la intensidad de la turbulencia atmosférica sigue siendo significativa. Por tanto, todos los valores representan en mayor o menor medida turbulencias atmosféricas intensas, lo que implica que cabe esperar que los errores de frente de onda obtenidos sean altos.

A continuación, se discutirán los resultados siguiendo el orden en que se realizaron las pruebas, empezando con la comparación en función del número de subaperturas y terminando en función del tamaño del campo de observación.

De la Tabla 7 se puede ver que el error de frente de onda obtenido decrece cuando disminuye el número de subaperturas del SH, especialmente hasta el caso de  $7 \times 7$  subaperturas, a partir de ahí se mantiene aproximadamente constante. La mayor diferencia se aprecia en situaciones de turbulencia más normal, con un  $r_0$  de 12 cm, para ese caso el error relativo disminuye desde un 47% hasta un 21%, o incluso un 19% para el caso de  $3 \times 3$  subaperturas. Esto conlleva un importante avance, especialmente en términos económicos, ya que con un SH de solamente  $3 \times 3$  subaperturas, las FCN son capaces de obtener mejores resultados que con los costosos SH que se diseñan actualmente para los sistemas de OA.

La explicación que se le puede dar a este resultado es que cuando se utilizan redes FCN, toda la información de la luz recibida por el SH se utiliza en la reconstrucción, en lugar de dos valores por subapertura como sucede con los métodos tradicionales. Esto implica que las características extraídas por la red no se vean limitadas por el número de subaperturas. Además, en la imagen que se introduce a la red FCN con toda la luz recogida por el SH, existen cambios abruptos en la imagen cuando se pasa de una subapertura a otra (en los bordes de cada una de ellas) como se ve en la figura 13 por ejemplo. Esos bordes generan error en la red cuando extrae las características de esas zonas, y puede ser la explicación de que, con menor número de subaperturas y, por tanto, de borde, la red logre obtener las mejores reconstrucciones.

Por otro lado, el caso de  $3 \times 3$  subaperturas tiene menos bordes que el de  $7 \times 7$ . En un principio, según la explicación dada anteriormente, cabría esperar errores menores en el caso de  $3 \times 3$ . Sin embargo, quizás sean demasiado escasas para que la FCN tenga suficiente información de la fase turbulenta en todas sus regiones, ya que tiene pocas subaperturas para comparar entre ellas las aberraciones recibidas en cada una.

Una última tendencia que se puede observar de la tabla 7 es que el error residual relativo disminuye con la intensidad de la turbulencia.

En cuanto a la segunda prueba, se observa que el número de píxeles por subapertura influye positivamente la reconstrucción de la FCN, en situaciones de turbulencia normal. Se trata de un resultado que cabría esperar, como la FCN obtiene la información de toda la luz recibida por cada subapertura, cuanto más detallada sea esa información (mayor número de sensores CCD), mejor será la reconstrucción obtenida. Sin embargo, pese a que no se aprecia una gran diferencia en el error residual relativo, cuando se analiza la similitud entre las reconstrucciones (ver el artículo completo [56]) sí que se ve que aumenta ligeramente, desde un 84% en el caso de  $24 \times 24$  píxeles a un 91% en el de mayor número de estos.

Por último, se comparó como influye el FOV de la observación y se comprobó que aumentar el campo de visión influye positivamente sobre las reconstrucciones realizadas por las redes, especialmente a mayor número de subaperturas. Hay que mencionar que por las limitaciones del simulador DASP, se tuvo que modificar ligeramente el número de píxeles por subapertura, disminuyéndolo. Aumentando el FOV se obtuvieron mejores reconstrucciones incluso con menor número de píxeles por subapertura. En términos de similitud, se pasó, por ejemplo, con  $15 \times 15$  subaperturas y  $8 \text{ cm}$  de  $r_0$  de 72% de similitud a 88%. Lo mismo sucede para el resto de caso, los errores y la similitud se mantienen, o incluso mejoran, disminuyendo ligeramente el número de píxeles por subapertura (lo que implica a priori una peor condición para las FCN), especialmente para bajos valores de  $r_0$ . Para turbulencias de menor intensidad, apenas se notaron diferencias reseñables al variar el campo de visión.



## 8 DISCUSIÓN DE LOS RESULTADOS DE CONGRESOS INTERNACIONALES

---

### 8.1 RECONSTRUCTOR CONVOLUCIONAL CARMEN PARA ÓPTICA ADAPTATIVA NOCTURNA.

En este subcapítulo se muestran y discuten los resultados del primer artículo de congreso realizado. Se trató de las primeras investigaciones realizadas a lo largo de las tesis relacionadas con la reconstrucción utilizando redes neuronales. En este caso, se comprobó el funcionamiento de las redes CNN aplicadas a problemas de OA nocturna, antes de haberse realizado la primera publicación mencionada anteriormente. A lo largo de este subcapítulo se discutirán los resultados más destacados correspondientes al artículo de congreso: “*Convolutional CARMEN: tomographic reconstruction for night observation*” [58].

Para la prueba se utilizó la configuración MOAO, que se caracteriza por utilizar una serie de estrellas guía no situadas necesariamente en el eje de observación, permitiendo recoger información de un campo de visión mayor que en el caso de la configuración SCAO. El sistema reconstructor deberá ser capaz de recoger la información de todas esas diferentes direcciones para estimar una corrección de la turbulencia en el eje de observación.

A lo largo de este estudio se comprobó el funcionamiento tanto de una red CNN, de una red MLP, así como el método tradicional de reconstrucción por mínimos cuadrados LS. Es importante mencionar que, como se trató del primer experimento realizado, consistió en una prueba de concepto para ver si las redes neuronales convolucionales podrían llegar a ser un buen método de reconstrucción, antes de realizar la prueba del primer artículo publicado. Lo importante no eran tanto los resultados obtenidos, como ver si realmente se podía tratar de un camino a seguir para próximas investigaciones.

Para la realización de simulaciones como el entrenamiento de las redes neuronales se realizó en un ordenador provisto de Ubuntu LTS 14.04.3 con una CPU Intel Xeon E5-1650 v3 @ 3.50 GHz, una memoria DDR4 de 128 GB, tarjeta gráfica Nvidia GeForce GTX TitanX y disco duro SSD. El lenguaje de programación utilizado fue Python, siendo Tensorflow el framework de redes neuronales escogido. Las simulaciones al igual que en los casos anteriores se realizaron con la plataforma DASP [31].

Para este experimento, se diseñó una configuración MOAO con 3 SH. En el caso de la red MLP las entradas consistían en los centroides de la luz medidos por los SH, para el caso de la CNN las entradas eran las imágenes completas de los 3 SH. Las salidas deseadas fueron los centroides objetivo, es decir, los centroides que se corresponden con la turbulencia situada en el eje de observación, ya que los medidos por los otros SH se corresponden con las direcciones donde se encontraban las estrellas guía.

Se utilizaron 186000 muestras para la realización del test de entrenamiento, con capas turbulentas variando desde  $0m$  hasta  $15500m$  de altitud en pasos de  $100m$ . En cuanto al parámetro  $r_0$ ; desde  $8cm$  hasta  $20cm$ . Para probar el funcionamiento de las redes, se generaron seis conjuntos de datos, tres de ellos consistentes en 4 capas cada uno con diferentes propiedades (ver tabla 10) y otros tres con turbulencias a alturas determinadas, en concreto  $5000m$ ,  $10000m$  y  $15000m$ .

Parámetro	Atmósfera			Unidades
Nombre del test	atm1	atm2	atm3	
$r_0$	0.16	0.12	0.085	m
<b>Capa 1</b>				
Altitud	0	0	0	m
Peso relativo	0.65	0.45	0.80	
Velocidad viento	7.5	7.5	10	m/s
Dirección viento	0	0	0	Grados
<b>Capa 2</b>				
Altitud	4000	2500	6500	m
Peso relativo	0.15	0.15	0.05	
Velocidad viento	12.5	12.5	15	m/s
Dirección viento	330	330	330	Grados
<b>Capa 3</b>				
Altitud	10000	4000	10000	m
Peso relativo	0.10	0.30	0.10	
Velocidad viento	15	15	17.5	m/s
Dirección viento	135	135	135	Grados
<b>Capa 4</b>				
Altitud	0	0	0	m
Peso relativo	0.65	0.45	0.80	
Velocidad viento	7.5	7.5	10	m/s
Dirección viento	0	0	0	Grados

*Tabla 10: Características de las capas turbulentas simuladas para el bloque de 3 conjuntos de datos de test. Se caracterizan por tener cada uno de ellos 4 capas turbulentas en lugar de tener una única a una altura fijada como el otro bloque de conjunto de datos utilizados.*

### 8.1.1 Resultados

En la Tabla 11 se muestran los resultados obtenidos. A diferencia de los estudios que se realizaron posteriormente, en este caso no se muestra el error de frente de onda residual ya que no se está reconstruyendo la fase. Se mide el WFE del frente de onda que se obtendría con los centroides objetivo. En una situación de reconstrucción ideal, su valor debería ser 0, ya que se obtendría un frente de onda plano. De manera que, cuanto menor sea el valor del WFE, mejor será la reconstrucción.

Test	Reconstructor	Ratio de Strehl	WFE (nm)
Atmósfera 1	LS	0.0444	630
	CNN	0.0321	545
	MLP	0.0416	544
	Sin corregir	0.0099	295
Atmósfera 2	LS	0.1988	340
	CNN	0.2766	301
	MLP	0.2783	300
	Sin corregir	0.0245	814
Atmósfera 3	LS	0.3008	295
	CNN	0.3891	259
	MLP	0.3908	257
	Sin corregir	0.0388	657
5000 m	LS	0.1993	338
	CNN	0.3637	267
	MLP	0.3686	268
	Sin corregir	0.0225	949
10000 m	LS	0.0647	480
	CNN	0.1731	361
	MLP	0.1739	361
	Sin corregir	0.0306	789
15000 m	LS	0.0572	506
	CNN	0.0690	471
	MLP	0.0653	474
	Sin corregir	0.0323	7268

Tabla 11: Resultados obtenidos por los diferentes reconstructores en términos de ratio de Strehl y WFE. También se muestran los valores que se obtendrían con el mismo frente de onda sin corregir.

### 8.1.2 Discusión de resultados

El objetivo fundamental de este estudio inicial era verificar si la IA podía ser utilizada como sistema reconstructor para OA, por lo que se planteó un problema inicial más sencillo que los publicados posteriormente.

Analizando los valores de WFE obtenidos por cada modelo de inteligencia artificial, se observa que en ambos casos son muy similares. En esta primera prueba sí que sorprende que, en la mayoría de las ocasiones, la MLP es la red que obtiene los mejores resultados en comparación con la CNN, no obstante, se trata de mejoras inferiores al 2%, por lo que tampoco se consideran demasiado relevantes.

Los resultados mostrados en la Tabla 11 resultaron claramente esperanzadores, ya que tanto la CNN como la MLP lograron mejorar en todos los casos las reconstrucciones realizadas mediante el método de LS. Como era de esperar, cualquiera de los métodos obtuvo errores menores de frente de onda que si no se hubiera aplicado ninguno de ellos. Tanto la topología de los modelos de redes utilizados como los resultados más en detalle se encuentran en el artículo de congreso [59].

Estos resultados fueron, por tanto, el punto de partida de la tesis para continuar desarrollando modelos de redes neuronales, buscando cada vez minimizar el error. También, con el avance de la investigación, se consiguieron resolver problemas más complejos como reconstruir la imagen de la fase turbulenta, no sólo los centroides objetivo.

## 8.2 RECONSTRUCCIÓN DE FASE TURBULENTO MEDIANTE REDES COMPLETAMENTE CONVOLUCIONALES EN OA SOLAR

En este subcapítulo se muestran y discuten los resultados del segundo artículo de congreso realizado. En él se realizaron pruebas aplicando redes neuronales completamente convolucionales como reconstructor de OA Solar sobre un sistema SCAO. Los resultados y datos del estudio completos se encuentran en el artículo de congreso: “*Early Fully-convolutional approach to wavefront imaging on solar adaptive optics simulations*” [60].

Este estudio consistió en la publicación de las primeras pruebas que se realizaron con redes FCN en óptica adaptativa solar. Fruto del primer artículo publicado, donde se había determinado las FCN como modelo de red más prometedora para ser utilizada como reconstructor en OA, se realizaron las primeras pruebas en solar utilizando para ello una de las configuraciones más simples, SCAO Solar. La continuación de la investigación resultó en el segundo artículo publicado comentado anteriormente.

Para este estudio se entrenaron las redes FCN con las imágenes de la información recibida por un SH de  $15 \times 15$  subaperturas y  $24 \times 24$  píxeles CCD en cada una de las subaperturas. Similar a las condiciones de sistema de OA del telescopio solar Gregor [57]. Tanto las simulaciones como el entrenamiento de las redes se realizaron en un ordenador provisto de Ubuntu LTS 14.04.3 con una CPU Intel Xeon E5-1650 v3 @ 3.50 GHz, una memoria DDR4 de 128 GB, tarjeta gráfica Nvidia GeForce GTX TitanX y disco duro SSD. El lenguaje de programación utilizado fue Python, siendo Tensorflow el framework de redes neuronales escogido.

Para comprobar las reconstrucciones realizadas, se elaboraron tres conjuntos de datos diferenciados por la intensidad de la turbulencia atmosférica, con valores de  $r_0$ : 8, 10 y 12 cm. Cada conjunto de datos consistió en 400 muestras con turbulencias a diferentes alturas. Las entradas de la red consistieron en la luz recibida por el SH directamente, siendo la salida la imagen de la fase de la turbulencia atmosférica. También se simuló las reconstrucciones que resultarían de aplicar el método de mínimos cuadrados, con el fin de determinar si mejoraban con las FCN o no. Los datos de estas simulaciones son los que luego también fueron utilizados en [52].

### 8.2.1 Resultados

Los resultados obtenidos por ambos métodos, tanto LS como redes FCN, se muestran en la tabla 12. En las figuras 34 y 35 se muestran algunos ejemplos de las fases originales y reconstruidas con el objetivo de realizar una comparación visual.

$r_0$ (cm)	WFE residual FCN (nm)	WFE Residual LS (nm)	WFE relativo FCN (%)	WFE relativo LS (%)
8	170.3	207.2	29	35
10	144.5	167.0	27	31
12	121.0	152.0	25	31

Tabla 12: Resultados medios obtenidos de la reconstrucción con redes FCN y el método de LS sobre las 400 muestras de cada test.



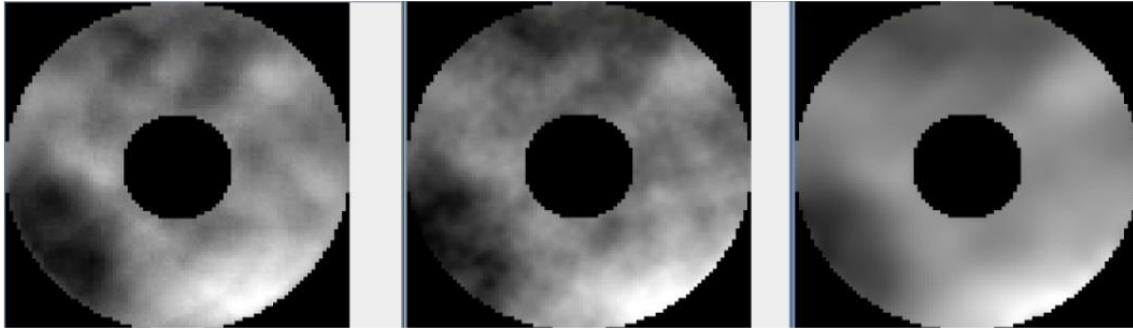


Figura 34: Ejemplo visual de la reconstrucción realizada para una muestra de test con  $r_0$  8 cm. La imagen del centro se corresponde con la fase original, a la izquierda la imagen de la reconstrucción con FCN y a la derecha la reconstruida con el método de LS.

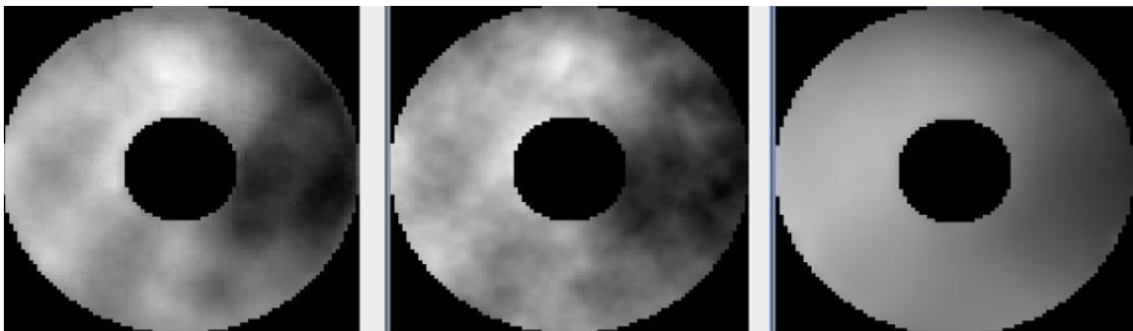


Figura 35: Ejemplo visual de la reconstrucción realizada para una muestra de test con  $r_0$  12 cm. La imagen del centro se corresponde con la fase original, a la izquierda la imagen de la reconstrucción con FCN y a la derecha la reconstruida con el método de LS.

### 8.2.2 Discusión de resultados

A partir de los resultados mostrados en la Tabla 12 se observa que la reconstrucción con las redes FCN mejora significativamente el error residual en comparación con las reconstrucciones realizadas mediante el método de mínimos cuadrados. En todos los casos se observa una mejora de entre el 4% y el 6% del error, pese a tratarse de una prueba de concepto donde quizás no se esperaban resultados tan satisfactorios y las topologías de la FCN aún tenían margen de mejora.

No obstante, como ya se comentó en la discusión de otros experimentos, era esperable el mejor rendimiento de las redes puesto que la información que reciben del SH es toda la luz recibida por las subaperturas, en lugar de dos valores por cada una de las subaperturas como lo sucede al método LS.

Observando tanto la figura 34 como 35, se observa de forma visual claramente como la reconstrucción situada a la izquierda, realizada con redes FCN, es mucho más precisa y fiel a la imagen original que la situada a la derecha. Situación esperable ya que el método LS aproxima la turbulencia mediante un número determinado de coeficientes de Zernike en lugar de obtener toda la imagen turbulenta como en el primer caso.

## 8.3 RECONSTRUCCIÓN MEDIANTE REDES NEURONALES PARA UN SISTEMA CON LA CONFIGURACIÓN GLAO SOLAR

Se trata de la tercera y última ponencia de congreso realizada durante el desarrollo de la tesis, en la que se realizó la reconstrucción para un sistema GLAO Solar. Hasta entonces, todas las pruebas realizadas eran con la configuración SCAO Solar, la más sencilla ya que solamente

cuenta con un único SH en el eje de observación. Los resultados fueron expuestos en el congreso “22th International Conference Computational and Mathematical Methods in Science and Engineering”.

La configuración GLAO presenta más dificultades a la hora de la reconstrucción ya que el sistema recoge información de la luz de varios SH apuntando en diferentes direcciones, con diferentes caminos ópticos y, a partir de ella, se reconstruye la turbulencia situada en el eje de observación del telescopio. El reconstructor debe entonces interpretar la información de los SH con una visión global para unificarla en una única salida. La ventaja de contar con mayor número de SH en diferentes direcciones permite realizar observaciones con mayor FOV que en el caso del SCAO. Este experimento se trata de una primera prueba para evaluar la capacidad de las redes neuronales artificiales de resolver problemas de OA más complejos que los sistemas SCAO.

Como se trató de la primera prueba realizada con una configuración diferente a SCAO, para este caso se utilizaron redes MLP. Con ellas se recoge la información de los centroides medida por los diferentes SH y se predicen un número determinado de los coeficientes de Zernike de la turbulencia, en el caso de esta prueba concreta, los 150 primeros excluyendo el 1, 2 y 3 (la pendiente, la inclinación y el pistón).

En concreto, en los conjuntos de datos generados se simularon 6 SH focalizados en diferentes direcciones, peinando una forma de observación rectangular, con 5 arcossegundos de distancia entre cada una de las zonas. Cada SH tenía  $15 \times 15$  subaperturas. Las entradas de la red consistieron entonces en 354 valores de pendiente por cada uno de los SH, ese valor no es el resultado de  $15 \times 15$  (*subpaerturas*)  $\times 2$  puesto que las subaperturas de las esquinas no reciben luz y no aportan datos a los valores de entrada.

El conjunto de datos de entrenamiento consistió en un set de datos con 26000 muestras, con capas turbulentas variando entre los 0 m y los 10000 m de altitud, en pasos de 100 m cada 200 muestras. El valor de  $r_0$  se varió entre 7 cm a 20 cm en pasos de 1 cm.

Los datos de test consistieron en un único conjunto de datos con 3000 muestras con un valor fijado de  $r_0$  de 10 cm.

Las simulaciones se realizaron con la plataforma DASP [32] en su versión GLAO solar. Tanto las simulaciones como el entrenamiento de las redes se realizaron en un ordenador provisto de Ubuntu LTS 14.04.3 con una CPU Intel Xeon E5-1650 v3 @ 3.50 GHz, una memoria DDR4 de 128 GB, tarjeta gráfica Nvidia GeForce GTX TitanX y disco duro SSD. El lenguaje de programación utilizado fue Python, siendo Keras el framework de redes neuronales escogido.

### 8.3.1 Resultados

En la Tabla 13 se muestran los resultados obtenidos, acompañados de la **¡Error! No se encuentra el origen de la referencia.** y la **¡Error! No se encuentra el origen de la referencia.** donde se presentan las imágenes de la fase reconstruida para dos muestras aleatoriamente escogidas del conjunto de datos.

Resultados obtenidos para $r_0 = 10\text{cm}$	WFE residual		WFE Relativo Residual (%)	Similitud (%)
	rad	nm		
	3.92	312.10	56	95

Tabla 13: Resultados obtenidos por la red MLP sobre el conjunto de datos con  $r_0$  10 cm.

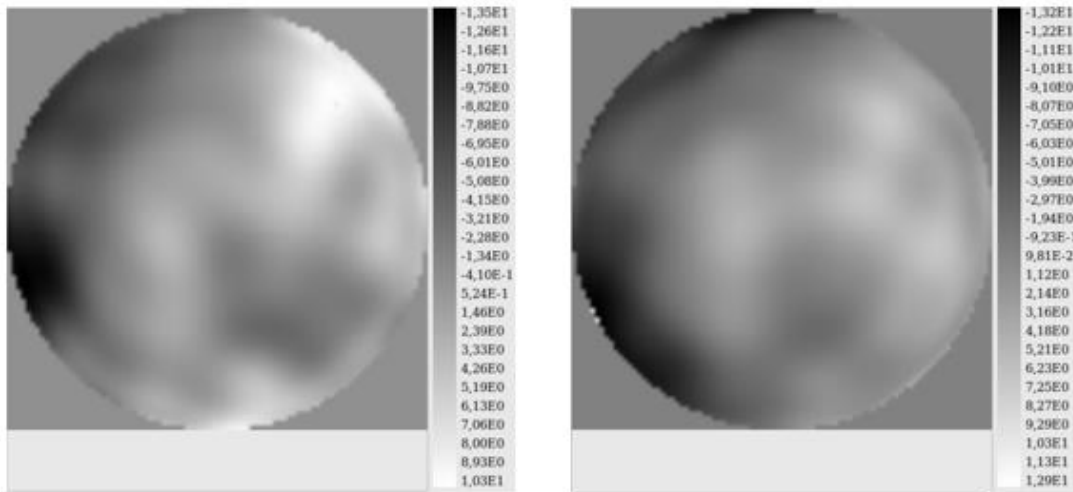


Figura 36: Comparación para una muestra del test de error residual 230 nm, por debajo del error residual medio obtenido sobre todo el conjunto de datos de test. La imagen de la izquierda se corresponde con la fase reconstruida.

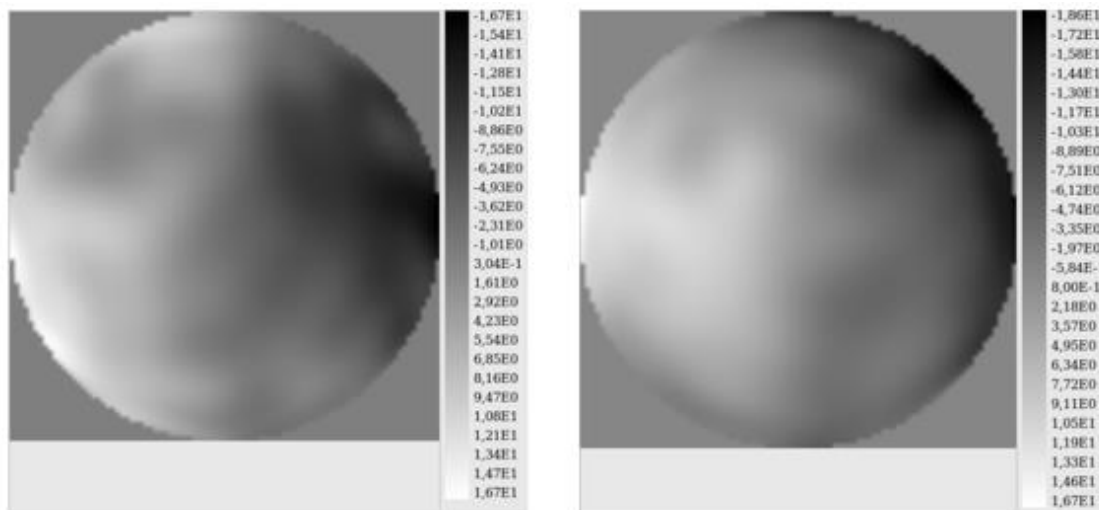


Figura 37: Comparación para una muestra del test con error residual 360 nm, por encima del error residual medio obtenido sobre todo el conjunto de datos de test. La imagen de la izquierda se corresponde con la fase reconstruida.

### 8.3.2 Discusión de resultados

Lo primero que llama la atención de esta prueba es que el error residual es superior a las pruebas que se estaban realizando anteriormente en SCAO. Hay dos factores fundamentales que influyen en este resultado; el primero es que se trata de una primera prueba de aplicación de redes a la configuración GLAO Solar, por lo tanto, el objetivo no era tanto obtener una reconstrucción precisa sino determinar si las redes neuronales podrían ser una alternativa a los reconstructores tradicionales para solucionar este problema.

En segundo lugar, se está reconstruyendo una aproximación de la fase, por lo tanto, recuperar la fase a partir de un número limitado de coeficientes de Zernike y compararla con la original implica tener un error debido ya a la aproximación, independientemente del que después sea debido a las redes neuronales.

Sin embargo, sí que se pueden extraer conclusiones prometedoras cuando se observan las fases visualmente en las figuras 36 y 37. En el caso con mayor error, la figura 37, las zonas

donde la turbulencia presenta máxima o mínima intensidad se corresponden en ambos casos con la misma zona aproximadamente de la figura, siendo sus formas también similares. En el caso de la figura 36 las similitudes son aún mayores con la fase original, como cabría esperar dado su error residual menor. Esta similitud entre imágenes queda patente analizando su similitud, si bien es inferior a otras pruebas estudiadas, sigue teniendo un porcentaje muy alto.

Por todo ello, de esta prueba se puede concretar que las ANN pueden ser una alternativa a los reconstructores para sistemas GLAO solares, pero aún deben ser más perfeccionadas en la búsqueda de un error residual menor.

## 9 OTROS RESULTADOS OBTENIDOS PENDIENTES DE PUBLICACIÓN

---

A lo largo del proceso de investigación de esta tesis se fueron obteniendo resultados en diferentes pruebas, en algunos casos más prometedores que otros, que no fueron publicados o llevados a congresos internacionales por diversos motivos, principalmente dos; por ser investigaciones similares a otras ya publicadas o por no haber dado tiempo a su redacción y presentación y están pendientes de ser publicados.

En este capítulo se muestra un resumen de los resultados más prometedores obtenidos pero que, por diversos motivos, aún no se han presentado en ningún congreso o revista de investigación.

### 9.1 RECONSTRUCTOR BASADO EN REDES NEURONALES ARTIFICIALES PARA OA SOLAR CAPAZ DE GENERALIZAR A DIFERENTES REGIONES SOLARES.

Todas las pruebas mostradas en las secciones anteriores mostraban resultados obtenidos por redes neuronales cuando se aplicaban sobre regiones del Sol ya conocidas. Es decir, las regiones del Sol sobre las que se entrenaban las redes neuronales y después se testeaban eran las mismas.

El problema de la generalización es una cuestión que surge al trabajar en óptica adaptativa solar. En el caso de la óptica nocturna, la luz recibida por los SH de los cuerpos celestes es similar en prácticamente todas las situaciones, son puntos de luz (en mayor o menor número) con una intensidad y grosor dependiente de cada situación. Sin embargo, en el caso solar, la imagen que se recibe en el SH cambia completamente de la región de la superficie solar que se esté observando. Este hecho se trata de un factor limitante, especialmente cuando se trabaja con redes convolucionales o completamente convolucionales, ya que esa modificación en los datos de entrada respecto a la esperada genera problemas a la hora de determinar la turbulencia.

La solución más sencilla a este problema sería pre-entrenar al modelo de red neuronal antes de realizar observaciones, de manera que la región sobre la que se vaya a realizar la observación ya sea previamente conocida por el sistema reconstructor. No obstante, se trata de una limitación que los reconstructores tradicionales basados en el método de reconstrucción por mínimos cuadrados LS no presentan. Por todo ello, se diseñó un modelo de red que fuese capaz de realizar reconstrucciones de la turbulencia atmosférica en regiones en las que no hubiera trabajado previamente.

Para esta prueba se decidió optar por una red MLP en un sistema SCAO solar, con un SH de  $15 \times 15$  subaperturas, que sería entrenada con un conjunto de datos de 195000 muestras de 4 regiones diferentes del Sol (escogiendo regiones de la imagen `imsol.fits` que posee la plataforma DASP de la superficie solar). Para la validación de la red durante el proceso de entrenamiento se utilizaron 6000 muestras de 10 regiones del Sol bien diferenciadas. Para las pruebas de rendimiento de la red se utilizaron tres conjuntos de datos caracterizados por el valor del parámetro  $r_0$ : 8, 10 y 12 *cm* respectivamente. Fueron generados utilizando otras 20 regiones solares, todas ellas diferentes a las de entrenamiento y validación.

Al tratarse de una red MLP, los datos simulados consistieron en las pendientes de cada subapertura medidas por el SH como datos de entrada y los coeficientes de Zernike como datos de salida. Para este caso, se guardaron los 153 primeros coeficientes de Zernike excluyendo la

inclinación, la pendiente y el pistón, ya que normalmente son corregidos por el espejo primario. Tanto las simulaciones como el entrenamiento de las redes se realizaron en un ordenador provisto de Ubuntu LTS 14.04.3 con una CPU Intel Xeon E5-1650 v3 @ 3.50 GHz, una memoria DDR4 de 128 GB, tarjeta gráfica Nvidia GeForce GTX TitanX y disco duro SSD. El lenguaje de programación utilizado fue Python, siendo Tensorflow el framework de redes neuronales escogido.

### 9.1.1 Resultados

Los resultados obtenidos de la prueba en términos de WFE residual se muestran en la tabla 14 mientras que en las figuras 38 y 39 se presentan algunos ejemplos visuales de las reconstrucciones obtenidas.

$r_0$ (cm)	WFE residual (nm)	WFE residual relativo (%)
8	169,51	27,7
10	140,86	27,7
12	117,78	27,8

Tabla 14: Resultados en términos de WFE residual de las reconstrucciones realizadas por la red MLP sobre 20 regiones del Sol desconocidas.

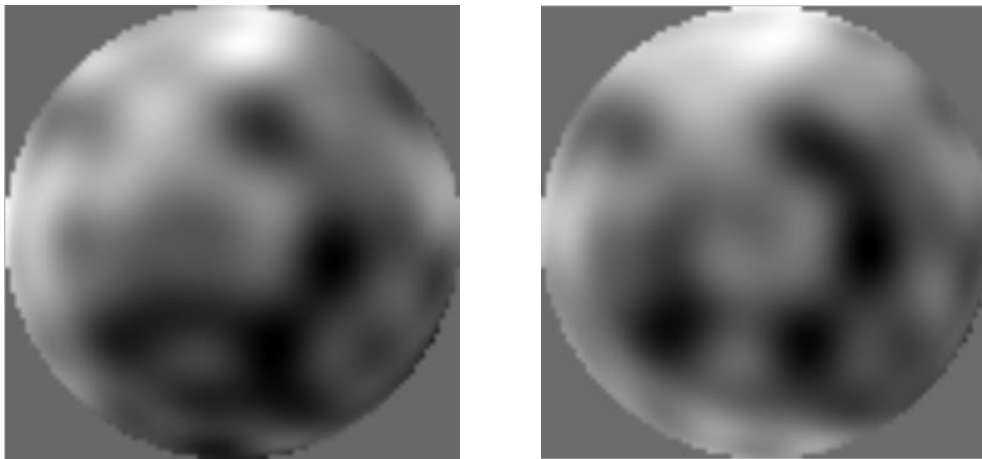


Figura 38: Comparación visual para un caso de  $n$  caso de  $r_0 = 8$  cm, la imagen de la izquierda si corresponde con la imagen reconstruida por la red mientras que la imagen de la derecha con la fase original. Se trata de un caso con WFE residual de 194.17 nm, por encima del valor medio obtenido en el test.

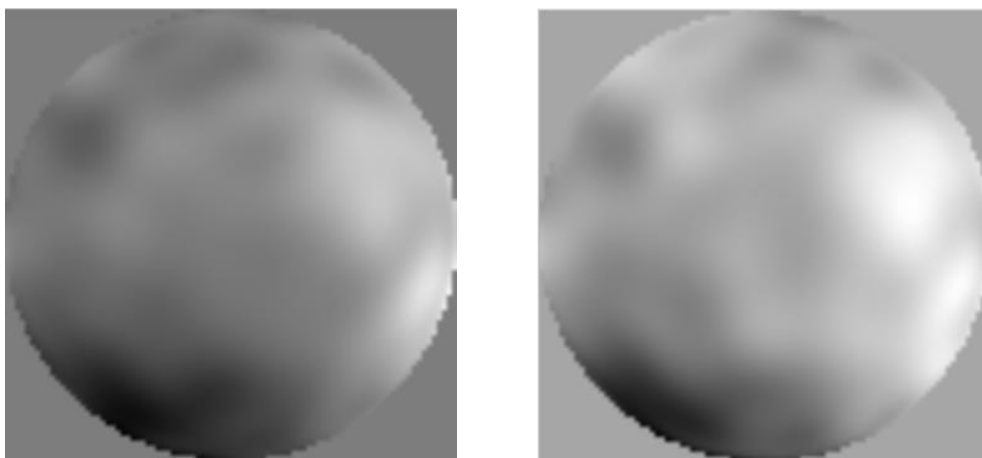


Figura 39: Comparación visual para un caso de  $n$  caso de  $r_0 = 10$  cm, la imagen de la izquierda si corresponde con la imagen reconstruida por la red mientras que la imagen de la derecha con la fase original. Se trata de un caso con WFE residual de 159.96 nm, más de un 13% superior al valor medio obtenido en el test.

### 9.1.2 Discusión de resultados

Tanto de las figuras como de los resultados mostrados en la Tabla 14 se puede ver como la red diseñada es capaz de generalizar, obteniendo valores de WFE residual del orden de los experimentos anteriores, pero sobre regiones del Sol completamente desconocidas. Para ello fue necesario entrenar la red con múltiples regiones, permitiendo así que obviar la información procedente de la región solar y centrándose únicamente en la proveniente de la turbulencia atmosférica.

Otro aspecto para destacar es la consistencia del error con la variación de intensidad, en la tabla 14 se observa que el WFE relativo está, en todos los casos, alrededor del 28% independientemente del valor de  $r_0$ .

Por último, la comprobación visual que permiten realizar las figuras 38 y 39 muestran como al igual que en pruebas anteriores, las zonas de máxima y mínima intensidad de las fases reconstruidas coinciden aproximadamente con las de la imagen original, teniendo formas similares.

De todos estos resultados se puede concluir que, al menos las redes MLP, son sistemas de IA que pueden ser utilizadas como reconstructores de OA Solar sin necesidad de ser entrenadas previamente sobre las regiones del Sol en las que sea necesario utilizarlas.





## 10 CONCLUSIONES Y LÍNEAS FUTURAS

---

A lo largo de esta memoria de tesis, se han presentado los pasos dados, junto con los resultados logrados, con el objetivo principal de conseguir un reconstructor basado en redes neuronales para óptica adaptativa solar. Este cometido se ha vertebrado en tres fases fundamentales, fruto de tres estudios publicados y descritos, que han constituido los fundamentos de esta tesis.

En primer lugar, se realizó un estudio en observaciones nocturnas, con el objeto de determinar qué tipo de red neuronal ofrecía los mejores resultados como sistema reconstructor. En concreto, se analizaron redes MLP, CNN y FCN.

Se verificaron tanto la calidad de las reconstrucciones como el tiempo de computación necesario para cada uno de los diferentes modelos:

- a) En cuanto a la calidad de las reconstrucciones realizadas, cuando se pretendía reconstruir la fase completa original, la FCN demostró que se conseguían las reconstrucciones de la fase con menor error respecto a la original, además de ser también el modelo más consistente, puesto que apenas había diferencia en el error cometido en función de la turbulencia atmosférica (solamente un 2% de diferencia entre los casos de alta intensidad de turbulencia y baja intensidad).
- b) En la segunda cuestión, se observó que la MLP era el modelo de red que alcanzaba los resultados con mayor rapidez: aproximadamente en la mitad de tiempo. Sin embargo, las entradas que necesitaba la MLP requerían un mayor preprocesamiento que en los supuestos de la CNN y la FCN. Por tanto, se concluyó que, si bien seguía siendo la más veloz, esa diferencia no era tan significativa.

Valorando estos buenos resultados, se eligieron las redes FCN como modelos prioritarios para efectuar las siguientes investigaciones.

En segundo lugar, una vez conocido el modelo de red que debía desarrollarse como reconstructor, se continuó la investigación determinando, a partir de los datos medidos por los sensores de frente de onda, cuáles de ellos serían los más interesantes para usarlos como entrada de estas redes. Para esto, se llevó a cabo un segundo estudio, esta vez simulando observaciones solares. Además, también se pretendía comprobar que, efectivamente, las FCN se comportaban de igual manera tanto en reconstrucciones diurnas como en nocturnas.

Sin valorar el tipo de entrada utilizada en la red, la primera conclusión distintiva que pudo extraerse fue que, en todos los casos, las redes FCN mejoraban, en términos de error de frente de onda residual, las reconstrucciones realizadas mediante el método de mínimos cuadrados, que es el más utilizado por los reconstructores actualmente. Es decir, se conseguían reconstrucciones de la fase turbulenta con mayor calidad.

En cuanto al objetivo propio de este segundo trabajo, se comparó entre ofrecer a la FCN como entrada, o bien la imagen de todas las subaperturas en conjunto del SH con la luz que recibían en cada momento sin preprocesamiento, o bien las correlaciones calculadas ya entre las diferentes subaperturas por el propio SH.

A la vista de los resultados, se evaluó que la red alcanzaba, cuando se le aportaba toda la información recibida por el sensor sin preprocesar, menores valores de error que en el segundo caso, aunque también es verdad que la diferencia entre las dos situaciones era mínima, pudiéndose trabajar en cualquiera de los dos supuestos. Teniendo en cuenta la mejor calidad de las reconstrucciones y el ahorro de tiempo que podría suponer evitar ese preprocesamiento de imágenes, se optó, como mejor entrada para los reconstructores de IA, por las imágenes de la luz recibida directamente por los sensores SH.

En tercer lugar, una vez conocido el modelo de red que se pretendía utilizar y los datos de entrada que se le debían aportar, parecía lógico continuar la investigación indagando qué tipos de configuraciones de sensores y observaciones propiciaban un mejor funcionamiento de la red y si este tipo de configuraciones ofrecía una ventaja añadida en el uso de redes neuronales como sistemas reconstructores –dejando aparte la propia ya demostrada, que suponía una reconstrucción de mayor calidad-.

Los resultados alcanzados en el tercer estudio publicado demostraron que así sucedía. Las reconstrucciones de mayor calidad realizadas por las redes FCN se producían al disminuir el número de subaperturas de los sensores SH. Los algoritmos de reconstrucción clásicos mejoran su precisión según incrementan el número de subaperturas, pero, como es obvio, este número está limitado por el tamaño de la pupila del telescopio y por factores económicos. Aumentar el número de subaperturas para un SH del mismo tamaño implica un gran incremento de costes, puesto que se necesita un mayor número de lentes para el mallado en el mismo espacio, por lo que deben ser más pequeñas y precisas.

No ocurría así en el caso de las redes, ya que, con sensores mucho más simples, se obtenían incluso mejores resultados que con los más complejos. Como ejemplo de alguno de los casos analizados, el error residual disminuía a la mitad bajo el mismo valor del parámetro  $r_0$  en un sensor de  $7 \times 7$  subaperturas con respecto a un sensor de  $15 \times 15$ . Un dato aún más sorprendente fue que, hasta con un sensor de solamente  $3 \times 3$  subaperturas, se conseguía mayor calidad en las reconstrucciones que en el supuesto mencionado anteriormente.

También se analizó la influencia del número de píxeles por subapertura y del campo de visión con el que se realice la observación.

- a) Se pudo concluir que aumentar el número de píxeles por subapertura mejoraba la calidad de la reconstrucción, especialmente en términos de similitud entre la fase reconstruida y la fase original, aunque tampoco era una diferencia muy importante. En el caso analizado con mayor diferencia, se pasaba de una similitud del 84% al 91%.
- b) En cuanto a si el campo de observación utilizado influía o no, se comprobó que, en las muestras analizadas con mayor campo de visión, las reconstrucciones de la red fueron más precisas, aunque tampoco se pudo establecer de los resultados logrados que se tratase de un factor muy diferencial, como sí sucedía en el caso de las subaperturas.

A lo largo de esta investigación, se ha demostrado satisfactoriamente, por todos los resultados evaluados, la viabilidad de la aplicación de redes neuronales como sistemas reconstructores de óptica adaptativa solar, sustentándose siempre sobre datos de simulación. Por tanto, las ANN surgen como potenciales candidatas para sustituir los algoritmos de reconstrucción que, actualmente, se emplean en los sistemas de óptica adaptativa por diversos motivos:

- Se han desarrollado modelos de redes neuronales que han obtenido errores residuales en la reconstrucción de la fase del frente de onda menores que los algoritmos de reconstrucción más utilizados sobre los mismos datos de simulación.
- Se ha comprobado la idoneidad de utilizar redes neuronales completamente convolucionales FCN como sistemas de reconstrucción: en primer lugar, por ser los modelos que obtuvieron menores errores de reconstrucción y, en segundo lugar, por las grandes ventajas que otorga la capacidad de usar las imágenes recibidas por los sensores de frente de onda directamente como entradas del sistema rector, posibilitando directamente las imágenes de fase como salida.
- Los sistemas de reconstrucción basados en IA han sido capaces de realizar reconstrucciones con gran precisión, utilizando tiempos de computación bajos, inferiores a los 5 ms, y ahorrando también diversos procesos habituales de preprocesamiento de los datos recibidos, ya que no eran necesarios en combinación con estas nuevas tecnologías.
- Basándose en datos de simulación, se ha demostrado que se obtienen los mejores resultados en combinación con sistemas de sensores más simples que los que se utilizan en los telescopios en la actualidad, con el consiguiente beneficio económico que implicaría el uso de estos sistemas.

En suma, en esta recapitulación se aprecian las conclusiones más relevantes que se han alcanzado en la investigación hasta el momento de la elaboración de este documento. Sin embargo, el desarrollo de estos sistemas continúa, aun con otras posibles líneas de investigación abiertas, dados los prometedores resultados conseguidos. Algunas de las que se están siguiendo en la actualidad por nuestro propio grupo de investigación son: mejora de los resultados obtenidos en combinación con configuraciones de OA más complejas como GLAO (con la intención de ampliarlo a MCAO), mejora de la generalización a regiones del Sol previamente no entrenadas en combinación con redes FCN, obtención de datos de observación reales para evaluar el funcionamiento de los sistemas (datos de telescopios o de bancos ópticos) fuera de entornos de simulación, etc.



## 11 COPIA COMPLETA DE LOS ARTÍCULOS PUBLICADOS




---

### 11.1 OVERVIEW AND CHOICE OF ARTIFICIAL INTELLIGENCE APPROACHES FOR NIGHT-TIME ADAPTIVE OPTICS RECONSTRUCTION

**AUTORES: FRANCISCO GARCÍA RIESGO, SERGIO LUIS SUÁREZ GÓMEZ, JESÚS DANIEL SANTOS RODRÍGUEZ, ENRIQUE DÍEZ ALONSO Y FERNANDO SÁNCHEZ LASHERAS**

Article

# Overview and Choice of Artificial Intelligence Approaches for Night-Time Adaptive Optics Reconstruction

Francisco García Riesgo <sup>1,2</sup>, Sergio Luis Suárez Gómez <sup>2,3</sup>, Jesús Daniel Santos <sup>1,2</sup>, Enrique Díez Alonso <sup>2,3</sup> and Fernando Sánchez Lasheras <sup>2,3,\*</sup>

<sup>1</sup> Department of Physics, University of Oviedo, 33007 Oviedo, Spain; garciariefrancisco@uniovi.es (F.G.R.); jdsantos@uniovi.es (J.D.S.)

<sup>2</sup> Instituto Universitario de Ciencias y Tecnologías Espaciales de Asturias (ICTEA), 33004 Oviedo, Spain; suarezsergio@uniovi.es (S.L.S.G.); diezenrique@uniovi.es (E.D.A.)

<sup>3</sup> Department of Mathematics, University of Oviedo, 33007 Oviedo, Spain

\* Correspondence: sanchezfermando@uniovi.es

**Abstract:** Adaptive optics (AO) is one of the most relevant systems for ground-based telescopes image correction. AO is characterized by demanding computational systems that must be able to quickly manage large amounts of data, trying to make all the calculations needed the closest to real-time. Furthermore, next generations of telescopes that are already being constructed will demand higher computational requirements. For these reasons, artificial neural networks (ANNs) have recently become one alternative to commonly used tomographic reconstructions based on several algorithms as the least-squares method. ANNs have shown its capacity to model complex physical systems, as well as predicting values in the case of nocturnal AO where some models have already been tested. In this research, a comparison in terms of quality of the outputs given and computational time needed is presented between three of the most common ANN topologies used nowadays, to obtain the one that fits better these AO systems requirements. Multi-layer perceptron (MLP), convolutional neural networks (CNN) and fully convolutional neural networks (FCN) are considered. The results presented determine the way forward for the development of reconstruction systems based on ANNs for future telescopes, as the ones being under construction for solar observations.

**Keywords:** artificial neural networks; machine learning; adaptive optics



**Citation:** García Riesgo, F.; Suárez Gómez, S.L.; Santos, J.D.; Díez Alonso, E.; Sánchez Lasheras, F. Overview and Choice of Artificial Intelligence Approaches for Night-Time Adaptive Optics Reconstruction. *Mathematics* 2021, 9, 1220. <https://doi.org/10.3390/math9111220>

Academic Editor: Paolo Crippa

Received: 3 May 2021

Accepted: 24 May 2021

Published: 27 May 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Nowadays, ground-based telescopes used for astronomical observations are a set of many system designed to obtain the best possible image of celestial bodies. Between all the machinery, optics systems, engineering components, many different sensors, as well as data processing systems and many other are set to work together for the same purpose [1].

One of the most important problems of ground-based telescopes are the aberrations formed on the received light, most of them produced by the atmospheric turbulence. The wavefronts of the received light travels for millions of light years almost without aberrations and when they go through the atmosphere they become completely deformed. Moreover, the turbulence is a random phenomenon which is constantly changing along its approximately 20 km altitude, so it implies a high difficulty when compensating its effects [2].

Adaptive optics (AO) are a set of methodologies that aim to correct the deviances suffered by the light received from celestial bodies in the telescopes. The reconstruction system (RS) is one of the most important parts of any AO structure, since it oversees the calculation of the corrections to avoid as much as possible the aberrations of the received image [3].

It is known the good results obtained when using CNNs with images due to their ability to extract the main features of the given input. In this case, the input that will be



given to the RS is the image of the light received by the wavefront sensor (WFS) that in this case will correspond to a Shack–Hartmann (SH) [4]. The main difference with previous RSs developments resides in the use of FCNs that allow to obtain a new image as output of the RS. Until now, the RSs commonly obtain as output a certain number of Zernike polynomials that describe the atmosphere at that moment [5]. The more Zernike modes, the more detailed the atmosphere will be. With this approach, the output obtained is a new image that completely describes the phase of the atmosphere at each moment, given much more information to the other systems that manage the aberrations presented in the received light.

In the latest years, artificial intelligence systems have been applied in a broad range of science branches [6,7]. In optical applications specifically, some kinds of artificial neural networks as the convolutional ones [8], have shown good results in image recognition, language processing [9], image classification [10], etc. Several studies have taken advantage of that developing. Previous research use the simpler kinds of ANNs to test the reconstruction as RS systems for nights observations, being CARMEN (complex atmospheric reconstructor based on machine learning) [11] and convolutional CARMEN [12]; both showing excellent results in comparison with the currently most used techniques as the least-square method (LS) [13]. They were developed as an MLP and a CNNs respectively [14]. In this work, the results obtained by the new RS based on FCNs are compared with the previous ones.

The purpose of this paper is to make a comparison between a new RS for night observations based on fully-convolutional neural networks (FCNs) [15] with other RS systems based on others kinds of artificial neural networks (ANNs) previously developed. In particular it is compared with a RS based on convolutional neural networks (CNNs) [12] and other based on the multilayer perceptron (MLP) [16]. In addition, results from this research improve previous approaches to the problem, as in [8] where images from the sun are used to estimate deformable mirror actuators, limiting these studies to the particular engineering of the considered scenario; in this paper, our target for the networks are turbulence phases, thus implementing a more general approach.

In this work, an introduction about adaptive optics, Shack–Hartmann wavefront sensing, and the simulation platform used to generate the data is presented in Section 2, along with concepts of artificial intelligence and ANNs and well as the computational resources required for this work. Next, obtained results are detailed in Section 3, showing their discussion in Section 4. Finally, some conclusions and future developments are shown in the last section.

## 2. Materials and Methods

Adaptive optics aim to correct the aberrations produced by the atmospheric turbulence in the light received in ground-based telescopes from celestial bodies. The Earth's atmosphere has constantly movements between different air masses with its own characteristics, as the speed of the wind, density, temperature, etc. The air masses can be schematically represented as turbulence layers at different heights. The performance is like the fluid's movements. The turbulences at greater scales are passed to turbulences at lower scales, consequently, new turbulence layers are being forming at each moment [17]. The way that the air masses affect the light depends on the parameters mentioned before, according to that, each layer has a different refraction index value. That affects the received light on its way through the atmosphere as it has been traversed a path with multiple lenses of various refraction index, one after the other.

An AO system includes all the techniques required to correct the astronomical images, from the measures of the turbulence to the correction of the light's phase in the telescope. That includes sensing the image that is being observed, reconstructing the turbulence atmosphere of that moment, and finally making the reconstruction trying to eliminate the aberrations presented. Each process is performed by different systems. A typical configuration of an AO system includes at least one WFS, the SH commonly is the most used one, that obtains information of the incoming light to know how the turbulence is,

then that information is passed to a RS system. There are many kinds of RS systems, it can be based on algorithms as the least-squares (LS) method, the learn and apply method [18], or it can use ANNs as the CARMEN reconstructor [19], etc. Finally, there is another system which aims at the correction of the wavefront received from the information given by the RS, it is the deformable mirror (DM). It consists of a mirror that can vary its surface shape to modify the wavefront received, trying to unmake the aberrations. There are different types of DM [5], which vary in the way the phase of the light is reshaped, being able to adjust the surface with piezoelectric actuators on a continuous surface, or using separate sections of reflective surface.

The principal components of an AO system can be combined in different ways, their order, and the number of elements of each type can be modified, depending on the observation that is wanted to be made. In the case of these researches, a single-conjugated AO system was chosen [20]. It is the simplest system in AO as it only has one WFS on-axis and one DM working in close loop. The WFS and the DM are on-axis of the reference star, being an SH the type of WFS chosen. Other more complex AO systems could be the multi-object AO system, which consists of more than one reference star, situated off-axis of the telescope, in different regions of the sky. Each reference star has a WFS associated, that work all together to correct the optic path on-axis of the telescope. The main advantage of this more complex systems is that they allow to correct a bigger field of view than the SCAO ones.

### 2.1. Shack-Hartmann WFS

In this sub-chapter, the SH is explained, as it is the WFS chosen for the research. The information that is given as input of our RS is directly the light received by the SH. Consequently, it is necessary to understand how the performance of the source of information is.

The SH is characterized by dividing the received wavefront into several subapertures [21,22]. It has a set of lenses with the same focal length, that focalized the incoming light in different subapertures as it is showed in Figure 1. In an ideal situation, the light of the celestial body would be received as plane wavefront that, after passing through the lens, will be focalized in the center of each subaperture. Due to the presence of the atmosphere turbulence, the wavefront presents aberrations that are represented by two characteristics. First of all, a spot is not received in each subaperture, actually receiving a blurry image which is not from plane wavefront. Secondly, the blurry image is not centered in the middle of the subaperture, it is placed in different position depending on the subaperture.

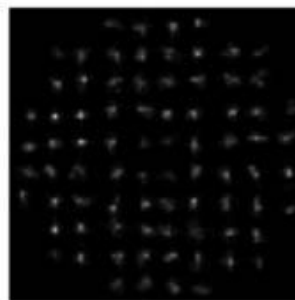


Figure 1. Example of the light received by the Shack-Hartmann wavefront sensor in an SCAO configuration with only one star on axis.

A significant magnitude that is measured by the SH is the centroid of each subaperture. It corresponds with the gravity center of the spot received and it is commonly used by the sensor to calculate the slopes of the received wavefront. The slopes are the main information used by most RS systems to reconstruct the original wavefront, as in the case



of the RS based on MLPs ANNs or the LS RS. That is not the case of the CNNs or our new FCNs, considering the properties of using all the image received by the SH. This means that they make the most of all the information received instead of the other RS where only two values of each subaperture are used. These methods allow to have no loss of information avoiding pre-processing data process.

## 2.2. Artificial Neural Networks

In this subchapter an introduction to the different kinds of neural network used in this research is detailed. An ANN is a complex set of interconnected computational units that imitates the learning process made by biological neural networks [23]. Each computational unit is called a neuron. As in the biological case, ANNs can learn from the data to offer outputs for a deterministic problem.

The main characteristics of an ANN are that neurons are set in layers, where the neurons of one layer are usually interconnected with the neurons of the adjacent layers. Each one performs an operation over the input received and passed the output through its connections to the neurons of the next layer, so its output is part of the input of the neurons of the next layer. The connections are characterized by a numerical value called weight so that they regulate the influence of the outputs of the previous layers in the input of the next one.

The ANNs can learn from the data so it is necessary to make a train or learning process before. It consists of pass thorough the ANN the inputs of a data set, in which the desired outputs are known, this process is needed before their application. The ANN modifies the weights of the connections between neurons to obtain the most similar outputs to the desired ones possible, thanks to different algorithms that measure the error. The most use one is the backpropagation algorithm [24] and it is the one used in this research.

The output of the  $j$ th neuron of a network; being  $g(\cdot)$  its activation function and  $y_j$  its inputs that correspond with the outputs of the previous layers, is expressed mathematically as follows [25]:

$$y_j(t) = g(y_i) = g\left(\sum_i \omega_{ji} \cdot x_i + b_j\right) \quad (1)$$

where the sum represents the local input of the neuron, that considerate all its connections with their corresponding weights and its bias represented by  $b_j$ . The bias is an optional term that affects the output of the neurons and can be modified during the training process as the weights.

There are several kinds of neural networks, depending on its structure and the structure of the information to process, that is the structure of the input and the output of the ANN. In our case, different kinds of inputs and outputs were used depending on the ANN used for the RS; the multi-layer perceptron, the convolutional neural network [26] or the fully-convolutional neural network [27]. That information is detailed in the subchapter of each system.

### 2.2.1. Multi-Layer Perceptron Neural Network (MLP)

This is the simplest kind of ANN used in the experiment. MLPs are characterized by having the neurons organized in three types of layers: input, hidden and output layers. The input layer receives the data, that are passed to a hidden layer where they are processed and then they are given to the output layer where the final activation functions are applied, and the output of the ANN are obtained [28]. The higher number of hidden layers the ANN has, the architecture more complex is, as there are more interconnections with their corresponding weights so there are more parameters to adjust [29,30]. Both input and output data supported by an MLP architecture are vector-shaped arrays.

It was performed in python using the Keras library [31], with feedforward architecture, using the backpropagation algorithm during the training process. Each of the neurons has an activation function that defines the operation made given an input signal. The model

chosen was implemented with the scaled exponential linear unit (SELU) function [32], implemented in Keras, since it showed the lowest error in the several tests made.

The selu function [33] is a variation of the exponential linear unit (ELU) with a scaling factor:

$$g(x) = \alpha * \begin{cases} \beta e^x - \beta & \text{if } x < 0 \\ x & \text{otherwise} \end{cases} \tag{2}$$

Being  $\alpha$  and  $\beta$  two fixed constants ( $\alpha = 1.67326324$  and  $\beta = 1.05070098$ ) whose objective is to preserve the mean and the variance of the inputs when they passed from a layer to the next one.

In our case the datasets include the slopes measured from the centroids by the SH, that were given to the MLP as inputs, and the Zernike’s coefficients of the first 153 modes of the turbulence of that moment, being the desired output. From the Zernike’s polynomials the turbulence phase profile is reconstructed, to know the quality of the reconstruction made by the MLP.

### 2.2.2. Convolutional Neural Networks

CNNs are the second kind of ANNs proved as RS in this experiment. These networks were developed to cope with higher dimensional problems, where tensor-shaped arrays are required as inputs or outputs, which is a problem out of reach to the size limitations of an MLP architecture [34]. That means higher information to process, for example, having an image as input data instead of a vector. An image is represented in data terms as a 2D matrix, where each position of the matrix represents the information of a pixel. A 2D matrix may suppose too much information for an MLP, as the number of neurons and interconnections that would be necessary to process the information in an MLP would be very high, making the problem impractical for a computer to handle. CNNs allow the use of input data with more than two dimensions where the information of the relative position of a pixel is relevant. Nevertheless, the output of a CNN is a value or an array of output values (as in the MLPs case). CNNs have shown good results working with images as document recognition, image classification [35,36], etc.

CNNs are implemented by two kinds of neuron layers, convolutional and pooling layers, in combination with an MLP at the end of the system. The convolutional layers work as filters; the input is passed, and the filters are convoluted over the full image in combination with the application of the activation function. These layers are often combined with pooling layers whose aim is to reduce the size of the image by selecting the most significant values over a group of pixels. Depending on the kind of pooling layer selected, they extract the mean value, the maximum or the minimum of the group of pixels. Due to that, the main characteristics of the input of the ANN are contained in the output of the convolutional block with the advantage that the size is much smaller. Once the mean features are extracted and the size of the data reduced, the resultant output of the convolutional block is passed to an MLP to obtain the final output of the ANN.

CNNs are characterized by performing the convolution operation, a mathematical operation between two functions that produces a third function which represents the amount of overlap one function  $g(x)$  when it is moved over another function  $f(x)$  [34]. The operation is typically denoted with an asterisk:

$$s(x) = (f * g)(x) = \int f(a)g(x - a)da \tag{3}$$

As  $x$  can only takes integer values, as the array’s positions of the inputs where convolutions are applied, a discrete convolution can be defined. In this case it is assumed that the functions are zero everywhere except the points where the values are stored:

$$s(x) = (f * g)(x) = \sum_a f(a)g(x - a) \tag{4}$$



In convolutional terminology, the second argument ( $g(x)$  in (3)) is moved over the first argument that would represent in a CNN the input of the neural network. The second argument is known as the kernel, that works as a filter over the input, giving an output that is called a feature map.

The inputs of a CNN are usually a multidimensional array, so multidimensional filters are probably needed. In the case of this research, the inputs consist of 2D arrays (image  $I$ ) and the kernels are 2D tensors as  $K$ :

$$S(i, j) = (I * K)(i, j) = \sum_m \sum_n I(m, n)K(i - m, j - n) \quad (5)$$

The convolution is a commutative operation so:

$$S(i, j) = (I * K)(i, j) = \sum_m \sum_n I(i - m, j - n)K(m, n) \quad (6)$$

Consequently, the discrete convolution consists in new tensors as the summatory of elementwise product of the input and kernel tensors. The Kernel's size is typically smaller than the input size, so the kernel is applied to sections of the input repeatedly throughout the sample, with a step size called stride. The resulting feature map has, as maximum, the same size as the input but, increasing the stride size, it is possible to reduce the dimensions of the output. To keep the original size, the sample limits have to be externally increased to apply the kernel the same times on the edge pixels as on the central part of the image, so zeros are added. This process is known as padding.

Once the convolutions are made to the inputs of a layer, the neurons typically apply its activation functions, the same kind of functions as in the MLP case.

CNNs need a training process where the interconnection weights between neurons are adjusted to obtain the best outputs. The datasets used for the CNNs in the research include the images of the information received by the SH at each moment as input data (an example of a sample is showed in Figure 1) and the Zernike's coefficients of the first 153 modes of the turbulence as output. In this case, the output is obtained taking advantage of all the information received by the SH instead of using its data pre-processed.

### 2.2.3. Fully-Convolutional Neural Networks

FCNs are designed to improve CNNs [13], allowing a tensor-shaped output. This architecture is typically conformed by two sections, a convolutional block and a deconvolutional one.

The main block of an FCN remains being the convolutional one. However, the classification block of the CNN is substituted by a deconvolutional block on the FCN in order to obtain a new image. Its performance is the opposite to the convolutional block, from a small image where the main features of the input are contained a new completely image is obtained as output by passing the information thorough deconvolutional and transpose-pooling layers.

For a training process in this type of neural network, it is necessary to have a dataset with input images and their corresponding output image for each one. The more images the dataset has, the better is the training process. The filters of the layers and the weights of the connections are modified with the backpropagation, each time that the error is backpropagated they are updated. For the first iteration they are randomly initialized.

In our case, the dataset used for the training process of FCNs consists of the same input as in the convolutional case, the images received by the SH, but the outputs were another image, in this case the turbulence phase's profile for each SH image.

### 2.2.4. Learning Process in Neural Networks

This is also called the train process, and is a stage needed before the use of an ANN where the system modifies the weights between neurons in order to model or to predict the best outputs to a determinate problem. There are two mean kinds of learning process

- Supervised learning process: the desired outputs to a determinate input data of a train dataset are known. The accuracy of the model is measured by a cost function that during the training process is pretended to minimize.
- Unsupervised learning: the outputs are not known so the algorithm tries to make relationships between the outputs, as in ANNs used as classifying algorithms, by extracting the main features over the data. In these cases, an objective function is usually tried to be minimized. An example of that function in classificatory algorithms can be a measure over the difference of a main parameter over the outputs.

In the case of this research, supervised learning is employed, which this section is dedicated to. In a supervised learning, the first thing to determinate is the loss function that measures the error committed in the ANN's outputs to adjust the weights of the ANN.

For this research, the loss function chosen is the mean squared error (MSE) [37], one of the most commonly selected loss functions that is defined as follows:

$$MSE = \frac{1}{n} \sum_n \|\hat{y} - y\|^2 \tag{7}$$

where  $\hat{y}$  represents the desired output and  $y$  the output given by the ANN being  $n$  the number of samples of train set. Other functions that can be used as loss functions are the mean squared error logarithm function, the mean absolute error function, etc. The learning process consists then in an optimization problem where it is intended to minimize the loss function by varying the weights values [38].

There are several algorithms based on gradient descent used to modify the ANN's weights values, the one used for this research is the back-propagation algorithm (BP) [39]. The BP is explained below for a simple case of an MLP with only one hidden layer but, from the explanation of this particular case, the main concept can be applied to any other type of neural networks, as the CNNs or the FCN.

Considering a three-layer MLP, the output of the  $k$  neuron of the output layer is expressed as:

$$z_k = g\left(\sum_j \omega_{kj}' y_j + b_k'\right) = g\left(\sum_j \omega_{kj}' g\left(\sum_i \omega_{ji} x_i + b_i\right) + b_k'\right) \tag{8}$$

where  $\omega_{kj}'$ ,  $b_k'$  represents the weights and bias from the output layer. In this case, the MLP used in the experiment is replicated where the same activation function  $g(\cdot)$  is applied to both layers and the MSE is considered as loss function, to calculate differences between the data patterns and those obtained by the network. The error  $E$  of the MLP's output according to (7) and (8):

$$E(\omega_{ji}, b_j, \omega_{kj}', b_k') = \frac{1}{n} \sum_n \sum_k \left\| \hat{z}_k^n - g\left(\sum_j \omega_{kj}' y_j^n + b_k'\right) \right\|^2 \tag{9}$$

The error is going to be minimize by gradient descent. For this case there are two different gradients, depending on the weights of the output layer and the hidden layer. For  $\epsilon$  the learning rate:

$$\delta \omega_{kj}' = \frac{dE}{d\omega_{kj}'} \epsilon \text{ and } \delta \omega_{ji} = \frac{dE}{d\omega_{ji}} \epsilon, \tag{10}$$

it is intended to find values of  $\Delta \omega_{kj}$  and  $\Delta \omega_{ji}$  that make a negative gradient since they minimize the loss function. Then, calculating the derivative of the loss function applying the chain rule the expressions of updating weights are obtained. For more simplicity, the term  $h_x$  represents the input of the  $x$  neuron according to (1):

$$h_x = \sum_y \omega_{xy} x_y + b_y \tag{11}$$



So,

$$\delta\omega'_{kj} = \varepsilon \sum_n \Delta_k^m y_j^n \text{ being } \Delta_k^m = \|z_k^m - g(h_k^m)\| \left( \frac{dg(h_k^m)}{dh_k^m} \right) \tag{12}$$

and

$$\delta\omega_{ji} = \varepsilon \sum_n \Delta_j^n x_i^n \text{ being } \Delta_j^n = \left( \sum_k \Delta_k^m \omega_{jk} \right) \left( \frac{dg(h_j^n)}{dh_j^n} \right). \tag{13}$$

The  $\Delta_x$  terms are called error signals. The concept of backpropagation is shown in the expressions as the first term to be calculated is  $\Delta_k^m$ , an error signal that is proportional to value of the loss function as it is the signal error of the output layer. Then the output layer weights are updating  $\delta\omega'_{kj}$  prior to the calculation of the error signal from the previous layer and continuing the process of updating the wights of the hidden layer. Clearly the errors signals are propagated from the output layer backward to the hidden layer.

The weights of an ANN are usually randomly initialized before being modified during the learning process. The batch size represents the number of samples of the train dataset to which the neural network is applied before the wights are modified. The learning process can be resume in the next steps:

1. The weights are randomly initialized before the training process.
2. Apply the ANN one batch:
  - a. The ANN is applied sample by sample.
  - b. The error signals are calculated for each sample.
  - c. The update of the weights for each sample is calculated according to (12) and (13).
3. The total update for the batch is calculated according to (12) and (13).
4. The weights are updated.
5. The loss function is calculated (7) and the process is repeated from the second step if the loss function is not optimized.

One of the main advantages of the BP algorithm is that it is a very general algorithm that can be applied for all kinds of problems, usually with good results. However, it does have some drawbacks, as it is slow to converge. The loss function commonly has many variables due to the large number of neurons, so the optimization problem is very slow to find the global minimum of the function or a local minimum with a similar value that provides good results.

### 2.3. DASP: Durham Adaptive Optics Simulation Platform

The datasets used along this research were made using the Durham Adaptive optics Simulation Platform (DASP) simulator [14]. The DASP simulator allows the user to make simulations of several AO System as single-conjugated AO, multi-object AO, multi-conjugated AO, etc. where all the parameters of the system and the characteristics of the turbulence can be modified.

The atmospheric turbulence is generated according to the Kolmogorov model implemented with Monte-Carlo simulations [15]. DASP generates different turbulence layers according to the setup previously chosen and simulates the propagation of the light emitted by the celestial body through them. The celestial body can be an artificial guide star, a natural guide star, more than one star as an asterism, a combination of the previous ones or an extended object as the Sun to simulate diurnal observations.

The Kolmogorov model approximates the performance of the atmospheric turbulence by dividing it in several layers and each one being characterized by some parameters as the Fried coherence length ( $r_0$ ), the velocity and the wind direction of the layer, etc. The main parameter is the  $r_0$ , that in physics terms corresponds with the diameter of the pupil of a telescope that, in the absence of turbulence, offers the same resolving power as a large telescope in the presence of it. The  $r_0$  value measures the effect of the atmospheric

turbulence and is usually used as a measure of the turbulence's intensity. It is measured in centimeters corresponding low values with high intensity turbulence. A normal day for observation would be represented by  $r_0$  values higher than 15 cm, being values between 10 and 15 cm a bad day for observation and lower than that a very bad day, probably a stormy day where observations would not be made.

Multiple parameters can be modified by the user in the DASP simulator, as the number of guide stars, the WFSs, the DMs, if the system works in close loop or not, several parameters of the turbulence as the Fried's coherence length  $r_0$ , etc. Further, the platform allows the user to extract much information of the simulation, as the profile of the atmospheric turbulence (for each layer or in common as the sum of all the layers), the information received by the WFS (the image, the position of the centroids, the slopes values ...) and much information of the reconstruction, as the voltages or the positions of the DMs. The least-square method is employed by the platform to make the reconstructions.

In the work presented, the platform was used to obtain the inputs and the desired outputs for each kind of ANN. To make a realistic comparison, the train and test datasets for each kind of neural network were obtained from the same simulation. Therefore, the Zernike's coefficients used as output of the MLP and the CNN corresponds with the approximation of the turbulence's profile phase used in the FCN case.

#### 2.4. Experimental Setup

The experiments were performed on a computer running on Ubuntu LTS 14.04.3, with an Intel Xeon CPU E5-1650 v3 @ 3.50 GHz, 128 Gb DDR4 memory, Nvidia GeForce GTX TitanX, and SSD hard drive. Python was the computer language used, the same as DASP is implemented. Keras was the framework used for this research, created by Google, this framework implemented in Python also allows an easy multi-GPU support.

The simulations were performed trying to replicate a real SCAO System, with a telescope of 4.2 m of pupil diameter. The SH used has  $10 \times 10$  subapertures with a CCD of  $8 \times 8$  pixels each subaperture. The profile of the phase turbulence is generated by DASP with  $80 \times 80$  pixels of size, like the one show in the Figure 2.



Figure 2 Image of the profile of the phase turbulence randomly extracted from the train set used in the research.

For the generation of the dataset, a high number of simulations were made where the images received by the SH, the values of the centroids, the image of the profile of the turbulence's phase and the first 153 Zernike's coefficient were recorded. In particular, the train dataset consisted of 800,000 images, having 100 samples of each simulation. The simulations were performance with a turbulence layer varying its height from 0 to 20 km of height in steps of 50 m and the  $r_0$  value varying from 7 to 17 cm in steps of 0.5 cm. The test dataset was much smaller, being the validation dataset formed by 4000 samples, having 50 samples of each simulation. The height of the turbulence layers was varied from 0 to 20 km in steps of 2 km and the  $r_0$  value from 8 to 16 cm in steps of 1 cm.

To measure the quality of the reconstructions made, 3 test set were simulated with 4000 images each one. In all of them, the atmospheric turbulence consisted in a turbulence layer height increased from 0 m to 20 km in steps of 500 m, 100 different samples of each situation were made. The difference between the dataset is its  $r_0$  value; it was fixed over each set with 8 cm, 10 cm and 12 cm respectively. These three data sets allow us to know how the RS behaves in very bad situations ( $r_0 = 8$  cm), bad situations ( $r_0 = 10$  cm) and in a bad day for observations ( $r_0 = 12$  cm). In a real telescope, observations will not be made in situations as the two first cases, and a normal day of observation correspond to  $r_0$  values around 16 cm, so we are trying or RS in extremely bad situations.

The quality of the reconstructions obtained by the RS are showed in the next section in terms of residual RMSE wavefront error (WFE). For a better comparison between the different topologies, the computational time needed to reconstruct all the samples of the test dataset was measured.

The RMSE WFE of a turbulence phase represents the statistical deviation from the perfect reference sphere and it is expressed in wavelength units.

$$\text{RMSE WFE} = \frac{2\pi}{\lambda} \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N}} \quad (14)$$

Being  $x_i$  each pixel of the turbulence phase,  $\bar{x}$  the mean value over all the pixels,  $N$  the total number of pixels and  $\lambda$  the wavelength of light source, for this research  $\lambda = 640$  nm. To measure the errors made in the reconstructions the residual RMS WFE was employed.

$$\text{residual WFE} = \frac{2\pi}{\lambda} \sqrt{\frac{\sum_{i=1}^N (x_i - y_i)^2}{N}} \quad (15)$$

where  $x_i$  are the pixels of the original turbulence phase and  $y_i$  are the pixels of the reconstructed turbulence phase. In the case of a perfect reconstruction, the surface resulting from the difference between each pixel of the original phase and the reconstructed one should have a zero RMS WFE.

The final topologies selected for each kind of ANN were chosen after the checking of several models. Obviously, for each case, the one which obtained the lowest error over the samples of the test sets was selected to comment.

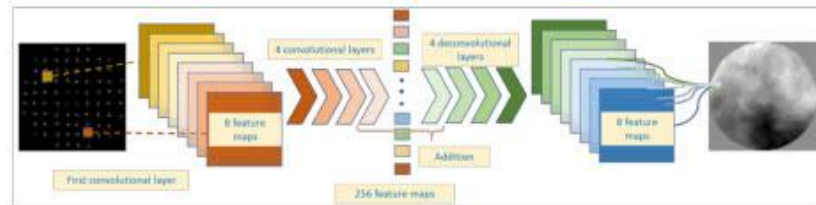
#### 2.4.1. MLP Model

The chosen MLP topology consisted in the input layer, with the centroids position from the SH, so it had the same number of neurons as the number of values given by the SH, 160 neurons. The SH only gives 160 centroids positions, that corresponds with 80 subapertures of the 100 in total; the subapertures of the corners and of the SH's center do not give information of the centroids since they barely get light (see Figure 1). Then, the information is passed through a hidden layer with 1500 neurons before the 153 neurons of the output layer that give the Zernike's coefficients.

#### 2.4.2. CNN Model

The CNN model chose for the experiment consisted in a convolutional block of 5 convolutional layers where the input images are reduced from 80 pixels by side to 256 features maps of 5 pixels by side. The reduction is made by expanding the strides of the convolutional layers, so pooling layers were not added. Each convolutional layer is formed by 8, 32, 64, 128 and 256 filters respectively, with strides selected to halve the input image size; except for the second layers, which let image shape unchanged. A schematic representation of the CNN model can be found in Figure 3.





**Figure 3.** Schematic representation of the topology of the FCN used in the research. An image is set as an input; through the convolutional block, kernels are applied to obtain feature maps. After processing with several convolutional layers, recovering of the information is performed with deconvolutions and adding from previous characteristics up to the desired output.

Then, after the convolutional block, the 256 filters with  $5 \times 5$  pixels of size are passed to a hidden layer composed by 1200 neurons before being the output layer return the 153 coefficients of the Zernike's polynomials.

As in the MLP case, all the layers have the selu function as activation function. The ANN was trained for 113 epochs with padding added.

#### 2.4.3. FCN Model

The FCN consist in two main blocks, a convolutional block and a deconvolutional block (see Figure 3). The convolutional block consists in the same as in the previous case, its output of 256 features maps of size  $5 \times 5$  pixels is given to 5 deconvolutional layers. As in the previous block, the size of the outputs is modified by having strides on the deconvolutional layers of size  $2 \times 2$  instead of using transpose pooling layers. All the convolutional transpose layers double the size of the input given except the last one, which receives 8 filters of size  $80 \times 80$  pixels and returns the final output of the ANN, the profile of the turbulence image of size  $80 \times 80$ . The rest of deconvolutional layers are formed by 128, 64, 32 and 8 filters.

An additional layer is added between the first and the second deconvolutional layer. The output of the first transpose layer is added with the output of the penultimate convolutional layer, as both outputs have the same number of filters and size,  $10 \times 10$  pixels.

As in the other ANNs, the selu activation function was the chosen one. Padding was added both in the convolutional block as in the other one. The FCN was trained for 115 epochs

### 3. Results

This section is divided by four subheadings, where the results for the MLP, the CNN and the FCN model are showed separately. The last one consists in an especial test made with the FCN.

#### 3.1. MLP Model

Table 1 shows the results obtained in terms of residual WFE and relative error for the MLP model. The ANN receives the positions of the centroids measured by the SH for each sample, given as output the first 153 coefficients of the Zernike's polynomials for the turbulence of that moment. Then, the profile of the turbulence phase is reconstructed from the coefficients. It is then compared in terms of the WFE with the one reconstructed from the correct coefficients obtained from the simulation by DASP.

As it has been explained, the network was applied in three test sets where the  $r_0$  value was fixed for each one; for each combination, a total of 3000 samples were taken into consideration:



**Table 1.** Results obtained for the MLP model. Each value is obtained as the mean value over the 3000 samples of the test dataset.

$r_0$ Value (cm)	Residual RMSE WFE (nm)	Relative Residual RMSE WFE (%)	Original RMSE WFE (nm)	Network RMSE WFE (nm)
$r_0 = 8$	100.84	28	357.53	289.28
$r_0 = 10$	90.65	30	297.43	230.20
$r_0 = 12$	84.54	33	255.67	187.42

Further, the average of the computational time per sample is showed in the Table 2 for each of the cases

**Table 2.** Means of the computational time needed by the MLP model per sample for each test dataset.

Dataset	Computational Time per Sample (ns)
$r_0 = 8$ cm	230
$r_0 = 10$ cm	238
$r_0 = 12$ cm	243

### 3.2. CNN Model

The results obtained for the convolutional neural network are showed in Table 3, with the corresponding computational time needed in Table 4. In this case, the input of the network were directly the images received on each subaperture of the SH, giving as output the first 153 coefficients of the Zernike’s polynomials; for each combination, a total of 3000 samples were taken into consideration.

**Table 3.** Results obtained for the CNN model. Each value is obtained as the mean value over the 3000 samples of the test dataset.

$r_0$ Value (cm)	Residual RMSE WFE (nm)	Relative Residual RMSE WFE (%)	Original RMSE WFE (nm)	Network RMSE WFE (nm)
$r_0 = 8$	86.58	24	357.53	299.47
$r_0 = 10$	104.91	35	297.43	208.81
$r_0 = 12$	112.05	44	255.67	155.84

**Table 4.** Means of the computational time needed by the MLP model per sample for each test dataset.

Dataset	Computational Time per Sample (ns)
$r_0 = 8$ cm	575
$r_0 = 10$ cm	600
$r_0 = 12$ cm	575

As in the MLP case, from the Zernike’s coefficient the image of the profile of the turbulence phase is reconstructed to calculate de residual WFE made by the ANN.

### 3.3. FCN Model

In this case the ANN receives as inputs the images received by the subapertures of the SH given directly as output the image of the profile of the turbulence phase. Therefore, the residual WFE showed in the results corresponds straightaway with the comparison between the real profile of the turbulence phase simulated by DASP and the one predicted by the FCN. The results are shown in Table 5; for each combination, a total of 3000 samples were taken into consideration.

**Table 5.** Results obtained for the FCN model. Each value is obtained as the mean value over the 3000 samples of the test dataset. In this case, the output of the ANN is directly compared with the simulated image of the phase’s profile by DASP.

$r_0$ Value (cm)	Residual RMSE WFE (nm)	Relative Residual RMSE WFE (%)	Original RMSE WFE (nm)	Network RMSE WFE (nm)
$r_0 = 8$	253.63	19	1310.93	1251.85
$r_0 = 10$	191.50	18	1088.87	1058.32
$r_0 = 12$	160.94	17	935.07	915.71

In Table 6 the computational time needed per sample by the FCN is showed.

**Table 6.** Means of the computational time needed by the FCN model per sample for each test dataset.

Dataset	Computational Time per Sample (ns)
$r_0 = 8$ cm	675
$r_0 = 10$ cm	700
$r_0 = 12$ cm	678

### 3.4. FCN Model for Profile Phase Approximation

An extra text was made to compare the performance of the FCN due to the difference in the output given by each kind of ANN. In the case of the FCN model, its outputs contain much more information (an image of  $80 \times 80$  pixels) than the others, as they only calculate 153 values corresponding with the 153 first Zernike’s coefficients; for each combination, a total of 3000 samples were taken into consideration. In this subsection, the same FCN topology as in Section 3.3 was tested but, the output given corresponds with the profile of the phase calculated from the first 153 Zernike’s modes simulated by DASP. Therefore, for that test, the same information is given by the FCN in comparison with the others ANN models.

The phase was reconstructed from the same test datasets as in previous chapters and then it is compared with the output of the FCN in terms of WFE and similarity. The results are shown in Table 7 while the computational times needed are shown in Table 8.

**Table 7.** Results obtained for the FCN model, when reconstructing the approximated profile of the phase. Each value is obtained as the mean value over the 3000 samples of the test dataset.

$r_0$ Value (cm)	Residual RMSE WFE (nm)	Relative Residual RMSE WFE (%)	Original RMSE WFE (nm)	Network RMSE WFE (nm)
$r_0 = 8$	146.68	37	394.19	302.52
$r_0 = 10$	177.23	44	395.21	256.69
$r_0 = 12$	203.72	52	394.08	224.09

**Table 8.** Means of the computational time needed by the FCN model per sample for each test dataset.

Dataset	Computational Time per Sample (ns)
$r_0 = 8$ cm	675
$r_0 = 10$ cm	670
$r_0 = 12$ cm	675

## 4. Discussion

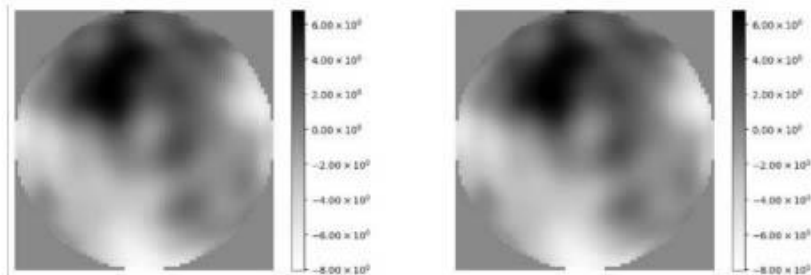
The results of different kinds of ANN models have been compared for the same problem in order to determine which one could be the most profitable model for night SCAO reconstructions.

In terms of the computational time needed to make the reconstructions, the results showed how the MLP is the quickest model, being approximately 40% quicker than the

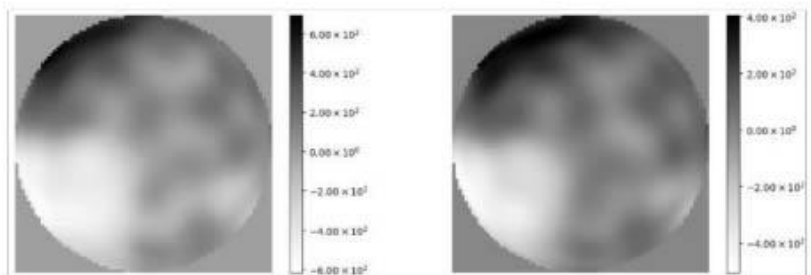
CNN model and the double than FCN ones. That result could be expected as the problem resolved by the MLP is the simplest one, passing from an input of 160 values (the positions of the not darkened subapertures centroids) to an output of 153 values. Therefore, it does not have to interpret an image with the corresponding waste of time. The CNN corresponds with the intermediate case when talking of computational time needed since, although it interprets the whole image of the subapertures, it does not generate a new image as output, a fact that the FCN does.

The other term to analyze is the quality of the reconstructions made by each of the models. For this analysis a separation between the different cases is needed since it is not the same problem reconstructing an approximation of the turbulence phase with the 153 first Zernike's mode than all the information contained in the turbulence phase. Therefore, the results showed in the Sections 3.1, 3.2 and 3.4 are firstly compared.

When the objective is to obtain an approximation of the information contained in the profile of the turbulence phase, the CNN showed the best results of the three compared. For the most turbulence case tested with an  $r_0$  value of 8 cm, the CNN commits a mean relative WFE of 24% being a 44% in the cases of lowest turbulence. An example randomly chosen over the 4000 samples of the test set is showed in Figures 4 and 5. The result is closely followed by the MLP model, that is able to achieve a mean relative WFE of 28% in the highest turbulence case and a 33% in the lowest one. Therefore, in cases of soft turbulences, the MLP achieves the lowest errors. The worst model in this kind of reconstruction is clearly the FCN one, as it achieves a 37% of relative error in the most turbulence case being increased to a 52% in the most turbulence. An example of that is showed in Figure 6.

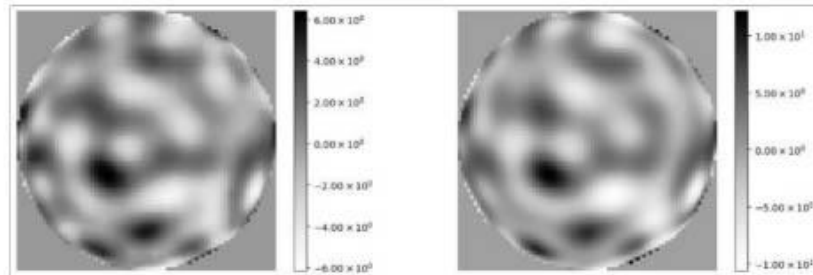


**Figure 4.** Comparison of a sample of the test set with  $r_0$  of 8 cm for the CNN model. On the left, the original reconstructed turbulence phase and, on the right, the one reconstructed by the CNN. With a naked eye both phases' profiles are very similar.



**Figure 5.** Comparison of a sample of the test set with  $r_0$  of 12 cm for the MLP model. On the left, the original reconstructed turbulence phase and, on the right, the one reconstructed with the Zernike's coefficients obtained by the MLP. Despite the error in the numerical values, with a naked eye both phases' profiles are very similar.



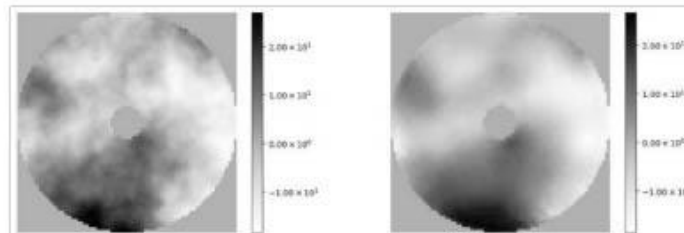


**Figure 6.** Comparison of a sample of the test set with  $r_0$  of 10 cm for the FCN model for the approximation of the turbulence phase. On the left, the original reconstructed turbulence phase and, on the right, the one obtained by the FCN. In this case, although the images are quite similar, the difference in the values are higher as the tonality differs between them.

It is important to remark that all the error committed is due to the reconstruction made by the ANNs and not to the approximation. The images that are compared to are obtained from the 153 Zernike's modes simulated by DASP. They do not correspond directly with the original turbulence phase simulated.

For these tests, both the CNN and the FCN have much information on their inputs in comparison with the MLP, to obtain an output that corresponds with an approximation made with not a lot of information. That fact could make that the convolutional block (that it's the same for both of them) is not able to extract the main characteristics to make relationship with the phase. That is observed especially in the FCN case, where the ANN has to extract characteristics from a lot information to obtain an extended output (an image of  $80 \times 80$  pixels) that contains a little of information. In that point resides the amount of error obtained by the FCN.

These three cases differ with the result showed in the Section 3.3 (example in Figure 7), where the FCN recovers the image of the phase profile that contains all the information possible. In this situation, the same FCN model achieves only a 19% of relative WFE for the  $r_0 = 8$  cm test and a 17% in the  $r_0 = 12$  cm case. Therefore, the FCN is able to relation the information received with the output when the last one contains enough information.



**Figure 7.** Comparison of a sample of the test set with  $r_0$  of 8 cm for the FCN model. On the left, the original turbulence phase simulated and, on the right, the one obtained by the FCN. In this case, the values between the images are similar corresponding with lower WFE seen in the results.

Furthermore, a change on the trend with the  $r_0$  value is observed between the first three cases analyzed and the last one. When the output contains less information, the error committed by the ANN increases when the turbulence intensity decreases, since for lights turbulences making relations among the information received and the main characteristics of the turbulence is more difficult. That is not the situation of the FCN, as the output contains enough information in all the cases to make the relations and the trend shows that the error decreases with the turbulence intensity, probably due to the fact that the phase of the turbulence phase needs less precision being the aberrations longer and more continuous.

## 5. Conclusions

In this paper, a comparison between the most used kinds of ANNs has been made, for one of the nocturnal AO configurations, the SCAO configuration. The research consists of a proof of concept in order to obtain a strategy for which type of neural network could fit the best in AO problems, to determine which one could be the most profitable to develop for future AO systems in next generation telescopes that are already under construction, for example, for the case of solar AO.

In AO, the most important parameters of a RS are such the computational time needed as the errors made in the reconstruction. The comparison is made in both terms, giving different results according to each factor.

In terms of computational time, the MLP model achieves up to three times faster computational times than the CNN, being even higher in the FCN case. The MLP do not have to process images, only arrays of numbers, so that is an expected result. However, in the comparison the time needed by the SH to calculate centroids of each subaperture is not being considered. The operation consists in correlation between the light received by the CCD pixels, not involving a high computational cost. Due to the simplicity of the operations that are made by MLPs compared to the other models, their reconstructions should continue to be the fastest ones.

On the other hand, the WFE errors of the reconstructions show another trend between the different models. When an approximation of the phase by 153 Zernike's modes is reconstructed, the CNN shows results that fit the most with the original phases when high turbulence intensities are present. The MLP model achieves similar results but the WFE is a bit higher. However, the less intensity of turbulence there is, the lower the WFE obtained by the MLP compared with the CNN one. Therefore, the CNN takes advantage of having all the image received by the SH as input when it is very aberrated by the turbulence, as more information of the last one is contained in the image. If the aberrations are lower, CNN will not be able to get lower WFE than the MLP.

This trend could be modified in future developments in Solar AO, as the images receives by the SH consists in a region of the sun, instead of a blurred point from the source of the light. Those images contain much more information even in low turbulence cases (as all the region of the sun is aberrated instead of a point). According to the test made along this research, the use of CNN when the SH's subapertures receives images with a wide amount of information achieves better results than the MLP.

Finally, the results obtained by the FCN are presented, where they show high values of WFE when the objective is to reconstruct an approximation of the turbulence phase based on 153 Zernike's modes, as in the previous cases. The FCN achieves 37% of relative error in high intensity turbulences, increasing the relative error in low intensities ones. In addition, another scenario was considered. This new experiment contains much more information of the turbulence than the approximations of the previous comparisons and, for this case, the FCN shows the best results of all the test made, achieving only a 19% of relative WFE for high turbulences. This error decreases until 15% in lower turbulence cases, much better than any other model.

From the last result it is possible to conclude that the FCN is the model that best fits to the AO system, as it reconstructs much more information than any other model and giving the lower relative error. Nevertheless, the objective of an RS system is to give the necessary turbulence information to a DM to correct the aberrations. The information is passed to a DM as an array of values, typically the needed voltages of the actuators. Anyway, whatever the format of the values, the result objective of an RS approximates the turbulence with some values, so the output of the FCN, despite having more information than any other ANN model, will have to be reduced for being useful. This process will increase the computational time needed by the RS but the error will be reduced, as in approximations the output of the FCN will be more similar to the original phase than in a completely reconstructed image with all the details.

Possible improvements and future work from this research are the improvement of ANNs models to decrease the actual error, as well as performing this study in other scenarios such as solar AO systems, to check if the expected performance of both ANNs with convolutional blocks would be improved working with extended images in the SH, with the corresponding challenges of the implementation of ANN models in these systems.

**Author Contributions:** F.G.R. conceived and designed the study. S.L.S.G. and F.G.R. contributed to the production of the models and analysis of the results. F.G.R. and E.D.A. performed the comparisons and the preparation of the manuscript. J.D.S. and E.S.L. contributed with interpretation of the results and with the preparation of the final version of the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Spanish Economy and Competitiveness Ministry with Grant Number AYA2017-89121-P.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author. The data are not publicly available due to its large size.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Lardièrre, O.; Andersen, D.; Blain, C.; Bradley, C.; Gamroth, D.; Jackson, K.; Lach, P.; Nash, R.; Venn, K.; Véran, J.-P.; et al. Multi-object adaptive optics on-sky results with Raven. In *Adaptive Optics Systems IV*; 2014; Volume 9148, p. 91481G. Available online: <https://doi.org/10.1117/12.2055480> (accessed on 20 May 2021).
- Tallon, M.; Foy, R. Adaptive telescope with laser probe-Isoplanatism and cone effect. *Astron. Astrophys.* **1990**, *235*, 549–557.
- Roggemann, M.C. Optical performance of fully and partially compensated adaptive optics systems using least-squares and minimum variance phase reconstructors. *Comput. Electr. Eng.* **1992**, *18*, 451–466. [[CrossRef](#)]
- Platt, B.C.; Shack, R. History and principles of Shack-Hartmann wavefront sensing. *J. Refract. Surg.* **2001**, *17*, S573–S577. [[CrossRef](#)] [[PubMed](#)]
- Tyson, R. *Principles of Adaptive Optics*; CRC Press: Boca Raton, FL, USA, 2010.
- Reddy, S.S.; Momoh, J.A. Short term electrical load forecasting using back propagation neural networks. In Proceedings of the 2014 North American Power Symposium (NAPS), Pullman, WA, USA, 7–9 September 2014; pp. 1–6.
- Reddy, S.S.; Jung, C.-M.; Seog, K.J. Day-ahead electricity price forecasting using back propagation neural networks and weighted least square technique. *Front. Energy* **2016**, *10*, 105–113. [[CrossRef](#)]
- Sanchez Lasheras, F.; Ordóñez, C.; Roca-Pardiñas, J.; de Cos Juez, F.J. Real-time tomographic reconstructor based on convolutional neural networks for solar observation. *Math. Methods Appl. Sci.* **2020**, *43*, 8032–8041. [[CrossRef](#)]
- Goldberg, Y. Neural network methods for natural language processing. *Synth. Lect. Hum. Lang. Technol.* **2017**, *10*, 1–309. [[CrossRef](#)]
- Ahmed, S.A.; Dey, S.; Sarma, K.K. Image texture classification using artificial neural network (ANN). In Proceedings of the 2011 2nd National Conference on Emerging Trends and Applications in Computer Science, Shillong, India, 4–5 March 2011; pp. 1–4.
- Osborn, J.; Guzman, D.; de Cos Juez, F.J.; Basden, A.G.; Morris, T.J.; Gendron, E.; Butterley, T.; Myers, R.M.; Guesalaga, A.; Sánchez Lasheras, F.; et al. Open-loop tomography with artificial neural networks on CANARY: On-sky results. *Mon. Not. R. Astron. Soc.* **2014**, *441*, 2508–2514. [[CrossRef](#)]
- García Riesgo, F.; Suárez Gómez, S.L.; Sánchez Lasheras, F.; González Gutiérrez, C.; Peñalver San Cristóbal, C.; de Cos Juez, F.J. Convolutional CARMEN: Tomographic Reconstruction for Night Observation. In Proceedings of the International Conference on Hybrid Artificial Intelligence Systems, Gijón, Spain, 11–13 November 2019; pp. 335–345.
- de Cos Juez, F.J.; Lasheras, F.S.; Roqueñí, N.; Osborn, J. An ANN-based smart tomographic reconstructor in a dynamic environment. *Sensors* **2012**, *12*, 8895–8911. [[CrossRef](#)] [[PubMed](#)]
- Suárez Gómez, S.L.; González-Gutiérrez, C.; Díez Alonso, E.; Santos Rodríguez, J.D.; Sánchez Rodríguez, M.L.; Morris, T.; Osborn, J.; Basden, A.; Bonavera, L.; González-Nuevo González, J.; et al. Experience with Artificial Neural Networks applied in Multi-Object Adaptive Optics. *Publ. Astron. Soc. Pacific* **2019**, *131*, 108012. [[CrossRef](#)]
- Riesgo, F.G.; Gómez, S.L.S.; Rodríguez, J.D.S.; Gutiérrez, C.G.; Alonso, E.D.; Rodríguez, F.J.L.; Fernández, P.R.; Bonavera, L.; Menéndez, S.d.C.F.; Juez, F.J.D.C. Early Fully-Convolutional Approach to Wavefront Imaging on Solar Adaptive Optics Simulations. In Proceedings of the International Conference on Hybrid Artificial Intelligence Systems, Gijón, Spain, 11–13 November 2020; pp. 674–685.
- Gómez, S.L.S.; Gutiérrez, C.G.; Rodríguez, J.D.S.; Rodríguez, M.L.S.; Lasheras, F.S.; de Cos Juez, F.J. Analysing the performance of a tomographic reconstructor with different neural networks frameworks. In *Advances in Intelligent Systems and Computing*; Springer: Berlin/Heidelberg, Germany, 2017; Volume 557, pp. 1051–1060.
- Zilberman, A.; Golbraikh, E.; Kopeika, N.S. Propagation of electromagnetic waves in Kolmogorov and non-Kolmogorov atmospheric turbulence: Three-layer altitude model. *Appl. Opt.* **2008**, *47*, 6385–6391. [[CrossRef](#)] [[PubMed](#)]



18. Osborn, J.; De Cos Juez, F.J.; Guzman, D.; Butterley, T.; Myers, R.; Guesalaga, A.; Laine, J. Using artificial neural networks for open-loop tomography. *Opt. Express* **2012**, *20*, 2420. [CrossRef] [PubMed]
19. Osborn, J.; Guzmán, D.; de Cos Juez, F.J.; Basden, A.G.; Morris, T.J.; Gendron, É.; Butterley, T.; Myers, R.M.; Guesalaga, A.; Lasheras, F.S.; et al. First on-sky results of a neural network based tomographic reconstructor: Carmen on Canary. In *Adaptive Optics Systems IV*; Marchetti, E., Close, L.M., Véran, J.-P., Eds.; International Society for Optics and Photonics: Bellingham, WA, USA, 2014; Volume 9148, p. 91484M.
20. European Southern Observatory. AO Modes. Available online: [https://www.eso.org/sci/facilities/develop/ao/ao\\_modes/.html](https://www.eso.org/sci/facilities/develop/ao/ao_modes/.html) (accessed on 2 April 2021).
21. Sidick, E.; Green, J.J.; Morgan, R.M.; Ohara, C.M.; Redding, D.C. Adaptive cross-correlation algorithm for extended scene Shack-Hartmann wavefront sensing. *Opt. Lett.* **2008**, *33*, 213–215. [CrossRef] [PubMed]
22. Neal, D.R.; Copland, J.; Neal, D.A. Shack-Hartmann wavefront sensor precision and accuracy. *Int. Symp. Opt. Sci. Technol.* **2002**, *4779*, 148.
23. Dwivedi, A.K. Artificial neural network model for effective cancer classification using microarray gene expression data. *Neural Comput. Appl.* **2016**, *1–10*. [CrossRef]
24. Chauvin, Y.; Rumelhart, D.E. *Backpropagation: Theory, Architectures, and Applications*; Psychology Press: Hove, UK, 1995; ISBN 0805812598.
25. Nielsen, M.A. *Neural Networks and Deep Learning*; 2015; Available online: <http://neuralnetworksanddeeplearning.com/> (accessed on 20 May 2021).
26. Kim, Y. Convolutional neural networks for sentence classification. *arXiv* **2014**, arXiv:1408.5882.
27. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
28. Gurney, K. *An Introduction to Neural Networks*; CRC Press: Boca Raton, FL, USA, 1997.
29. Reddy, S.S.; Momoh, J.A. Economic Dispatch using Improved Hopfield Neural Network. In Proceedings of the 2015 North American Power Symposium (NAPS), Charlotte, NC, USA, 4–6 October 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 1–5. [CrossRef]
30. Reddy, S.S.; Jung, C.-M. Short-term load forecasting using artificial neural networks and wavelet transform. *Int. J. Appl. Eng. Res.* **2016**, *11*, 9831–9836.
31. Gulli, A.; Pal, S. *Deep Learning with Keras*; Packt Publishing Ltd.: Birmingham, UK, 2017.
32. Nair, V.; Hinton, G.E. Rectified linear units improve restricted boltzmann machines. In Proceedings of the 27th International Conference on Machine Learning (ICML-10), Haifa, Israel, 21–24 June 2010; pp. 807–814.
33. Keras Special Interest Group Keras Layer Activation Functions. Available online: <https://keras.io/api/layers/activations/> (accessed on 20 May 2021).
34. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef] [PubMed]
35. Mirowski, P.W.; LeCun, Y.; Madhavan, D.; Kuzniecky, R. Comparing SVM and convolutional networks for epileptic seizure prediction from intracranial EEG. In Proceedings of the 2008, MLSP 2008, IEEE Workshop on Machine Learning for Signal Processing, Cancun, Mexico, 16–19 October 2008; pp. 244–249.
36. Nagi, J.; Ducatelle, F.; Di Caro, G.A.; Cireşan, D.; Meier, U.; Giusti, A.; Nagi, F.; Schmidhuber, J.; Gambardella, L.M. Max-pooling convolutional neural networks for vision-based hand gesture recognition. In Proceedings of the 2011 IEEE International Conference on Signal and Image Processing Applications (ICSIPA), Kuala Lumpur, Malaysia, 16–18 November 2011; pp. 342–347.
37. Benvenuto, N.; Piazza, F. On the complex backpropagation algorithm. *IEEE Trans. Signal Process.* **1992**, *40*, 967–969. [CrossRef]
38. Sarz-Molina, A.; del Brío, B. *Redes Neuronales y Sistemas Borrosos*; Rama: Madrid, Spain, 2006; p. 436. ISBN 978-84-7897-743-7.
39. Rumelhart, D.E.; Durbin, R.; Golden, R.; Chauvin, Y. Backpropagation: The basic theory. In *Backpropagation: Theory, Architectures, and Applications*; Lawrence Erlbaum Associates Publishers: Hillsdale, NJ, USA, 1995; pp. 1–34.






## 11.2 FULLY CONVOLUTIONAL APPROACHES FOR NUMERICAL APPROXIMATION OF TURBULENT PHASES IN SOLAR ADAPTIVE OPTICS

**AUTORES: FRANCISCO GARCÍA RIESGO, SERGIO LUIS SUÁREZ GÓMEZ, ENRIQUE DÍEZ ALONSO , CARLOS GONZÁLEZ-GUTIÉRREZ Y JESÚS DANIEL SANTOS RODRÍGUEZ**

Article

# Fully Convolutional Approaches for Numerical Approximation of Turbulent Phases in Solar Adaptive Optics

Francisco García Riesgo<sup>1,2</sup>, Sergio Luis Suárez Gómez<sup>2,3</sup>, Enrique Díez Alonso<sup>2,3</sup>, Carlos González-Gutiérrez<sup>2,4</sup> and Jesús Daniel Santos<sup>1,2,\*</sup> 

<sup>1</sup> Department of Physics, University of Oviedo, 33007 Oviedo, Spain; garciariefrancisco@uniovi.es

<sup>2</sup> Instituto Universitario de Ciencias y Tecnologías Espaciales de Asturias (ICTEA), 33004 Oviedo, Spain; suarezsergio@uniovi.es (S.L.S.G.); diezenrique@uniovi.es (E.D.A.); gonzalezcarlos@uniovi.es (C.G.-G.)

<sup>3</sup> Department of Mathematics, University of Oviedo, 33007 Oviedo, Spain

<sup>4</sup> Computer Sciences Department, University of Oviedo, 33024 Gijón, Spain

\* Correspondence: jdsantos@uniovi.es



**Citation:** Riesgo, F.G.; Gómez, S.L.S.; Alonso, E.D.; González-Gutiérrez, C.; Santos, J.D. Fully Convolutional Approaches for Numerical Approximation of Turbulent Phases in Solar Adaptive Optics. *Mathematics* 2021, 9, 1630. <https://doi.org/10.3390/math9141630>

Academic Editor: Amir Mosavi

Received: 27 May 2021

Accepted: 5 July 2021

Published: 10 July 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** Information on the correlations from solar Shack–Hartmann wavefront sensors is usually used for reconstruction algorithms. However, modern applications of artificial neural networks as adaptive optics reconstruction algorithms allow the use of the full image as an input to the system intended for estimating a correction, avoiding approximations and a loss of information, and obtaining numerical values of those correlations. Although studied for night-time adaptive optics, the solar scenario implies more complexity due to the resolution of the solar images potentially taken. Fully convolutional neural networks were the technique chosen in this research to address this problem. In this work, wavefront phase recovery for adaptive optics correction is addressed, comparing networks that use images from the sensor or images from the correlations as inputs. As a result, this research shows improvements in performance for phase recovery with the image-to-phase approach. For recovering the turbulence of high-altitude layers, up to 93% similarity is reached.

**Keywords:** fully convolutional neural networks; artificial intelligence; artificial neural networks; adaptive optics; solar physics

## 1. Introduction

Artificial intelligence methods have led to major improvement in several fields of the science of engineering [1]. As a mathematical tool to perform numerical approximations, they are broadly used to represent complex physical systems, whose analytical models are way too complex to work with, or even improve other numerical approaches both in terms of time consumption and performance. Among the broad range of techniques considered as artificial intelligence, some of them are used satisfactorily as a solution to cutting-edge problems, such as deterministic artificial intelligence [2], which implies an alternative to stochastic methods [3]. Other options include neural networks, since the variety of architectures of neural networks allows for managing different types of problems, from those that require a numerical approximation as a regression problem to those that are classification problems.

The flexibility of the networks not only relates to what they are expected to obtain but also the manner of the information of the problem being measured. Thus, several types of information may be used to work with neural networks; data vectors are for the simplest models of networks, but two-dimensional arrays, such as images, or even higher dimensional tensors can be used to feed a network.

Artificial neural networks, as learning algorithms [4], considered under a supervised learning process will have their answers verified against true information measurements, usually real data or simulated data, and then adjust their inner parameters to better fit the problem by means of an optimization algorithm. As a consequence, a network can

be considered a malleable model that fits a given problem by solving a multidimensional optimization problem.

While this seems to be a clear advantage, the search for local minima, given a particular architecture, does not have a perfect answer when expecting models to work toward real-time computation. Moreover, this issue becomes larger when considering all the possible hyperparameters that tune the topology and architecture of a neural network. These different approaches to the solution of a problem make it difficult to choose an answer. In particular, the application of neural networks considered in this work for adaptive optics (AO) has the same issue.

In general, AO is a set of techniques related to astronomical telescope imaging, where atmospheric distortions modify the images taken by those telescopes. This is due to the changing nature of the atmosphere, whose turbulences are usually modeled by statistically based models such as those of Von Karman [5] or Kolmogorov [6,7]. Adaptive optics includes a set of techniques and elements that allow for correcting the turbulent effects, from sensing the turbulences with wavefront sensors to correction of the wavefront of the image in deformable mirrors and other processes, such as estimation of the turbulent effects with a reconstruction algorithm.

In recent years, the latest issue is one that has been broadly studied with the application of artificial neural networks [8]. As it was previously mentioned, it is not easy to select an adequate network, and thus this problem was one of the main issues addressed in previous works [9]. An attempt for an answer in the case of nocturnal AO has been deliberated in previous papers, yet a study for the solar situation is necessary [10,11]. Although in previous works, some networks such as multilayer perceptrons (MLPs) and convolutional neural networks (CNNs) were checked [12,13], this work presents improvements through the use of fully convolutional networks (FCNs), comparing the differences between using solar correlations as inputs for the networks and the full image from the solar Shack-Hartmann (SH) wavefront sensors.

This paper is organized as follows. Section 2 includes the artificial intelligence methods described above, as well as some insights about solar adaptive optics modeling and details about the simulation platform. Section 3 shows the results of the selected topologies for the two cases considered, and Section 4 includes a detailed discussion about the obtained results. Finally, in Section 5, our conclusions are put forward.

## 2. Materials and Methods

### 2.1. Artificial Intelligence

Artificial intelligence methods have become, in recent times, fundamental in daily applications for both scientific and engineering purposes. Their key feature is their capability to learn, supervised or unsupervised, from a great amount of data and extract patterns to provide a response to a problem [14]. In this research, artificial intelligence was used as the reconstruction system of a solar AO system, based on the results and previous performance proven for nocturnal AO. For these purposes, artificial neural networks (ANNs) were used. Choosing one type of architecture or another highly depends on the problem to be solved. Usually, multilayer perceptrons (MLPs) are used to process vectorial inputs, or low-dimensional tensors that can be easily vectorized [15]. Convolutional neural networks (CNNs) rely on the use of kernels or filters that are trained to extract and process the most relevant features from an image-like input, as well as for tensor inputs [16]. They usually have a segment where the data are resized to be smaller while maintaining the information, allowing one to vectorize the data and add an MLP segment at the end of the network. Other alternatives include fully convolutional networks (FCNs) [17], which take the full input, and after processing the main characteristics, the deconvolutional layers are set to recover from the characteristics in the shape of a tensor.

In this work, FCNs were tested for phase reconstruction. The ANNs were used with wavefront sensor images of the frames of the Sun and with correlations, in order to check the input data that best fit with the FCNs' performance.



### 2.1.1. MLPs and Backpropagation

A multilayer perceptron is based in the computation of responses of individual computation units of neurons, which are sorted in layers; each of them computes the following calculus [18]:

$$y_j = f \left( \sum_{i=1}^n w_{ij} x_i + b_j \right), \quad (1)$$

where for each neuron  $j$ , the input values from  $x_i = 1 \dots n$  are weighted by the values  $w_{ij}$ , and then a bias  $b_j$  is added. All these computations are then modified by an activation function  $f$ . The outputs are those defined as  $y_j$ .

The activation functions work as a normalizing modification of the output, usually a hyperbolic tangent or sigmoid, whose images are restricted to  $(-1, 1)$  and  $(0, 1)$ , respectively [19]. Other activation functions that are commonly used are rectified linear unit (ReLU), parametric linear unit (PReLU), etc. [20].

Neurons are assigned to layers, which are sorted so all the neurons of each layer can be densely connected to all the neurons of the following layer, which is conducted by means of assigning the weights to those connections. Once the last layer gives an output, it is considered as the output of the network to the inputs given.

At first, the network's weights are set as random (thus, sometimes, they are set following some rules, as some predefined filters); consequently, the network's response to an input would be random, meaning the model must be trained. To estimate the differences between the obtained output and the desired value, a loss function is applied. Usually, measures such as mean squared error (MSE) or RMSE (root mean square error) are applied for regression problems. Cross-correlation is often used for classification purposes [21].

The backpropagation algorithm is set to change the values of the weights according to the information obtained from the loss function over a training set [22,23]. This is then a multivariate optimization problem over the weights as variables, whose dimension depends explicitly on the size of the network. The simplest way to approach this problem is to set a stochastic gradient descent algorithm [24], where for each iteration of the samples from the training set, the weights are updated as follows:

$$w_{ij} = w_{ij} - \frac{\eta \partial C}{\partial w}, \quad (2)$$

where  $C$  is the loss function, and its partial derivative is computed to upgrade the weight values in the direction of the negative gradient. The parameter  $\eta$  is called the learning rate and is used to regulate the magnitude of the steps taken in the direction of the negative gradient.

Other optimizers include the momentum of inertia or several modifications, such as Nesterov (Adam), Adagrad, and Adadelta, to improve the convergence of the method towards local minima [25].

### 2.1.2. Convolutional Neural Networks

For multidimensional inputs, such as those matrixes that represent images, or tensors, vectorization to apply an MLP may not be adequate. In these situations, convolutional layers are used to process the full input, by using kernels or filters [26]. These are set in the convolutional layer, obtaining a map of characteristics from each kernel applied, which is applied by a discrete convolution, given by

$$S(i, j) = (I * K)(i, j) = \sum_m \sum_n I(m, n) K(i - m, j - n), \quad (3)$$

for an image  $I$  and a kernel  $K$ , over  $i, j$ , the sizes of the kernel as a matrix, as well as the matrix  $I \in M(n, m)$ . At first, the values of  $K$  are set as random or as predefined filter values but, over iterations, are intended to be minimized with the same procedure explained for the MLPs, taken as an optimization problem.

Convolution layers also include the application of an activation function. At first glance, characteristic maps maintain the size of the original matrix for each of the characteristic maps, so in order to downsize the data, usually max-pooling is applied [16].

After the desired number of convolutional layers, the processed data are vectorized to be the input of an MLP section.

### 2.1.3. Fully Convolutional Neural Networks

However, if the desired output should be a multidimensional array, other options can be considered. The topology of an FCN allows having these outputs, particularly after the succession of convolutional layers, a section of deconvolutional layers whose calculus consists of the transpose of discrete convolutions over the correspondent kernels or convolution window.

In FCNs, the classification block presented at the end of a common CNN is substituted by the deconvolutional block, allowing the ANN to obtain a new image as output from the several maps of characteristics extracted from the input by the convolutional block; an example of an FCN topology is shown in Figure 1. Its performance consists of a new extended image from small images with the main features of the inputs. Therefore, for this type of ANN, the principal block of the ANN is the convolutional one as in the previous case.

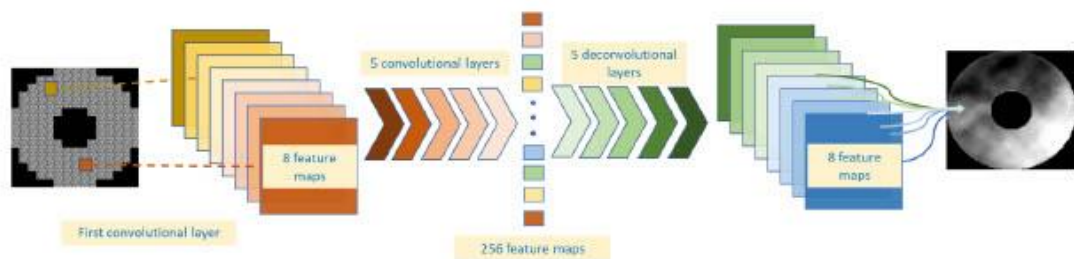


Figure 1. Schematic representation of the FCN used throughout this research. The comparison is made in both cases with the same FCN topology.

One of the main features of FCNs is not just plain processing downsizing through convolutions and then recreating bigger matrixes through deconvolutions, it is the design where features from the convolutional section are then fused into the adequate parts of the deconvolutional section. Thus, this allows passing spatial information deeper to sections where it may have been lost.

FCN is the ANN topology used for this research, as both the input and the output data used for the comparison consist of images. As input data, the image of the wavefront received by the Shack–Hartmann wavefront sensor or the image of the calculated cross-correlations for several moments is given, being the profile of the turbulence phase for each moment of the desired output.

## 2.2. Solar Adaptive Optics

Adaptive optics is the branch of astronomy that corrects and improves the quality of images taken by telescopes. Due to the fluid nature of the atmosphere, the refractive index varies rapidly, meaning any light that passes through it has its wavefront displaced due to the change in velocity of the photons when passing through the different refractive indexes [27].

Kolmogorov models are some of the most used models to model atmospheric turbulences. They are based in the layers of the turbulence, characterizing it through the heights of the layers, their relative forces, winds, the energy generated by external-scale processes that influence and cause phenomena on smaller scales  $l_0$ , large-scale air movements  $L_0$ ,



and Fried's coherence parameter  $r_0$ , which generalizes some of the properties that include aspects such as composition and temperature. The Kolmogorov spectra are given by

$$\Phi_n(K) = 0.033C_n^2 |\vec{K}|^{-11}, \quad (4)$$

where  $C_n$  is the refractive index structure constant.

Fried's coherence length ( $r_0$ ) represents a very useful parameter in astronomy terms as it contains information about the turbulence intensity, the wavelength, and the propagation path for an incoming wavefront. It is defined as follows:

$$r_0 = [0.423k^2 \sec(\gamma) \int_0^L C_n^2(h) \left(\frac{h}{L}\right)^{\frac{5}{3}} dh]^{-\frac{3}{5}}, \quad (5)$$

where  $\gamma$  represents the zenith angle.

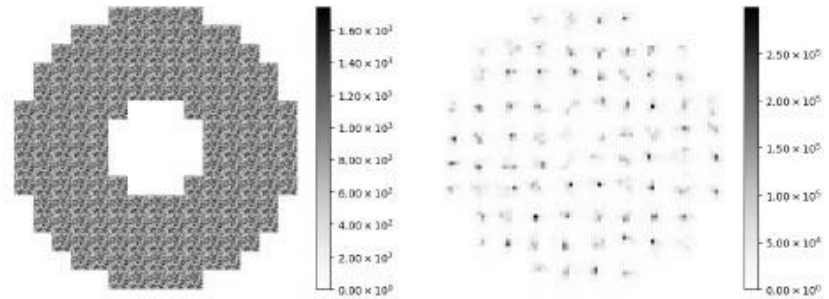
The  $r_0$  value can be defined as the diameter of a pupil where the RMS wavefront aberration value is 1 radian, and it is expressed in length terms, commonly in cm. This means that the lower the  $r_0$  value, the more intense the aberrations. To put this in context, a common astronomical observation day would be represented by an  $r_0$  value higher than 16 cm. If the value is lower than 12 cm, astronomical observations are usually postponed, due to bad turbulence conditions.  $r_0$  values less than or equal to 8 cm correspond to extremely bad atmospheric turbulence, as it could be a stormy day.

To correct those aberrations, mechanical systems to refract the beams of light at the adequate wavelengths return the images to their original shape. Usually, the distortions are measured by Shack–Hartmann wavefront sensors (SHWFSs). In night-time observations, SHWFSs consist of an array of convergent lenses, consequently having as many estimates of the turbulent profile of the atmosphere as lenses in the sensor; as each lens causes the light to converge, the estimation, which is a blurred concentrated spot of light, whose gravity center is called a centroid, corresponds to the deviation from the center of the CCD sensor.

Then, typically a system of AO has some SHWFSs that measure the distortions and a deformable mirror that reflects the adequate wavefront, and they need a reconstruction algorithm that estimates the turbulent profile. For these algorithms, the most traditional approach is least squares (LS), but AI alternatives have been proven successful approaches [28].

For solar observations, however, the situation differs in several ways: First, the energy received by the atmosphere from the Sun directly disturbs the atmosphere, having more turbulent layers as a result of the excess of energy. The reconstruction must take this into account, since one point in time may be completely different to the next, while in night observation, variations are slower as the turbulence is stabler. From all of the turbulent profile, most of the variations occur in the ground layer, also due to the heat radiated by the Earth and the telescope's pupil during daylight hours.

Another major issue, which changes the way of measuring the profile, is the fact that the Sun's proximity to Earth implies that it cannot be considered a point source. Wavefront measurement should be conducted without guide stars. The estimation cannot be carried out, then, with centroids, as the SHWFS receives on each subaperture an extended image, being necessary to calculate the correlations between the images that appear in the subapertures, in order to estimate the variations in the turbulent profiles. An example is shown in Figure 2.



**Figure 2** Information received by the SHWFS in diurnal observation on the left and in night-time observations on the right. In solar observations, due to the source’s proximity, an extended image is received by each subaperture of the SHWFS, saturating all the pixels instead of blurred spots as happens in the night scenario, meaning that the algorithms used in night AO have to be modified in diurnal AO.

The way to calculate the correlations is by using an image from one subaperture as the reference image at the beginning of the observation; thus, the variations detected by sensors are measured by the comparisons with the reference image. From the first images, any can be used as the reference one, but commonly, one of the brightest images should be chosen. A cross-correlation between the “live” image, measurement from the sensor at each individual instant, and the reference image is estimated according to the following [29]:

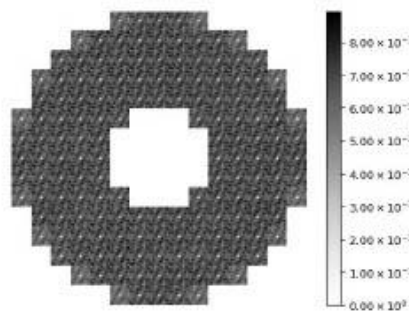
$$CC_{SR}(\Delta) = \int \int I_S(x) \times I_R(x + \Delta) dx \tag{6}$$

where  $I_S(x)$  is the live image and  $I_R(x)$  is the reference image obtained from the SH, where  $x$  denotes the spatial coordinate in the lower contour of the sensor and  $\Delta$  denotes a spatial lag, the result of obtaining the image at a different subaperture.

The actual calculation of the covariance can be conducted by a Fourier transform given by

$$CC_{SR}(\Delta) = F^+ [F^- [I_S(x)] \times F^{-*} [I_R(x)]], \tag{7}$$

where  $F^-$ ,  $F^+$  are direct and inverse Fourier transforms, respectively.  $F^{-*}$  stands for the complex conjugate of a Fourier transform [30]. Results of correlated images from a solar SH are exemplified in Figure 3.



**Figure 3** Image of the cross-correlations calculated by the SHWFS from a randomly chosen sample of the training dataset. A spot is situated approximately in the center of each subaperture that corresponds to the maximum of the CC.

The position of the maximum of the cross-correlation CC is a linear measure of the displacement between the reference image and the given subaperture; therefore, the



aberration of the wavefront will already be identified, even more so when the gradients or inclinations are measured, which will be carried out in the same way as conducted in night adaptive optics from the centroids of each subaperture. Once the values of the coefficients are known, the received wavefront has already been reconstructed, meaning the next step that will be carried out will be to obtain the values of the positions that each DM must occupy to correct the aberration suffered.

### 2.3. Durham Adaptive Optics Simulation Platform (DASP)

The DASP consists of a simulation platform developed by Durham University for night and solar observations [31]. The software allows the user to determine the parameters of the simulation; several AO configurations can be implemented such as SCAO, MCAO, and MOAO. Moreover, in the simulations, the number of components can be selected by the user, such as the number of SHWFSs and the number of DMs. The features of each component can also be chosen, such as the telescope diameter, the number of subapertures of the SHWFS and its number of pixels, or the number of actuators of the DM. Due to the quality of the simulations carried out by the DASP, the platform is currently used in simulations for forthcoming real instruments, such as the Extreme Large Telescope (ELT) or the 12 m Chinese Large Optical Telescope [31].

The DASP simulates the atmospheric turbulence according to Kolmogorov's model, which is implemented by Monte Carlo simulations. Some parameters of this model can be chosen by the user, such as the  $r_0$  value that determines the intensity of the whole turbulence, as well as the number of turbulence layers, the weight of each one over all of the turbulence, their height, and their wind direction and velocity.

A very interesting feature of this platform is that multiple pieces of information of the intermediate processes of the AO system can be saved, allowing for comparisons of the research. For each situation, the information received by the SHWFS, the cross-correlations calculated by the SHWFS (in solar observations), the slope measurements on each subaperture, the turbulence phase, Zernike's coefficients of the turbulence phase, or the voltage for the DM's actuators can be extracted from the software.

For this research, the DASP was used as the tool to generate the datasets needed to train, validate, and test the ANN models. Simulations were conducted according to the parameters shown in the next section, saving the image of the wavefront received by the SHWFS and its corresponding correlations, which are the input of the ANN models. The phase profile for each moment was also saved, being the desired output of our AI reconstruction system. Thus, the dataset to train and test both models was utilized in all simulations to make a realistic comparison.

### 2.4. Experimental Setup

The training of the networks, as well as the simulations, was performed on a computer running on a Ubuntu LTS 18.04.4, with an Intel Xeon CPU W-3235 @3.3Ghz, 512 Gb DDR4 memory, Nvidia GeForce RTX 2080 Ti, and SSD hard drive. Python was the computer language used, along with TensorFlow and Keras. The Gregor Solar Telescope [32] configuration was replicated in the simulations, with a 1.5 m pupil diameter, working in a Solar SCAO configuration. As the objective is to compare the behavior of the FCNs depending on the input data, all the simulations were conducted over the same region of the Sun.

The training dataset consisted of 80,000 simulated images of the information received by the SHWFS and the calculated correlations with their corresponding profile of the turbulence phase, which is the desired output of both FCN models. All the simulations related to two turbulence layers, the first one being situated at 0 m height with a weight of 10% of the turbulence, and the second one being another turbulent layer whose height varied from 1 to 20,000 m in steps of 100 m each for the remaining 90%. The  $r_0$  values of the turbulence varied from those corresponding to extremely turbulent scenarios to values from a calm observation day. In numerical terms, they range from 8 to 16 cm of  $r_0$  in steps of a cm. For each situation, 50 consecutive simulations were conducted, resulting in a total

of 80,000 simulations. To validate the models during the learning process, a test dataset was constructed with 3000 situations where both the height and the  $r_0$  value varied between the same terms, but in higher steps.

The ANN models were tested with 12 additional test datasets that consisted of randomly generated turbulence phases unknown to the networks. Each dataset corresponded to a fixed  $r_0$  value from 7 to 19 cm and a determinate range of the height for the atmospheric turbulence. One half consisted of turbulence between 0 and 3000 m that represents the ground layer, and the other half between 9000 and 12,000 m. The datasets allowed analyzing the performance of the FCNs with several turbulence situations and comparing which input data fit better according to the situation. Each dataset was composed of 600 samples.

The ANNs' performance was analyzed by visual comparison and analytical error. For the second case, the residual wavefront error (WFE) was employed:

$$\text{residual WFE} = \frac{2\pi}{\lambda} \sqrt{\frac{\sum_{i=1}^N (x_i - y_i)^2}{N}} \quad (8)$$

where  $x_i$  represents the turbulence phase's pixels of the FCN's output and  $y_i$  is those of the desired output. This error can also be measured in percentage by dividing (8) by the RMSE WFE of the simulated profile phase.

#### Selected FCN Topology

To make a realistic comparison, the same network topology was employed for both types of input data considered. The topology consisted of a convolutional block with 6 convolutional layers, where the input shape of the image of  $360 \times 360$  was reduced, applying strides of sizes 1, 2, or 3 pixels in both directions, to 256 feature maps of size  $5 \times 5$ . They were passed to a deconvolutional block formed by 6 deconvolutional layers that, with the same size of strides as those from the previous block, obtained, as final output, an image of size  $90 \times 90$  pixels, as the simulated turbulence's phase. Padding was added in all the layers, and the ANN was trained with Nesterov's gradient descent optimization algorithm, with a 0.02 learning rate and a momentum value of 0.7. All the layers applied the same activation function, the hyperbolic tangent. A schematic representation of the selected FCN can be seen in Figure 1.

### 3. Results

This section presents two different scenarios showing the performance of the FCN when comparing the two types of input images in scenarios of low-height turbulences, trying to replicate the ground layer, in one subsection, and high-height turbulences in the other.

The data shown in the next table correspond to the mean of the residual WFE of the reconstruction, the mean WFE residual of the phase's profile over the 600 samples of each test, and the mean value of the similarity of the reconstructions. The similarity is calculated as the mean value of the quotient of the RMSE WFE of the reconstruction between the simulated one for each sample.

The recall time needed per sample by the FCN to reconstruct the turbulence's profile is also shown.

#### 3.1. Ground Layer Turbulences

The errors committed in the reconstruction of ground layer turbulences by the FCNs are shown in Table 1. All the test datasets consist of 600 images of ground layers between 0 m and 3 km of height; therefore, the results given consist of the mean value over all the samples of the test. Then, several figures of randomly chosen samples of each test dataset are shown for a visual comparison.

**Table 1.** Results obtained for the ground layer reconstructions. WFE net refers to the RMSE WFE calculated over the images given by each FCN as output, while WFE residual refers to the RMSE WFE of the image that is obtained as the difference between the output of one ANN and the original simulated image. Both terms are measured in radians (rad). SHWFS refers to the ANN that uses as input the information received by the subapertures of the sensor, while CC refers to the ANN that uses the cross-correlations (CC) as inputs. The time column shows the mean time over the dataset per sample for each ANN, measured in milliseconds (ms).

$r_0$ (cm)	WFE Net (rad)		WFE Residual (rad)		Similarity (%)		Time (ms)	
	SHWFS	CC	SHWFS	CC	SHWFS	CC	SHWFS	CC
7	7.29	7.22	2.55	2.85	84.3	83.2	4.53	4.33
9	5.42	5.41	1.72	1.99	82.2	80.4	4.45	4.38
11	5.18	5.12	1.51	1.78	85.7	83.8	4.48	4.33
13	4.47	4.49	1.28	1.50	81.1	77.1	4.47	4.35
15	3.82	3.83	1.11	1.31	85.9	83.6	4.41	4.38
17	3.50	3.49	1.05	1.20	72.9	69.7	4.50	4.37

In Figure 4, several examples of the reconstructions carried out by both FCNs with their corresponding original image are shown for a visual comparison. There is an example for each used test dataset.

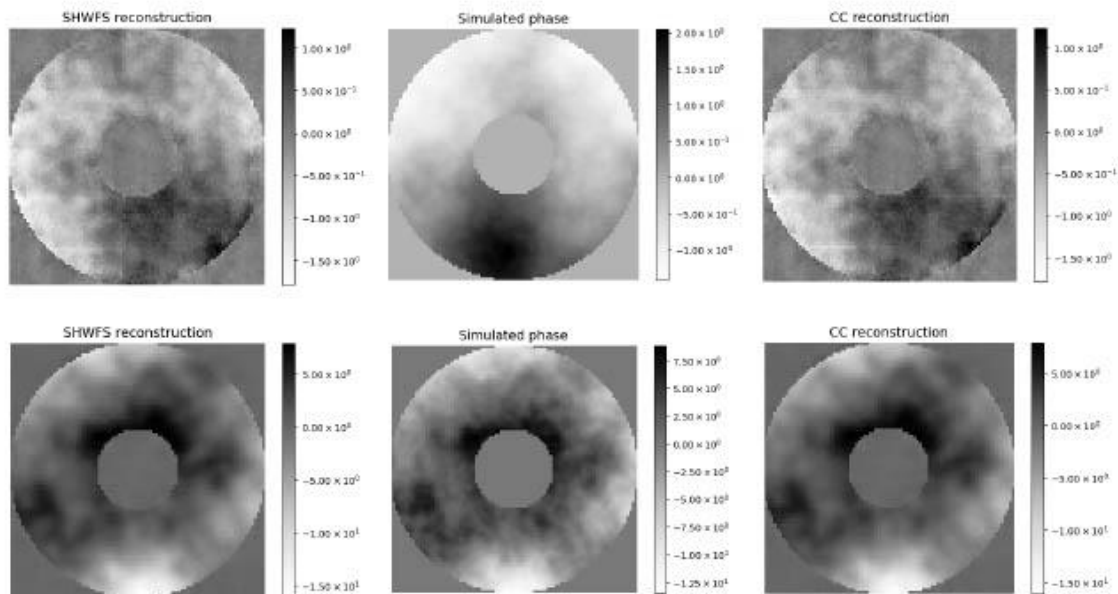
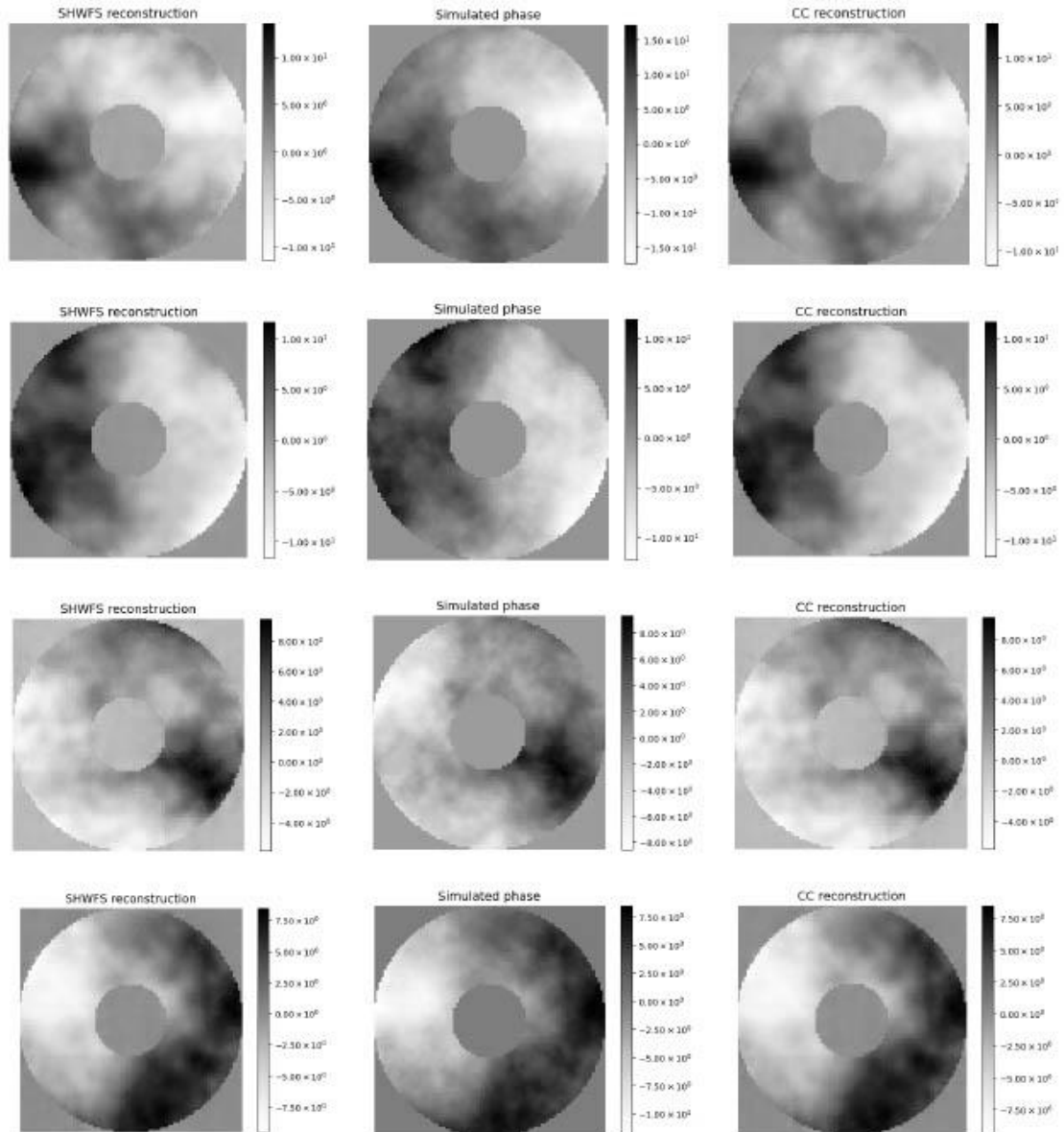


Figure 4. Cont.





**Figure 4.** Comparison of the reconstructed phases for ground layer turbulences with the original one. Each row corresponds to an example of a sample of each test dataset used with  $r_0$  values from 7 to 17 cm in steps of 2 cm and its corresponding reconstructions. On the left, the reconstructions carried out by the FCN whose inputs are directly the information received by the SHWFS are shown. In the center of the image, the original turbulence phase simulated is shown, while on the right, the reconstructions carried out from the cross-correlations are shown.

### 3.2. High-Altitude Layer Turbulence

In this subsection, the results of the reconstruction for high-altitude layers applied over simulations with turbulent profiles of altitudes between 9 and 12 km in height are shown.

The results are presented in the same way as in the previous subsection: first, Table 2 shows the analytical values of the errors, similarity, and time needed per reconstruction, and then Figure 5 allows for a visual comparison of a randomly selected example from each test dataset.

**Table 2.** Results obtained for the high-altitude layer reconstructions. WFE net refers to the RMSE WFE calculated over the images given by each FCN as output, while WFE residual refers to the RMSE WFE of the image that is obtained as the difference between the output of one ANN and the original simulated image. Both terms are measured in radians (rad). SHWFS refers to the ANN that uses as input the information received by the subapertures of the sensor, while CC refers to the ANN that uses the cross-correlations (CC) as inputs. The time column shows the mean time over the dataset per sample for each ANN, measured in milliseconds (ms).

$r_0$ (cm)	WFE Net (rad)		WFE Residual (rad)		Similarity (%)		Time (ms)	
	SHWFS	CC	SHWFS	CC	SHWFS	CC	SHWFS	CC
7	9.05	8.97	1.69	1.78	94.1	94.0	4.50	4.37
9	7.97	7.96	1.25	1.30	95.1	95.0	4.52	4.33
11	6.64	6.69	0.95	1.02	95.6	95.7	4.47	4.33
13	5.02	5.03	0.80	0.86	95.6	95.4	4.53	4.32
15	4.63	4.69	0.72	0.79	95.9	95.5	4.41	4.32
17	4.50	4.53	0.70	0.74	96.6	95.8	4.57	4.35

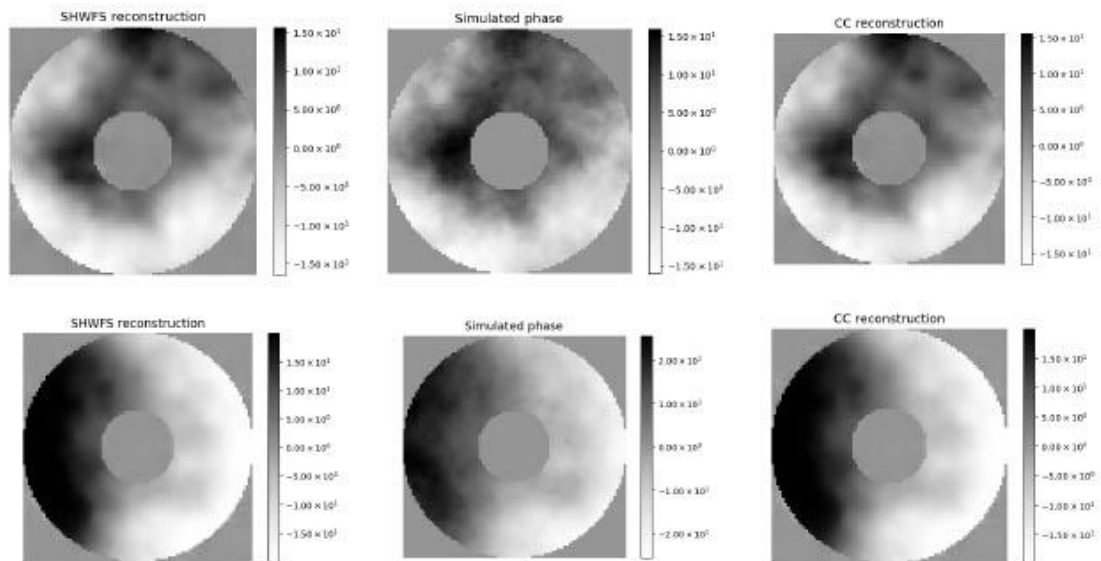
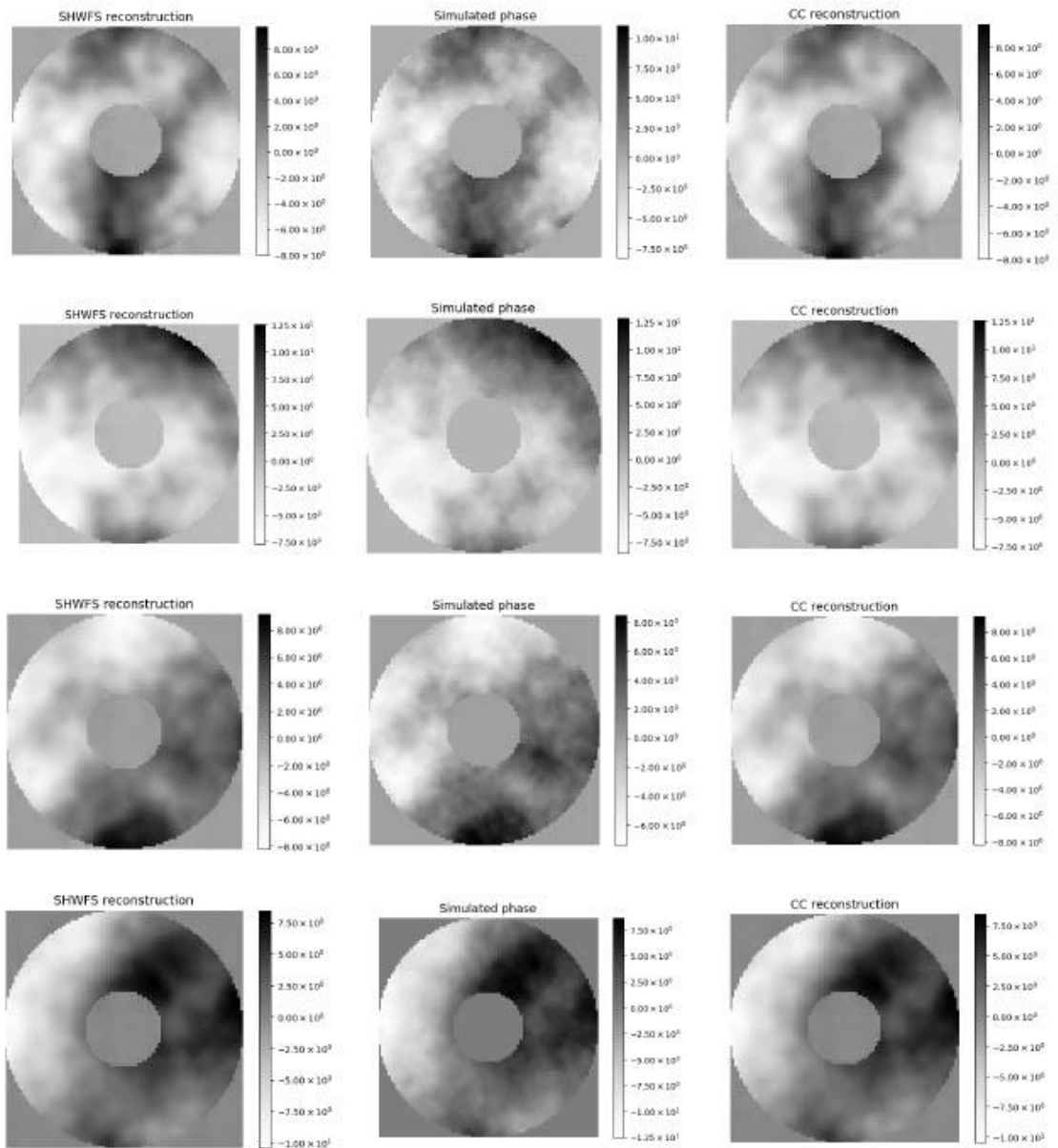


Figure 5. Cont.



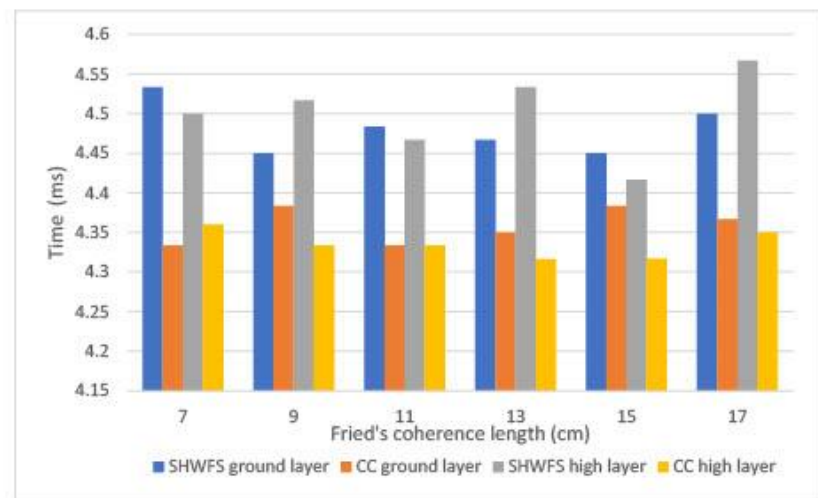
**Figure 5.** Comparison of the reconstructed phases for high-altitude layer turbulences with the original one. Each row corresponds to an example of a sample of each test dataset used, with  $r_0$  values from 7 to 17 cm in steps of 2 cm, and its corresponding reconstructions. The images are placed following the order presented in Figure 4.

#### 4. Discussion

The performances of two different approaches to recover the turbulent profile’s phase in the context of solar AO were presented in this research. Information received by the SHWFS or its calculated cross-correlations was used as input data for an FCN. The main objective is to know if the loss of information due to pre-processing the information received

affects the quality of the reconstruction of a reconstruction system, particularly when using an ANN.

Both methods were analyzed in terms of the computational time needed by the RS, where a visual representation of the obtained results is shown in Figure 6. The figure depicts how the use of cross-correlations minimally decreases the time needed per sample by the RS independently of the type of turbulence layer reconstructed. The time is reduced, in mean terms, by approximately 3% for the ground layer and 4% for the second case, being irrelevant values for the selection of the best model. Moreover, considering the case of cross-correlations, the recall time needed was not considered, meaning the method would be slower than the one with SH images.



**Figure 6.** Required time needed per epoch for both models in the two cases of turbulence profiles considered. SHWFS represents the FCNs whose inputs are directly the information received by the wavefront sensor, while CC reconstruction corresponds to the cross-correlations. To make comprehension easier, the bars for each case are in the same order as the legend.

The quality of the reconstruction was analyzed in analytical terms of the WFE and similarity, and a visual comparison could be also made, as shown in Figures 4 and 5.

In both methods, it is observed that the ground layer is more difficult to reconstruct given the low values of similarity of the turbulence's phases. This matches with the initial assumption as ground layers have always been a big challenge for AO systems, especially in solar observations. As it has been mentioned in Section 2, in diurnal observations, due to the impact of the solar energy reflected by the Earth, ground layers have more energy and faster variations. This is reflected in Figures 7 and 8, where a clear trend is seen when comparing the results for ground layer turbulences and altitude layer turbulence for both methods. In the case of the SHWFS neural networks, the residual WFE committed is, in mean terms, for the six tests, 34% lower when the turbulence layer is situated at high altitudes, while in the CC reconstruction, that value is 39%. The height of the turbulence has more influence in the reconstruction when all the information received by the Shack-Hartmann sensor is used as the input of the neural network. This fact can be expected since cross-correlations are processed data that usually have approximately the same aspect, being robust to variations in the turbulence, which does not happen in the case of the light received by the sensor. These variations affect the response of the ANN and affect the quality in the reconstructions, which strongly depends on each scenario, in contrast with the CC FCN case.



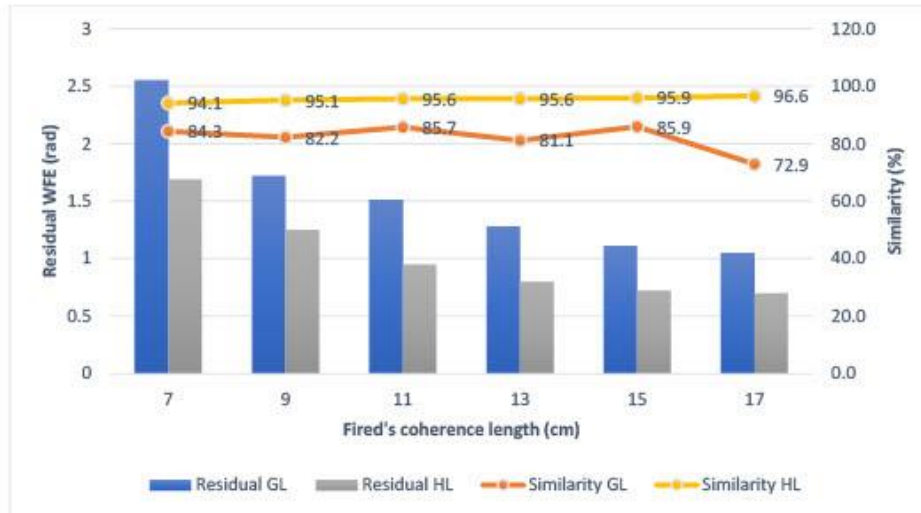


Figure 7. Influence of the height in the reconstructions, using the information received by the sensor as inputs for the FCN for the several  $r_0$  values tested. To make comprehension easier, the bars for each case are in the same order as the legend and each case has its own markers.

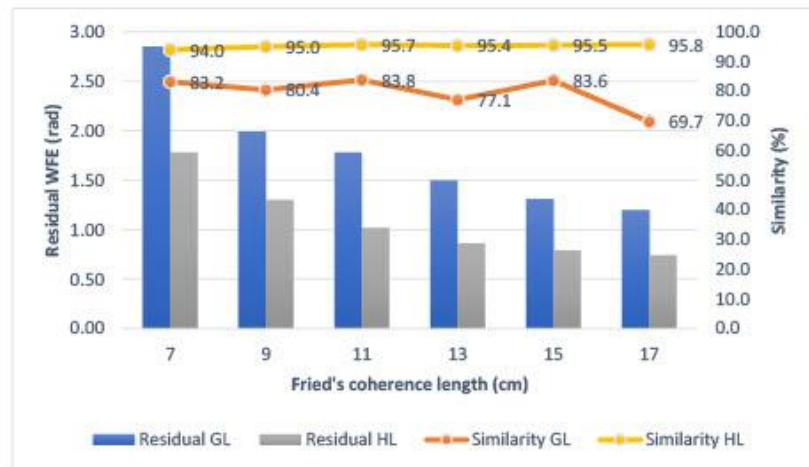


Figure 8. Influence of the height of the layers (ground layers (GL) and high-altitude layers (HL)) in the reconstructions carried out using the cross-correlations as inputs for the FCN for the several  $r_0$  values tested. To make comprehension easier, the bars for each case are in the same order as the legend and each case has its own markers.

The same analysis can be conducted when considering the influence of the  $r_0$  value in the quality of the reconstructions. The error committed lowers as the  $r_0$  value increases for both cases, something expected as the intensity of the turbulence also decreases. However, the difference in the error committed for an  $r_0$  value compared with the previous one is slightly smaller in the case of the CC network. For the first case, for example, between 7 and 9 cm, in the SHWFS case, the error is 33% lower for 9 cm, as in the CC case, where it is 30%. Thus, the CC FCN appears to be more robust to turbulence variations than the

SHWFS one, since the second one receives much more information of the turbulence for each moment.

The results explained can visually be analyzed in Figures 4 and 5. The profiles of the turbulence phases reconstructed in Figure 4 have less definition than the Figure 5 ones, as they correspond to ground layer turbulence. Moreover, as the rows of the figures are analyzed, the definition is clearly higher in the rows below (higher  $r_0$  situations) than in those above.

Finally, we present the comparison of the results obtained for both methods, using the light received by the Shack–Hartmann sensor as input (SHWFS method) or the cross-correlations (CC method). In Figure 9, the representation of the similarity obtained by each method in the reconstructed turbulence phases for all the situations tested can be seen. The results are very similar in both methods, but a clear trend can be seen; for every case, the reconstruction carried out using all the information received by the Shack–Hartmann sensor improves the similarity to the one carried out from the cross-correlations. Nevertheless, that difference is more accentuated for the cases of ground layer turbulences than for the other ones. The loss of information in the process of establishing cross-correlations between the subapertures means the ANN is not able to reconstruct the turbulence phase in such a similar way. For the cases of ground layer reconstructions, the SHWFS method obtains increases in the similarity between 1.1% and 4%, being the mean for the six tests, obtaining an average improvement of 2.7% in similarity terms. That difference is lower for the high-altitude layer reconstruction, the average being only 0.25%.

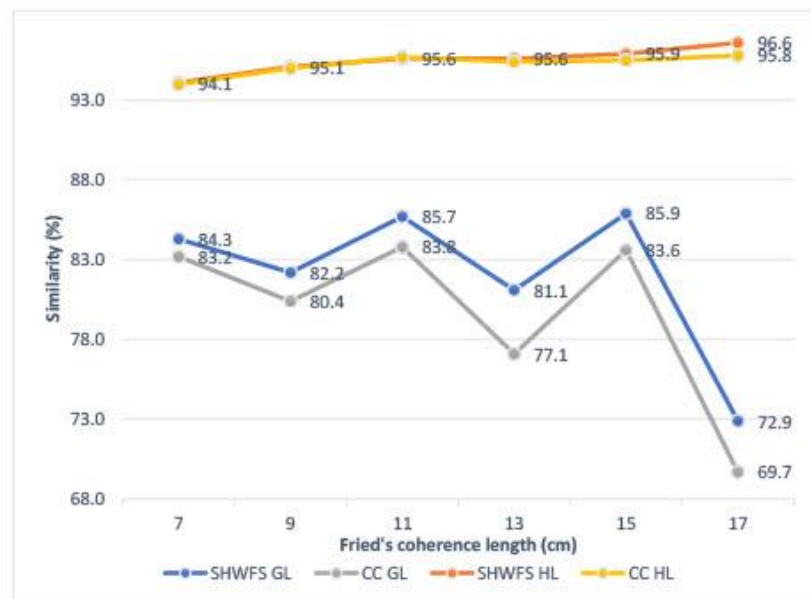


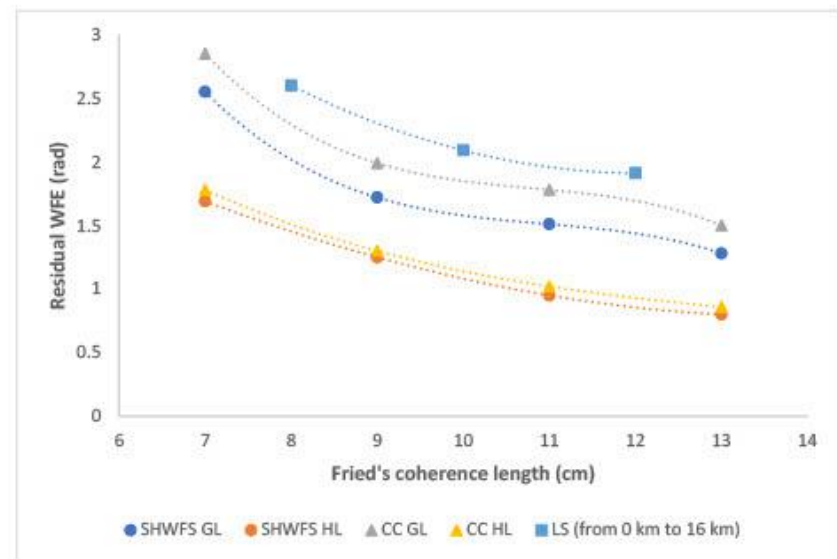
Figure 9. Comparison of the similarity obtained by both methods for ground layer (GL) and high-altitude layer (HL) situations. Note that the markers in the figure are fixed for each ANN model.

In order to analyze the quality of the results obtained by both ANN models, some data of other AO reconstructors used in previous research [33] are compared. In Table 3, the residual WFE obtained in reconstructions applying the least squares method (LS) [10], the most commonly used algorithm in real AO systems, is shown. The characteristic of the AO system is the same as in this research in terms of the telescope's pupil diameter and subapertures. The data given are calculated as an average for turbulence layers from 0 to 16 km in height over 400 samples; therefore, in this case, there are no specific reconstruction

error values for ground layer or high-altitude layer reconstructions. Comparing the residual WFEs of Table 3 with the ones obtained for the high-altitude layer reconstructions using a polynomial approximation of order 3, both methods based on FCNs provide lower WFEs. The FCN obtains, for high-altitude layers with an  $r_0$  of 8 cm, 44.3% less residual WFE than the LS method for several turbulences until 16 km in height for the case of the SHWFS image as the input. That value decreases to 42.6% for the case of CC as the input data. In the case of ground layer reconstructions, the SHWFS FCN reduces the LS error by 21.8%, while the CC FCN reduces it by 11.9%, keeping in mind that the LS data are an average over altitudes between 0 and 16 km in height. Anyway, even in the worst case where the FCNs only reconstruct ground layer turbulences, the LS residual WFE is reduced at least by 11%, as it can be seen in Figure 10, where the polynomial approximation calculated for FCN models is always lower than the LS one.

**Table 3.** Average of the residual wavefront error obtained by the least squares (LS) method in simulations of a Solar SCAO AO system with a telescope of 1.5 m of diameter and  $15 \times 15$  subapertures over 400 samples for 3 different tests. Each test has a fixed  $r_0$  value and one turbulence layer varying from 0 to 16 km. The data are extracted from [33].

$r_0$ (cm)	8	10	12
WFE Residual LS (rad)	2.60	2.09	1.91



**Figure 10.** Comparison of the residual WFE obtained by both ANN methods in comparison with the most used algorithm in real Solar AO system based on the least squares (LS) method. The data for the LS reconstruction are extracted from [33]. A polynomial trend line of order 3 has been added for each case. Note that the range of altitudes for the LS method is higher than for the other ones. The markers in the figure are fixed for each model for easier comprehension.

The difference remains when comparing higher  $r_0$  values that are closer to real observation situations. Comparing the residual WFE obtained by these two new models with a 12 cm  $r_0$  value with the LS one, the error obtained by the SHWFS ANN decreases by 21.8% for the ground layer comparison, while the FCN with CC as input also decreases by 11.7%. If the high-altitude layer is compared, the SHWFS FCN obtains 48.6% less error, and the CC FCN 47.1% less, according to Figure 10. Therefore, both FCN models achieve



approximately only about half the error when operating in the worst turbulence situations. It is important to realize that the altitude layer cases for the FCNs were simulated between 9 and 12 km, while for the LS between 0 m and 16 km, being the mean of the altitudes, 8 km height was used.

Considering the comparisons in Figures 4 and 5, the results shown in Figure 9 can be visually confirmed, as for each row, independently of the figure, the image on the left is more defined, similar to the center one, than the image on the right panel. Both images are of excellent quality reconstructions as both are parallel to the one in the center, having the high-phase difference zones in the same region of the image, and the same for the low-phase difference zones; however, the left one is always more detailed and precise than the right one.

Further, the recall time rounds to 4.5 ms per sample, very close to real-time computation. Note that the expected times for each correction of the whole system are usually considered under 2 ms [34], this approach being a solid candidate for real implementation. This can be supported by the fact that, although both the correlation-based approach and the SH image-based approach have similar recall times, the first one requires using time and resources to compute the correlations, adding a reason to determine the SH image-to-phase approach as the best choice. In general, those approaches save time that is usually required for other processes of AO, such as translating the SH images to correlations or centroids, and translating the outputs (usually Zernike coefficients or ideal centroids) to the phase [35]. The latest improvements are difficult to measure since they rely directly on the algorithm used.

## 5. Conclusions

In this research, AO reconstruction for solar scenarios was addressed. The proposed method relies on the numerical approximations performed by FCNs, which were trained successfully with simulated data, to predict turbulent profile phases. Two scenarios were considered: obtaining the phases from solar SH images, or from the images of the calculated correlations. The obtained results show how, considering performance, the network that uses images directly from the SH recovers phases that are numerically closer to the actual phase. Note that, in shape, both are quite similar. Moreover, the resolution that a DM can achieve for representing a phase is usually lower than the precision obtained by the methods presented here.

Anyway, both FCN methods obtain lower values of the residual WFE in comparison with the most common reconstruction algorithm currently used, the LS method. From the research, it can be extracted that the ANNs improved the residual WFE by 1% (lower) in the case of CC images and by 21% in the case of SH images.

However, the pre-processed images were shown to be more robust to variations than the solar SH images, allowing the CC FCN to be less susceptible to changes in the input data, since, after the cross-correlations are calculated, the resultant images have a similar pattern. As the CC FCN also obtained good-quality reconstructions, this renders cross-correlations a significant option as input data for ANN-based reconstructors when they are applied to non-trained situations, such as new turbulence conditions or observations in previously unknown solar regions.

In terms of similarity, both reconstructors based on FCNs showed very good performances in the reconstruction of high-altitude layers, achieving similarities higher than 93% in the worst case. This makes FCNs a very interesting option in more complex solar adaptive systems such as in multi-conjugated adaptive systems (MCAO), where several DMs are conjugated and each one only corrects determinate altitude layers; ANNs could take care of high-altitude layers due to their good performance for those cases.

Considering time requirements, both approaches are close to real-time computation, besides allowing for avoiding other algorithms that estimate intermediate values required for the reconstruction.

Possible future work related to this research involves the implementation of an optic bench for the subsequent implementation in a real solar telescope, such as the European Solar Telescope (EST). An optic bench will provide some realistic data to check the performance of the ANNs with real SH images or cross-correlations. However, with an optic bench, only a few several turbulence situations can be obtained, depending on the device available. Another future implementation is to develop the system for more complex AO configurations, such as ground layer adaptive optics (GLAO) or MCAO.

**Author Contributions:** F.G.R. and S.L.S.G. conceived and designed the study. C.G.-G. and F.G.R. contributed to the production of the simulations and the training of models. F.G.R. and E.D.A. performed the comparisons and analysis as well as the preparation of the manuscript. J.D.S. and S.L.S.G. contributed with interpretation of the results and with the preparation of the final version of the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Spanish Economy and Competitiveness Ministry with Grant Number AYA2017-89121-P.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author. The data are not publicly available due to their large size.

**Conflicts of Interest:** The authors declare no conflict of interest.

#### Abbreviations

AI	Artificial Intelligence
ANN	Artificial Neural Network
AO	Adaptive Optics
CC	Cross-Correlations
CC GL	It refers to the ANN which reconstructs the ground layer that has as input the image of cross-correlations
CC HL	It refers to the ANN which reconstructs the high-altitude layer that has as input the image of cross-correlations
CCD	Charge-Coupled Device (camera used by the sensor)
CNN	Convolutional Neural Network
DASP	Durham Adaptive Optics Simulation Platform
DM	Deformable Mirror
FCN	Fully Convolutional Neural Network
GLAO	Ground Layer Adaptive Optics
LS	Least Squares
MCAO	Multi-Conjugated Adaptive Optics
MLP	Multilayer Perceptron
MOAO	Multi-Object Adaptive Optics
RMSE WFE	Root Mean Squared Error Wavefront Error
$r_0$	Fried's Coherence Length
SCAO	Single-Conjugated Adaptive Optics
SH, SHWFS	Shack-Hartmann Wavefront Sensor
SHWFS GL	It refers to the ANN which reconstructs the ground layer that has as input the information received by the Shack-Hartmann sensor
SHWFS HL	It refers to the ANN which reconstructs the high-altitude layer that has as input the information received by the Shack-Hartmann sensor
WFE	Wavefront Error
WFE net	Wavefront error of the output of the network that has
CC	The image of cross-correlations as input
SHWFS	The image received by the Shack-Hartmann sensor as input



WFE residual	Wavefront error of the difference between the simulated image and the output of the network that has:
CC	The image of cross-correlations as input
SHWFS	The image received by the Shack–Hartmann sensor as input
WFE residual LS	Wavefront error of the difference between the simulated image and the reconstructed image by the least squares algorithm

## References

- Russell, S.J.; Norvig, P. *Artificial Intelligence: A Modern Approach*; Pearson Education Limited: Kuala Lumpur, Malaysia, 2016.
- Sands, T. *Deterministic Artificial Intelligence*; IntechOpen: London, UK, 2020.
- Sands, T. Development of Deterministic Artificial Intelligence for Unmanned Underwater Vehicles (UUV). *J. Mar. Sci. Eng.* **2020**, *8*, 578. [[CrossRef](#)]
- Smeresky, B.; Rizzo, A.; Sands, T. Optimal Learning and Self-Awareness Versus PDI. *Algorithms* **2020**, *13*, 23. [[CrossRef](#)]
- Batchelor, G.K. *The Theory of Homogeneous Turbulence*; Cambridge UP: London, UK, 1953.
- Zilberman, A.; Golbraikh, E.; Kopeika, N.S. Propagation of electromagnetic waves in Kolmogorov and non-Kolmogorov atmospheric turbulence: Three-layer altitude model. *Appl. Opt.* **2008**, *47*, 6385–6391. [[CrossRef](#)] [[PubMed](#)]
- Golbraikh, E.; Branover, H.; Kopeika, N.S.; Zilberman, A. Non-Kolmogorov atmospheric turbulence and optical signal propagation. *Nonlinear Process. Geophys.* **2006**, *13*, 297–301. [[CrossRef](#)]
- Osborn, J.; Guzman, D.; de Cos Juez, F.J.; Basden, A.G.; Morris, T.J.; Gendron, E.; Butterley, T.; Myers, R.M.; Guesalaga, A.; Sánchez Lasheras, F.; et al. Open-loop tomography with artificial neural networks on CANARY: On-sky results. *Mon. Not. R. Astron. Soc.* **2014**, *441*, 2508–2514. [[CrossRef](#)]
- Suárez Gómez, S.L.; González-Gutiérrez, C.; Díez Alonso, E.; Santos Rodríguez, J.D.; Sánchez Rodríguez, M.L.; Morris, T.; Osborn, J.; Basden, A.; Bonavera, L.; González-Nuevo González, J.; et al. Experience with Artificial Neural Networks applied in Multi-Object Adaptive Optics. *Publ. Astron. Soc. Pac.* **2019**, *131*, 108012. [[CrossRef](#)]
- Rimmele, T.R. Solar adaptive optics. In Proceedings of the Adaptive Optical Systems Technology, Munich, Germany, 7 July 2000; Volume 4007, pp. 218–232.
- García Riesgo, F.; Suárez Gómez, S.L.; Santos, J.D.; Díez Alonso, E.; Sánchez Lasheras, F. Overview and Choice of Artificial Intelligence Approaches for Night-Time Adaptive Optics Reconstruction. *Mathematics* **2021**, *9*, 1220. [[CrossRef](#)]
- Suárez-Gómez, S.L.; González-Gutiérrez, C.; Sánchez-Lasheras, F.; Basden, A.G.; Montilla, I.; De Cos Juez, F.J.; Collados-Vera, M. An approach using deep learning for tomographic reconstruction in solar observation. In Proceedings of the Adaptive Optics for Extremely Large Telescopes 5, Instituto de Astrofísica de Canarias (IAC), Tenerife, Spain, 25–30 June 2017.
- Sánchez Lasheras, F.; Ordóñez, C.; Roca-Pardiñas, J.; de Cos Juez, F.J. Real-time tomographic reconstructor based on convolutional neural networks for solar observation. *Math. Methods Appl. Sci.* **2020**, *43*, 8032–8041. [[CrossRef](#)]
- LeCun, Y.; Bengio, Y. Convolutional networks for images, speech, and time series. In *The Handbook of Brain Theory and Neural Networks*; A Bradford Book: Cambridge, MA, USA, 1995; pp. 255–258.
- Gardner, M.W.; Dorling, S.R. Artificial neural networks (the multilayer perceptron)—A review of applications in the atmospheric sciences. *Atmos. Environ.* **1998**, *32*, 2627–2636. [[CrossRef](#)]
- Giusti, A.; Ciresan, D.C.; Masci, J.; Gambardella, L.M.; Schmidhuber, J. Fast image scanning with deep max-pooling convolutional neural networks. In Proceedings of the Image Processing (ICIP), 2013 20th IEEE International Conference, Melbourne, Australia, 15–18 September 2013; pp. 4034–4038.
- Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
- Chundi, G.S.; Lloyd-Hart, M.; Sundareshan, M.K. Training multilayer perceptron and radial basis function neural networks for wavefront sensing and restoration of turbulence-degraded imagery. In Proceedings of the 2004 IEEE International Joint Conference on Neural Networks, Budapest, Hungary, 25–29 July 2004; Volume 3, pp. 2117–2122.
- Keras Special Interest Group Keras Layer Activation Functions. Available online: <https://keras.io/api/layers/activations/> (accessed on 11 March 2021).
- Nair, V.; Hinton, G.E. Rectified linear units improve restricted boltzmann machines. In Proceedings of the 27th International Conference on Machine Learning (ICML-10), Haifa, Israel, 21–24 June 2010; pp. 807–814.
- Zhang, Z.; Sabuncu, M. Generalized cross entropy loss for training deep neural networks with noisy labels. In Proceedings of the Advances in Neural Information Processing Systems, Montréal, QC, Canada, 3–8 December 2018; pp. 8778–8788.
- LeCun, Y.; Boser, B.; Denker, J.S.; Henderson, D.; Howard, R.E.; Hubbard, W.; Jackel, L.D. Backpropagation applied to handwritten zip code recognition. *Neural Comput.* **1989**, *1*, 541–551. [[CrossRef](#)]
- Chauvin, Y.; Rumelhart, D.E. *Backpropagation: Theory, Architectures, and Applications*; Psychology Press: Hove, UK, 2013.
- Ruder, S. An overview of gradient descent optimization algorithms. *arXiv Prepr.* **2016**, arXiv:1609.04747.
- Nesterov, Y. *Lectures on Convex Optimization*; Springer: Berlin/Heidelberg, Germany, 2018; Volume 137.
- Suárez Gómez, S.L.; García Riesgo, F.; González Gutiérrez, C.; Rodríguez Ramos, L.F.; Santos, J.D. Defocused Image Deep Learning Designed for Wavefront Reconstruction in Tomographic Pupil Image Sensors. *Mathematics* **2021**, *9*, 15. [[CrossRef](#)]
- Davies, R.; Kasper, M. Adaptive optics for astronomy. *Annu. Rev. Astron. Astrophys.* **2012**, *50*, 305–351. [[CrossRef](#)]

28. Osborn, J.; Guzman, D.; de Cos Juez, F.J.; Basden, A.G.; Morris, T.J.; Gendron, É.; Butterley, T.; Myers, R.M.; Guesalaga, A.; Sanchez Lasheras, F.; et al. First on-sky results of a neural network based tomographic reconstructor: Carmen on Canary. In Proceedings of the SPIE Astronomical Telescopes and Instrumentation, Montréal, QC, Canada, 25–26 June 2014; Marchetti, E., Close, L.M., Véran, J.-P., Eds.; International Society for Optics and Photonics: Bellingham, WA, USA, 2014; p. 91484M.
29. Von Der Lühe, O.; Widener, A.L.; Rimmele, T.; Spence, G.; Dunn, R.B. Solar feature correlation tracker for ground-based telescopes. *Astron. Astrophys.* **1989**, *224*, 351–360.
30. Bracewell, R.N.; Bracewell, R.N. *The Fourier Transform and Its Applications*; McGraw-Hill: New York, NY, USA, 1986; Volume 31999.
31. Basden, A.G.; Bharmal, N.A.; Jenkins, D.; Morris, T.J.; Osborn, J.; Peng, J.; Staykov, L. The Durham Adaptive Optics Simulation Platform (DASP): Current status. *SoftwareX* **2018**, *7*, 63–69. [[CrossRef](#)]
32. Berkefeld, T.; Schmidt, D.; Soltau, D.; von der Lühe, O.; Heidecke, F. The GREGOR adaptive optics system. *Astron. Nachrichten* **2012**, *333*, 863–871. [[CrossRef](#)]
33. Riesgo, F.G.; Gómez, S.L.S.; Rodríguez, J.D.S.; Gutiérrez, C.G.; Alonso, E.D.; Rodríguez, F.J.I.; Fernández, P.R.; Bonavera, L.; Menéndez, S.d.C.; de Cos Juez, F.J. Early Fully-Convolutional Approach to Wavefront Imaging on Solar Adaptive Optics Simulations. In Proceedings of the International Conference on Hybrid Artificial Intelligence Systems, Gijón, Spain, 11–13 November 2020; pp. 674–685.
34. González-Gutiérrez, C.; Sánchez-Rodríguez, M.L.; Calvo-Rolle, J.L.; de Cos Juez, F.J. Multi-GPU Development of a Neural Networks Based Reconstructor for Adaptive Optics. *Complexity* **2018**, *2018*, 5348265. [[CrossRef](#)]
35. Guzmán, D.; de Cos Juez, F.J.; Myers, R.; Guesalaga, A.; Lasheras, F.S. Modeling a MEMS deformable mirror using non-parametric estimation techniques. *Opt. Express* **2010**, *18*, 21356–21369. [[CrossRef](#)] [[PubMed](#)]





### 11.3 COMPARATIVE STUDY OF SHACK-HARTMANN CONFIGURATIONS FOR ATMOSPHERIC TURBULENCE RECONSTRUCTIONS IN SOLAR ADAPTIVE OPTICS

**AUTORES: FRANCISCO GARCÍA RIESGO, SERGIO LUIS SUÁREZ GÓMEZ, JESÚS DANIEL SANTOS  
RODRÍGUEZ, CARLOS GONZÁLEZ-GUTIÉRREZ, YOLANDA MARTÍN HERNANDO, LUZ MARÍA MONTOYA  
MARTÍNEZ, ANDRÉS ASENSIO RAMOS, MANUEL COLLADOS VERA, MIGUEL NÚÑEZ CAGIGAL Y FRANCISCO  
JAVIER DE COS JUEZ**



Contents lists available at ScienceDirect

Optics and Lasers in Engineering

journal homepage: [www.elsevier.com/locate/optlaseng](http://www.elsevier.com/locate/optlaseng)

## Comparative study of Shack-Hartmann configurations for atmospheric turbulence reconstructions in solar adaptive optics

Francisco García Riesgo<sup>a,b,\*</sup>, Sergio Luis Suárez Gómez<sup>a,c</sup>, Jesús Daniel Santos Rodríguez<sup>a,b</sup>, Carlos González Gutiérrez<sup>a,d</sup>, Yolanda Martín Hernando<sup>e</sup>, Luz María Montoya Martínez<sup>e</sup>, Andrés Asensio Ramos<sup>e</sup>, Manuel Collados Vera<sup>e</sup>, Miguel Núñez Cagigal<sup>e</sup>, Francisco Javier De Cos Juez<sup>a,f</sup>

<sup>a</sup> ICTEA (Instituto de Ciencias y Tecnologías Espaciales de Asturias), Spain

<sup>b</sup> Department of Physics, University of Oviedo, Spain

<sup>c</sup> Department of Mathematics, University of Oviedo, Spain

<sup>d</sup> Department of Computer Science, University of Oviedo, Spain

<sup>e</sup> IAC (Instituto Astrofísico de Canarias), Spain

<sup>f</sup> Department of Mining Exploitation and Prospecting, University of Oviedo, Spain

### ARTICLE INFO

#### Keywords:

Adaptive optics  
Artificial intelligence  
Deep learning  
Shack-Hartmann  
Atmospheric turbulence  
Neural networks  
Fully-Convolutional Neural Networks

### ABSTRACT

The correction of the phase variations induced by the atmospheric turbulence is one of the most complex problems that an Adaptive Optics (AO) System must deal with, as it must calculate the properties of all the atmosphere traversed by the light from several measures taken by ground-based telescopes. Traditional reconstructors systems used in AO are based on computational algorithms where its reconstruction quality improves with the number of measures made by the telescopes' sensors. That means that sensors are getting greater and greater with their corresponding higher financial expense. Artificial Intelligence (IA) has become in recent years a real alternative to traditional computational methods as reconstructors for AO systems. Fully-convolutional neural networks (FCNs) specifically have shown great performances working in Solar AO, demonstrating their ability to obtain a lot of valuable information from the recorded images for the wavefront phase evaluation. Along this research, the influence of the properties of the telescope's sensors and of the observations in the reconstructions made by the FCNs' is measured, to obtain the configuration that best suits the performance of artificial neural networks (ANN). The presented results determine the way forward for the future sensors for telescopes with reconstruction systems based on ANNs, to obtain higher quality reconstructions employing fewer economic resources.

### 1. Introduction

Telescopes have evolved from a simple combination of mirrors, lenses and crystals to a complex set of technology-improved systems working together, that pursues a better understanding of the universe, by improving the quality of images from celestial bodies. amongst the most recent technologies, the improvements that some of these developments brought made those systems as Adaptive Optics (AO), engineering components, systems of sensors, etc. became necessary by its own merits.

AO is the field of optics whose aim is to correct in real time the distortions caused by the atmosphere in the received images on ground-based telescopes. In the last 30 years [1], with the improvement of computa-

tional systems, AO has experienced a tremendous development. Several AO systems were implemented in large ground-based telescopes and are still operating. On these telescopes, and on those planned for near-future implementation, rest the responsibility of the new discoveries that will be made during the next decades.

In the case of the technology for diurnal observation, its development for complex systems is still being researched since the majority of the telescopes that will perform solar observations in the upcoming years are still under construction, as the EST (European Solar Telescope) [2], a new four-metre class telescope. The development of these new large solar telescopes results on a new challenge for AO systems, they present new configurations as the Multi-Conjugate Adaptive Optics (MCAO) that implies that the amount of information to process in real time greatly

\* Corresponding author at: ICTEA (Instituto de Ciencias y Tecnologías Espaciales de Asturias), Spain.  
E-mail address: [garciafrancisco@uniovi.es](mailto:garciafrancisco@uniovi.es) (F.G. Riesgo).

<https://doi.org/10.1016/j.optlaseng.2022.107157>

Received 2 March 2022; Received in revised form 9 June 2022; Accepted 13 June 2022

Available online 24 June 2022

0143-8166/© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)



escalates in comparison with night-time AO systems. Some older solar telescopes with simpler configurations are currently using Solar AO systems, that are already well tested and implemented, as the one-metre class SST (Swedish Solar Telescope) [3,4] or the DST (Dunn Solar Telescope).

The field of Solar physics is less developed due to the extra difficulties involved in making daytime observations. The principal attractive is the Sun's atmosphere where many processes that affect live on Earth take place. amongst them, the most relevant are the Sun's luminosity variations, changing Earth's climate and how the magnetic fields are generated and dissipated in the Sun's atmosphere. It is necessary to obtain high quality images of the Sun's atmosphere and the Sun's surface to understand its activity.

The corrections made by Solar AO systems allow the telescopes to correct the aberrations suffered by the light on its way from the source to the telescope, obtaining a higher quality image; most aberrations are produced when the light passes throughout the Earth's atmosphere. The atmospheric turbulence is a random phenomenon that is continuously changing and producing aberrations on the wavefronts of the light. In a current AO system, there are some different parts that work together to obtain the corrected image. One of them is the Reconstructor System (RS) that is briefly introduced here since it is the focus of this research. When the light arrives at the telescope, measurements are made by several sensors that are sent to the RS to calculate an estimation of how the atmosphere was when the light passed through it. If it were completely known, the received image could be perfectly corrected with an ideal AO system.

In recent years, Artificial Intelligence (AI) methods have been applied in several scientific fields as mathematical tool that allows to simplify complex physical systems by numerical approximations, improving their performance. Artificial Neural Networks (ANNs) are one of the most developed fields in AI. Their good performance is well known in image recognition, language processing, image classification, etc. [5,6]. Several science branches have taken advantage of these improvements and ANNs are currently applied in diverse areas, such as prediction systems, car industry to make self-sufficient cars, simulation platforms of diverse nature, etc. AO is one of the fields where ANNs have been applied showing a great performance in night observations, as the CARMEN reconstructor [7–9].

In this research a new reconstructor system (RS) is presented for extended images, particularized to solar observations, based on Fully-Convolutional Neural Networks (FCNs), a type of ANN characterized by its good performance working with images. The main objective of the research is to establish a comparison of how the reconstruction changes depending on the variations of some parameters of the AO system; the number of subapertures of the Shack-Hartmann wavefront sensor, the number of pixels of each subaperture and the field of view of the observation are the ones compared. The properties of FCNs allow to obtain reconstructions working with simpler AO sensors than the ones currently used, requiring Shack-Hartmann wavefront sensors with less subapertures and pixels per subaperture, that will imply lower economic contributions and a simpler control.

This manuscript is composed by 4 more sections; the material and methods section consist in a description of the main concepts of adaptive optics, Shack-Hartmann wavefront sensor, the simulation platform used to generate the data and artificial intelligence, explaining how the problem can be solved with their use. Next, the architecture and some parameters of the FCN used are detailed in the Section 3, showing the results of the RS and their discussion in Section 4. Finally, some conclusions and future developments are presented in the last section.

## 2. Material and methods

In this section an introduction to the main concepts of Solar adaptive optics and artificial neural networks is presented.

### 2.1. Solar adaptive optics (Solar AO)

The term Adaptive Optics comprehends a set of techniques whose aim is the real-time correction of the images received by ground-based telescopes. Aberrations are present in the images formed of celestial bodies causing that the quality of the images decreases considerably. The principal source of aberrations is the atmospheric turbulence existing in the part of the way where the light travels through the Earth's atmosphere.

The Earth's atmosphere consists of multiple air masses of several sizes and shapes that are in constant relative in movement. Each one has its own properties: size, wind's direction and velocity, temperature, height, etc. The relative movements between the air masses, like fluids movements, cause the atmospheric turbulence to be a random phenomenon in constant change, which makes the work of AO systems difficult to reach real time corrections, without knowing how the turbulence will be in the next moment. The turbulence of larger scales is passed to turbulences at lower scales [10]. The air masses are commonly represented as turbulence layers at different heights, so the path that light makes through the atmosphere consists of passing through several turbulence layers, each one with its own refraction index value. The way that the light goes through the atmosphere to the telescope can be modelled as it has traversed a path with multiple lenses of various refraction indexes, one after the other [1].

Corrections performed by an AO system are made after the light is received by the telescope, before it is observed. The system must do several processes that includes to sense the image, to reconstruct the estimated atmospheric turbulence that the light has traversed and finally with the previous information to calculate the corrections to apply, to eliminate as much aberration effects as possible.

As it was commented before, the air masses of the turbulence are grouped by layers. According to the Kolmogorov's model [10] when describing a turbulence layer there are two main parameters, its height, and its intensity that is given by the Fried's coherence length (commonly designated as  $r_0$ ) [11]. The physical interpretation given to  $r_0$  corresponds to the diameter of a telescope that, in absence of turbulence, offers the same resolution as a large telescope in presence of that specific atmosphere. So, the lower  $r_0$  value, the higher turbulence intensity;  $r_0$  is commonly given in centimetres, where 12 cm represents a worse day for observation where the turbulence has high intensity and values between 15–18 cm represents a normal day for observation. An extremely bad turbulence condition is presented when  $r_0$  is lower than 10 cm, under these circumstances an observation would usually be suspended.

Another important parameter when performing an observation is the Field of View (FOV). This value represents the maximum angular size of the object as seen from the entrance pupil. Commonly the FOV is determined by the field over which the optical system exhibits good performance.

There are several AO Systems according to its configuration. However, they usually have the same components; at least one Wavefront Sensor (WFS) where the aberrations presented in the received wavefront are measured, a Reconstructor System (RS) which estimates how the atmosphere is at each moment and calculates a correction for the aberrations and at least one Deformable Mirror (DM) that consists in a mirror which surface can be modified where the received wavefront is reflected and is applied to it the corrections calculated by the RS. In this work an SCAO system was chosen (Simple Conjugate Adaptive Optics). It is the simplest system in AO since it has one WFS and one DM on-axis with the celestial body working on close loop (see Fig. 1).

The reconstructors that are usually used nowadays in AO Systems are based on matrix vector multiplication applying least squares methods or minimum variance techniques [12–14]. In the last years some reconstructors based on ANNs have been successful in night AO systems, as the CARMEN reconstructor or the CARMEN Convolutional reconstructor [15]. The work presented is an evolution of these last kind of reconstructors but in this case applied for Solar observations.



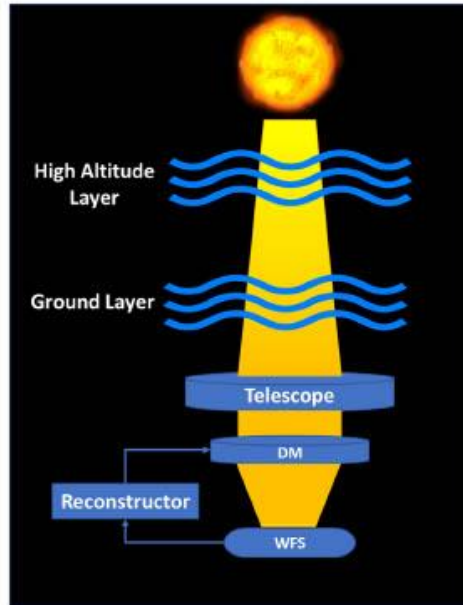


Fig. 1. Schematic representation of a Solar SCAO configuration.

There are several difficulties when making a Solar observation compared with a night one, for the AO system. These are the most important:

- First, the Sun is an extended object when it is observed from Earth since it is much nearer than any other star. Therefore, the telescope's pupil and all the subaperture pixels of the SH are completely saturated by the Sun's surface (see Fig. 2) instead having only a spot of light in the centre. The SH must commonly make some pre-processing on the data received before passing it to the RS.
- In diurnal observations, due to the solar energy emitted, the atmospheric turbulence has more intensity, implying higher turbulences than at night hours. It is expected then that the  $r_0$  is lower at day.
- Solar AO must work in the visible spectral range while night systems tend to work in the infrared where both sensing and correction are easier.
- Some alternative wavefront reconstruction algorithms have been developed for night time astronomy.

#### 2.1.1. Shack-Hartmann wavefront sensor

The Shack-Hartmann (SH) is the most used WFS in Adaptive Optics both in night and solar observations, and the one chosen in this work. It is composed of a matrix of lenses with the same focal length, being focalised each one in a different subaperture. The subapertures are located after the lenses, composed by another matrix of separated CCDs (charge-coupled device) or other photon sensors. Each lens focalises the received light on its corresponding subaperture. In an ideal situation, the light of the celestial body would be received as a plane wavefront that, after passing through the lens, will be focalized in the centre of each subaperture, being equal for all of them. Due to the presence of the atmospheric turbulence, the wavefront presents aberrations that can be calculated by the comparison of the images received by the subapertures.

The SH divides the wavefront in discrete sub-pupils and process the data before passing them to the RS. It calculates cross-correlations [16] between the images received by each subaperture, the number can be limited to reduce the time needed. Then, the maximum of the cross-

correlations is located for each subaperture. The SH is able to obtain from that value the gradient or tilt of each subaperture [17]. The gradient is usually used by commonly RS as the Least-Square method.

The principal difference between the reconstructors based on convolutional neural networks and the others resides on the property of CNNs of using images as inputs. The images received by each subaperture are directly used as input of the RS and the pre-processing data procedure made by the SH is not necessary, avoiding the loss of information and the time needed.

The most common reconstructor systems used nowadays in real telescopes are based in the Least-Squares reconstruction method (LS). When using this reconstructor, the result will be better the more subapertures the sensor has. Nevertheless, the maximum number of subapertures is limited by the diameter of the pupil and the cost of the sensor depends on them, being higher when the number of subapertures increases. For this reason, it is interesting to obtain new reconstructors that can achieve good reconstructions even with a low number of subapertures. The other parameter that has influence on the information received is the number of pixels per subaperture, that represents the number photon sensors that the SH has in each subaperture.

#### 2.2. Fully-convolutional neural networks

Artificial Neural Networks are an interconnected group of processing units, that tries to mimic the behaviour of biological neural networks [18]. These structures are characterized for being able to learn from the data and give an answer to a determinate problem. They are formed by individual computational units, called neurons. The neurons are sorted in different layers, forming a neural structure. The output of one layer is the input of the neurons of the next layer since neurons are linked to those from the previous and following layer. These connections are regulated by weights whose values can be modified during the training process.

The training process is the method that allows an ANN to learn from the data and takes place before the ANN is used. Each neuron applies an activation function over the input given by the previous ones. The process consists of the application of the network to a train dataset that its correct outputs are known. Then, using algorithms that measure the error, the ANN is adjusted to minimize the error of the outputs of train dataset varying some parameters, as the connections weights until the error is decreased to a minimum value. The most used algorithm is the backpropagation algorithm [19] and it is the one used in this work.

The kind of neural networks selected is the Fully-Convolutional Neural Networks (FCNs) [20,21] that are based on Convolutional Neural Networks. An FCN consists in several layers of neurons joined in two blocks, a convolutional block and a deconvolutional block (see Figs. 3 and 4). That allows having both input and output data as a multidimensional array. The convolutional block can be formed by several convolutional layers and pooling layers. The convolutional layers work as filters extracting the main features of the input given. The pooling layers reduce the size of the input by selecting the most significant value over a group of pixels, which is chosen by a set criterion, being the maximum value usually selected. In this work no pooling layers were used; the size of the input was reduced by using convolutional layers with strides size higher than one pixel. At the end of the convolutional block, the features are fussed in the deconvolutional section until obtain the final image as output. This deconvolutional block usually is formed by deconvolutional and transpose pooling layers.

For the case of this research, the input data given to the FCNs are the images of wavefronts received by the SH (as the right image of the Fig. 2) obtaining as output the profile of the turbulence phase for each moment (as it is represented in the schemes of Figs. 3 and 4).

During a common training process, a train dataset is passed through the FCN with input images and their desired outputs are known. The error is backpropagated through the different layers to be minimized. The filters of the layers and the weights of the connections are modified

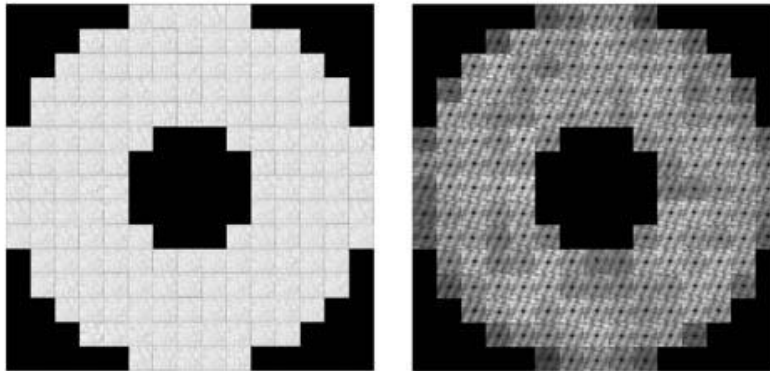


Fig. 2. Example of the information received by an SH of  $15 \times 15$  subapertures. On the left, there is an image of the information received without any pre-processing. On the right, the same information after the cross-correlations have been made by the SH.

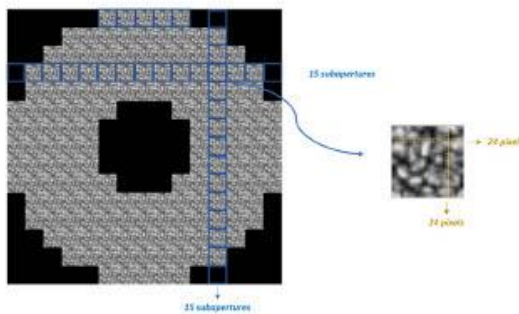


Fig. 3. Description of the main parameters compared in an example of input image. The squares in blue represent the subapertures in the SH image and the ones in yellow the pixels over each subaperture.

with the backpropagation, each time that the error is backpropagated they are updated. The network is applied to the dataset several times to train it, being the number of times that it is passed selected by the user. In the experimental setup section, the main characteristics of the neural networks topologies and the datasets used are detailed.

### 2.3. DASP: durham adaptive optics simulation platform

All the dataset used for the training, validation and test processes were simulated using THE Durham Adaptive optics Simulation Platform (DASP) [22]. DASP tries to mimic the data obtained in real telescopes, allowing to make simulations in diurnal or nocturnal observations. In

the case of diurnal ones, it only can replicate a SCAO AO system or a MCAO one. SCAO was the chosen one for this research.

The atmospheric turbulence is generated according to the Von-Karman statics implemented with Monte-Carlo Simulations [10]. The incident wavefront phase is aberrated with several thin perturbed layers, DASP simulates it agreeing to the previously chosen setup and generates the image of the SH as a convolution of Sun's surface image with the generated turbulence layers.

The platform allows the user to select multiple parameters of the simulation. For each case, a parameter file is needed that specify relevant ones, dimension the system and determines the output to be produced [22]. In the case of this research, these parameters include amongst others the Fried's coherence length ( $r_0$ ), the outer scale ( $L_0$ ), the height of each layer, the wind velocity and the direction of each layer and the contribution of each layer to the total turbulence. Further, much information can be extracted from the simulation, as the information received by the SH, the correlations, the phase of the turbulence, the DM position, etc. In the datasets simulated for the research only the SH images and the image of the turbulence's phase has been saved, whose constitute the input and the output of the FCNs. The base of the simulated SH images is a synthetic solar granulation image of the Sun's surface with a resolution of  $0.0139''/\text{pixel}$ . All the datasets used both for training, testing and validation were simulated with this platform.

### 2.4. Experimental setup

The experiments were performed with an Ubuntu LTS 14.04.3 server, which has 128Gb of DDR4 memory, Nvidia GeForce GTX TitanX, an Intel Xeon CPU E5-1650 v3 @ 3.50 GHz and SSD hard drive. As computer language, Python was chosen for its performance with packages of Arti-

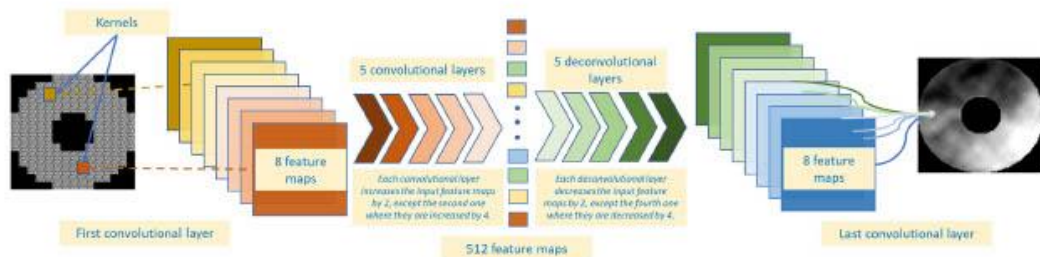


Fig. 4. Schematic representation of an FCNs selected. Due to the number of layers, it corresponds with the Model 1, 2 and 3 since the differences between them are the kernels size and the strides of the layers.



**Table 1**  
Shack-Hartmann configurations used to compare the influence of the number of subapertures. Values correspond with the number of pixels per side, assuming that images of the subapertures and of the phase of the turbulence have a square shape. The table is divided to indicate the cases where the two neural network models were used.

	15 × 15 subapertures	10 × 10 subapertures	7 × 7 subapertures	3 × 3 subapertures
Pixels per subaperture	24 × 24	36 × 36	48 × 48	112 × 112
Total pixels of the Shack-Hartmann	360 × 360	360 × 360	336 × 336	336 × 336
Total number of pixels of the phase	90 × 90	90 × 90	84 × 84	84 × 84
FCN model	Model 1		Model 2	

cial Intelligence as Tensorflow 2 [23] and Keras [24]. Also, the adaptive optics simulator DASP [22] is written with Python 2.

Different simulations were generated for the training, validation and test process of the FCNs. The influence of three different parameters in the reconstruction was studied, these parameters are: the number of subapertures, the number of pixels per subaperture and the FOV of the observation. For each, one train dataset, one dataset for validation and three test datasets were simulated. The simulations presented are based on the SCAO configuration of the Gregor Solar Telescope [25], with 1.5 m optical aperture. Furthermore, each dataset has the same number of images to allow a realistic comparison: train sets had 80,000 while each test dataset had 6000 images. One sun region was set for all the obtained images. Results are shown in the next section, calculated as the mean value over the 6000 simulations.

Considering the parameters of each set, they are as follows: the train dataset consisted of 80,000 simulated images. Each simulation had two different turbulence layers, the first one is always situated at 0 m of height and had a weight of the 10% over all the atmospheric turbulence. The other layer was set at different heights, varying from 0 to 20,000 m in steps of 200 m of height and having the remaining 90% of the weight. That structure of the dataset giving the 90% of weight to the second layer allows to train the network for several heights' turbulence layers. The  $r_0$  value ranged from 8 cm, where the turbulence is stronger, to 16 cm per each altitude simulated in steps of 1 cm. For each case, 100 different situations were created. So, 100 steps of height by 8 steps of  $r_0$  value by 100 samples for each situation gives in the total the 80,000 simulated images. The validation dataset consisted of 6000 simulated images where the turbulence layer varied from 0 m to 15,000 m of height in steps of 1000 m, having the same weight as in the previous case. The  $r_0$  value ranged from 8 to 16 cm per altitude in steps of 1 cm. For this dataset 50 simulations were made for each configuration.

The three test datasets were simulated with an atmospheric turbulence with the same layers structure as the explained before, excepting the heights' step, it was modified being the test one of 500 m of altitude. The  $r_0$  value was set the same value for all the simulations of the set.

Once the shared parameters of the simulations for all the studies have been described, the own features of the FCNs and the datasets of simulations used for each one are detailed in the next subsections. Several FCNs with different topologies were tested before the ones selected for each case. The input parameters of the FCNs, presented differences for each case studied, that could be quantified as 5 times bigger of the inputs size, from the smallest to the biggest case considered, which prevented the use of a single topology of the FCN.

#### 2.4.1. Influence of the number of subapertures

For the analysis of the influence of the number of subapertures, four SH configurations were employed, with 15 × 15, 10 × 10, 7 × 7 and 3 × 3 subapertures, respectively. To compare the influence of the subapertures, we keep constant the total number of pixels of the input image. The configurations of the SH compared are shown in the Table 1.

Two different topologies of FCNs were used for the four cases as is showed in Table 1, since there were two possible inputs shapes 360 × 360 and 336 × 336 pixels. The total number of pixels, the number of subapertures, the number of pixels per subaperture and the total number of pixels of the phase image must satisfy some relations according to the DASP

simulation platform to carry out the simulation [22]. Only few combinations are allowed so they were selected to obtain the final input shape as similar as possible to perform the most reliable comparison, giving rise in the two input shapes mentioned above.

The first two cases used the same network as the input shape was 360 × 360 pixels in both, and another one was used for the last two with an input shape of 336 × 336 pixels:

- The Model 1 is formed by a convolutional block with 6 layers with kernels of size 11 × 11, 7 × 7, 5 × 5, 5 × 5, 3 × 3 and 3 × 3, respectively without pooling layers, and a deconvolutional block that is formed by another 6 layers with kernels 3 × 3 and the rest 5 layers, 5 × 5, respectively. Padding is added in all the layers and the hyperbolic tangent is used as activation function. The input image is reduced from 360 × 360 to 5 × 5 pixels in the convolutional block by using strides of shape 2 × 2 or 3 × 3. Then the image is increased to its final shape of 90 × 90 pixels by the deconvolutional block by using strides of shape 2 × 2 and 3 × 3.
- The Model 2 is similar to the first one in all the parameters, except on the strides of the layers. In this case, the input image is reduced from 336 × 336 to 7 × 7 by using strides of shape 2 × 2 and 3 × 3 and then the image is increased to 84 × 84 pixels with strides of shape 3 × 3 and 2 × 2.

#### 2.4.2. Influence of the number of pixels

New datasets were simulated to analyze the influence of the number of pixels per subaperture. They all consist in three subapertures and the same FOV but varying the total number of pixels. All the configurations selected are shown in Table 2.

To check the performance on the simplest configuration of the previous subsections, 3 subapertures were chosen. The first case selected is 3 × 3 subapertures with 112 × 112 pixels per subaperture, where the datasets from the previous point were used and it allow us to reuse the FCN that had shown a good performance. This selection outnumbers the usual configuration used in a real SH, so it was considered the biggest case as the other ones had less pixels; 48 × 48 and 28 × 28, respectively.

Unlike the previous subsection, for this case a unique model of FCN with its own topology was made for each configuration; the simulated images were so different in size terms so the same network could not be used, as it could only be optimized for one of the situations. The selected models are explained behind:

- Model 2 is the same as the used in the previous experiment with the same name as the datasets are the same too. To sum up, it consists in an FCN with 6 convolutional layers and 6 deconvolutional layers without pooling layers.
- Model 3 is similar to the first one, but the kernels and strides of the layers were modified according to the new input and output sizes of the images. It is composed by the same number of layers, with kernels of size 9 × 9, 7 × 7, 5 × 5 and all the last ones with size 3 × 3, respectively for the convolutional block. The deconvolutional block is formed by 6 deconvolutional layers with kernels of size 3 × 3 for the first layer and 5 × 5 for the rest. Images are reduced from 144 × 144 pixels to 3 × 3 pixels using strides of size 2 × 2 and 3 × 3 and then an increase is made to obtain outputs of 36 × 36 pixels of phase by using strides in the deconvolutional block of the same shape. Hyperbolic



**Table 2**  
Shack-Hartmann configurations used to compare the influence of the number of pixels per subaperture.

Pixels per subaperture	112 × 112 pixels	48 × 48 pixels	28 × 28 pixels
Total Pixels of the SH	336 × 336	144 × 144	84 × 84
Total number of pixels of the phase fcn model	Model 2	Model 3	Model 4

**Table 3**  
Shack-Hartmann configurations used to compare the influence of the FOV. All the simulations that have been made for previous comparisons were remade with 10 arcseconds of FOV. The FCN's models correspond with the ones used in previous studies.

	15 × 15 subapertures	10 × 10 subapertures	7 × 7 subapertures	3 × 3 subapertures		
Pixels per subaperture	20 × 20	28 × 28	42 × 42	24 × 24	48 × 48	96 × 96
Total number of pixels of the SH	300 × 300	280 × 280	294 × 294	72 × 72	144 × 144	288 × 288
Total number of pixels of the phase fcn model	Model 1	Model 2	Model 3	Model 4		

tangent function is used as activation function as in the previous models.

- The Model 4 is the smallest, since the input image had size of 84 × 84 pixels. Model 4 is composed by 5 convolutional layers, so only one less layer than the others, and kernels of size 9 × 9, 7 × 7, 5 × 5, 3 × 3 and 3 × 3, respectively, followed by 4 deconvolutional layers with kernels 3 × 3 the first one and 5 × 5 all the rest. In this case the input image is reduced to 7 × 7 pixels by applying strides of size 2 × 2 and 3 × 3. Then, it is increased by the deconvolutional block by using strides of 3 × 3 pixels in the second layer. The activation function selected for this case is the hyperbolic tangent function.

**2.4.3. Influence of the FOV**

New simulations were generated to determine the influence of the FOV of the observation. All the previous studies were made with simulations of 6 arcseconds of FOV. So, to take advantage of the previous results, all the simulations were remade with 10 arcseconds as it allowed us to compare case by case the influence of the FOV with the previous selected configurations.

The FCNs used to evaluate the influence of the FOV were the same as in previous subsections for each case. Due to the limitations of the DASP simulator, the shape of the FCN's inputs was different to previous cases as they had to obey some relations between the FOV, the number of pixels and the number of subapertures. Anyway, we decided to maintain the characteristics of the FCNs to make a realistic comparison since the simulations were similar enough in terms of the inputs shape to allow us to use the same topology. The configurations of the simulations are shown in Table 3 with the chosen model of FCN from each one.

The test datasets used in the comparison had the same characteristics as in the previous situations, they were formed by 3 datasets for each simulation, having fixed  $r_0$  values of 8, 10 and 12 cm each one.

**2.5. Residual WFE**

The residual RMSE WFE is the chosen magnitude to evaluate the quality of the reconstructions. The residual WFE allows to measure how similar are two pictures by calculating the RMSE of the difference pixel by pixel of both images. In this case, as the images correspond to the turbulence's phase, it is expressed in wavelengths units. The surface obtained by the difference pixel by pixel of two identical images has a zero residual WFE value, as it increases, the difference between images also increases. The error can be calculated as follows:

$$\text{residual WFE} = \sqrt{\frac{\sum_{i=1}^N (x_i - y_i)^2}{N}}$$

Being  $x_i$  the pixels of the original image,  $y_i$  the reconstructed ones and  $N$  the total number of pixels of both images.

The residual WFE is often presented in radians, in those cases it is calculated as:

$$\text{residual WFE (rad)} = \frac{2\pi}{\lambda} \sqrt{\frac{\sum_{i=1}^N (x_i - y_i)^2}{N}}$$

Where  $\lambda$  represents the wavelength of the incoming light.

**3. Results and discussion**

The results are divided in three subsections, to show the three objectives of the research: the influence of the number of subapertures, the number of pixels per subaperture and the FOV of the observation.

Results from all the cases are given considering the mean of value of the residual WFE and with the percentage of the similarity between the phase reconstructed by the FCN and the original phase, the one simu-

**Table 4**  
Comparison of the influence of the variation on the number of subapertures. The most turbulent layer for each case is shadowed in light grey, whose intensity increases as the intensity of the turbulence decreases.

Subapertures	$r_0$ (cm)	Residual WFE (nm)	Similarity (%)	WFE original (nm)	WFE Network (nm)
15	8	304.79	72	621.52	445.65
	10	235.56	80	516.47	411.43
	12	206.91	85	443.26	376.41
10	8	249.88	91	526.82	477.48
	10	206.11	94	437.63	410.63
	12	180.65	96	375.62	359.7
7	8	204.52	93	662.9	615.95
	10	155.18	96	550.69	526.82
	12	101.07	97	472.7	458.38
3	8	202.93	91	684.39	623.91
	10	146.43	96	568.2	542.73
	12	93.90	94	499.76	471.11

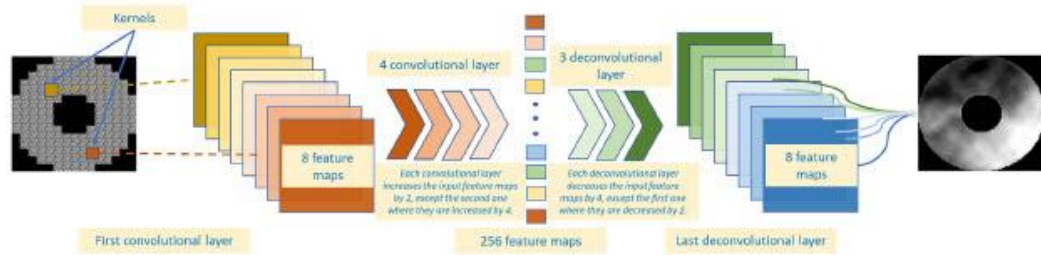


Fig. 5. Schematic representation of the topology of the FCN selected as Model 4. The difference between the other models apart from the sizes of their kernels and strides is that Model 4 has one less convolutional layer and only four deconvolutional layers.

lated by DASP. Some figures of the reconstructed and the original phase are also shown for a visual comparison.

### 3.1. Influence of the number of subapertures

In Table 4 the quality of the reconstructions performed by the FCN for each SH's configuration for different intensity of turbulence layers is compared. Note that all the  $r_0$  values are very low, so the best situation with 12 cm of  $r_0$  is still a significantly turbulent situation. It is expected, then, that errors obtained were high even the reconstructor made a well performance. An example of the reconstruction made by the FCN is shown in Fig. 5.

The results show that the networks improve the quality of the reconstructions when the number of subapertures decreases. There also exists a clear trend in the similarity of the reconstructions, which increases from 72% to 93% when the subapertures are reduced from  $15 \times 15$  to  $7 \times 7$ , less than a half, at high atmosphere turbulence conditions. Also, note that  $15 \times 15$  subapertures is a number that could be found perfectly in a reconstructor system of a solar telescope of 1.5 m of pupil, as the one selected in the simulations.

Another important parameter that has not been mentioned yet is the time taken by the network to perform the reconstruction. For the cases showed above, practically there are not differences being always the mean computational time needed per sample over each test under 4.6 ms. The test that showed the lowest value was the one that considers 7 subapertures one with  $r_0$  value of 10 cm, where the recall time mean value over that test was 4.43 ms. The slowest one was 10 subapertures with 8 cm of  $r_0$  that needs 4.54 ms per sample.

Therefore, models 1 and 2 show similar execution times, being the second one slightly faster since its inputs are smaller than those of the first model.

When using FCNs, all the information of the image given as input is analyzed, instead of using the standard SH method, where only 2 values, the slope information, can be recovered from each subaperture. This is an advantage because no information of the input is lost. That fact allows

the FCNs to achieve good performance in the reconstructions without a high number of subapertures. Furthermore, the input of the FCN is the complete image with all the subapertures, where there are abrupt changes in the boundaries of the subapertures since each one is saturated by the Sun's image (see Fig. 2). These limits are misunderstood by the FCN causing that the system achieves better reconstructions with less subapertures, contrary to the working of other reconstruction methods.

The trend mentioned before is not saw when the results for  $7 \times 7$  and  $3 \times 3$  subapertures are considered. Despite with  $7 \times 7$  subapertures there are more subaperture's limits, the information received by an SH with only  $3 \times 3$  subapertures in total is not enough to reconstruct the turbulence's phase in its all regions, as the FCN does not have enough information of the aberrations received due to not have sufficient subapertures to compare them.

Moreover, the less is the intensity of the turbulence, the higher is the similarity of the reconstruction with the original image. All the cases correspond with extreme conditions, especially  $r_0 = 8\text{cm}$ . So, it is expected that under lighter circumstances it would work better. That fact is very interesting for close loop adaptive optic systems, as the images received are firstly corrected with the information of the aberrations of the previous moment before being received by the WFS, so the intensity of the aberrations measured by the SH are much slower.

### 3.2. Influence of the number of pixels

The quality of the reconstructions was measured using three test datasets with the same characteristics as in the first study, with  $r_0$  values of 8, 10 and 12 cm each one. Results are shown in Table 5.

The importance of the number of subapertures lies in the information received from the SH. The higher number of pixels, the more information extracted by the FCN. Consequently, the phase may be reconstructed with more precision. In the cases of high values of  $r_0$ , the atmosphere is more uniform, so reconstructions can be made even having low number of pixels.

Table 5  
Results of the comparison of the influence of the variation on the number of pixels per subaperture. The most turbulence layer for each case is coloured in light grey, and the colour becomes more intense as the intensity of the turbulence decreases.

Pixel per subap.	$r_0$ (cm)	Residual WFE (nm)	Similarity (%)	WFE original (nm)	WFE Network (nm)
112	8	208.50	91	684.39	619.93
	10	151.20	95	568.20	541.94
	12	130.51	94	499.76	471.11
48	8	179.85	89	634.25	567.40
	10	126.53	97	526.02	510.90
	12	105.05	98	452.01	444.85
24	8	213.27	84	615.15	518.06
	10	124.94	97	510.11	494.19
	12	101.07	99	438.48	432.91



**Table 6**  
Results obtained for reconstructions made with simulations with 10 arcseconds of FOV. All the simulations of previous subsections were remade with 10 arcseconds to compare case by case the influence of the FOV with the previous selected configurations.

Subapertures	Pixel per subaperture	r0(cm)	Residual WFE (nm)	Similarity (%)	WFE original (nm)	WFE Network (nm)
15	20	8	217.25	88	591.28	518.06
		10	167.12	96	491.01	472.70
		12	147.22	97	421.77	432.12
10	28	8	257.04	88	666.88	587.30
		10	190.20	94	553.88	522.04
		12	165.53	96	475.89	457.58
7	42	8	222.82	93	635.84	592.07
		10	175.08	95	528.41	503.74
		12	150.41	97	453.60	439.28
3	96	8	154.38	94	582.52	547.51
		10	112.21	97	483.85	468.73
		12	94.70	98	415.41	405.86
	48	8	164.73	93	637.43	590.48
		10	118.57	97	529.21	512.49
		12	98.68	98	454.40	444.06
	24	8	139.26	91	650.17	592.87
		10	77.99	97	539.55	530.80
		12	59.68	98	463.15	458.38

Analysing the results; they show that the quality of reconstructions made by ANNs is influenced positively by the total number of pixels of the SH, as expected. In Table 5 can be seen how the original phase and the one retrieved by the FCN are more similar when using higher number of pixels, especially in the cases of worst turbulence conditions. For  $r_0$  values of 12 cm, there is practically no difference when varying the number of pixels between 122 per subaperture and 24, but it is expected that if the number of subapertures is reduced even more, the similarity will decrease, as the amount of information received by the FCN may not be enough. However, when the  $r_0$  value represents high turbulence conditions, as  $r_0 = 8$  cm, it is seen how the quality of the reconstruction decreases around 7% between the case with more subapertures in comparison to the one with less subapertures.

As in the previous subsection the computational time needed was also analyzed for the 3 ANN models. Every value given is calculated as the mean time needed per sample over all the tests. The ANN Model 3 was applied for the case with  $48 \times 48$  pixels per subaperture, obtaining a mean computational time of 4.41 ms, 4.39 ms, and 4.43 ms for 8 cm, 10 cm and 12 cm of  $r_0$ , respectively. The ANN Model 4 was applied for the last case with  $24 \times 24$  pixels per subaperture, the mean times needed per sample were 4.18, 4.22 and 4.23 ms for 8, 10 and 12 cm of  $r_0$ .

Comparing these results with the previous subsection the only case that showed an improvement in terms of computational time is the  $24 \times 24$  pixels per subaperture one, being approximately 4% faster than the big one of  $112 \times 112$  pixels per subaperture. It may be due to the smaller size of the input images or to the topology of the ANN model, as Model 4 is composed by one less convolutional and deconvolutional layer. From this study can be concluded that the number of layers of the ANN model affects the computational time, since the intermediate case of  $48 \times 48$  pixels per subaperture obtains similar computational times to the bigger one, despite being half the size.

### 3.3. Influence of the FOV of the observation

The influence of the FOV in the reconstructions made by FCNs is analyzed comparing the results obtained with 10 arcseconds of FOV with the results obtained in previous subsections, all of them obtained over simulations with 6 arcseconds of FOV. The quality of the reconstructions made with the new datasets can be seen in Table 6 as the results obtained with 6 arcseconds of FOV are shown in Tables 4 and 5.

To analyze the results, the cases with 8 cm of  $r_0$  are principally commented since they are where most differences are obtained. Beginning with the case with most subapertures, for an SH of  $15 \times 15$  subapertures the similarity of the reconstruction is improved from 72% with 6 arc-

seconds of FOV to 88% with 10 arcseconds when using the same FCN model. It should be noted that in the new simulations, the number of pixels per subaperture has been reduced from 28 to 20 due to DASP simulator relations. A better reconstruction is obtained even initially having a worse situation. To sum up, in this case increasing the FOV of the simulation from 6 arcseconds to 10, the quality of the reconstruction has improved a 16% even reducing the number of pixels per subaperture, which implies worse reconstructions as it has been discussed in the previous subsection.

In the case of 20 subapertures, the result has decreased from 91% to 89% but as before, the number total of pixels of the SH with 10 arcseconds of FOV was 280 versus the 360 pixels with 6 arcseconds of FOV, in this case the worse result is attributed to the reduction of the number of pixels. So, if the quality of reconstructions is almost the same despite that great difference in pixels, it is possible to reaffirm that the FCN reconstruct better with higher values of FOV.

As results of 7 subapertures shows there are practically no differences between the two situations. It is expected that the reconstruction will improve for the FCN in this case since it is made with less pixels per subapertures again, 42 versus 48, that suppose in total 294 vs 336.

In the cases made with SH of 3 subapertures the same tendency is observed, as for all the cases a better reconstruction is retrieved, even in the biggest one, where despite having 16 pixels less per subaperture (a 14% less), the similarity is improved from 91% to 93% for an  $r_0$  value of 8 cm.

Finally, from situations with  $r_0$  values higher than 8 cm, the quality of the reconstructions is approximately the same with 6 or 10 arcseconds. In some cases, as in the last one with 3 subapertures, the similarity is even higher for the situations with 10 arcseconds. So, in conditions of normal intensity turbulence, the new reconstructor can achieve phases of the turbulence closer to the original ones in simulations with 10 arcseconds of FOV despite having less pixels per subaperture.

## 4. Conclusions

A new reconstructor system for Solar AO based on FCNs has been tested throughout this paper. The influence on its reconstructions of some observation's parameters as the number of subapertures of the SH, the number of pixels per subaperture and the FOV of the observation has been studied to know the benefits of the new reconstructor and how the most suitable configurations are (Fig. 6).

The first conclusion that can be extracted from the study is that a better reconstruction is not due to have more subapertures as happens with commonly used reconstructor systems in Solar AO. The FCN only



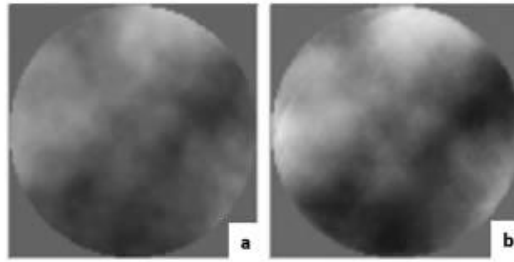


Fig. 6. Example of the reconstruction made by the FCN for a  $7 \times 7$  subapertures SH. Fig. 5.a represents the original image (simulated with the DASP simulator) and Fig. 5.b the final reconstructed phase obtained by the FCN. It corresponds to a situation of  $r_0 = 8$  cm. It is important to note how the reconstructed phase has approximately the same shape, having the turbulence in the same region of the image, as the original one.

Table 7  
Configuration with best results obtained from the cases studied. A higher optimized configuration can be modelled using the results extracted from each of the cases studied.

Parameters with best results from the cases evaluated in the study		
Number of subapertures	Pixels per subaperture	FOV
$7 \times 7$	$42 \times 42$	10

need enough subapertures to interpret the turbulence in the different regions of the pupil but, from that number on, stems to the characteristics of Solar observations, the FCN misunderstands the information received and the quality of the reconstructions decreases. In our study the most adequate number was  $7 \times 7$  subapertures.

In previous works made by our group, we had obtained that LS working with an SH with these characteristics obtains a mean similarity in the reconstructions of 85% over a dataset of 6.000 images for  $r_0$  values of 8 cm. Consequently, it is important to highlight that our new reconstructor improves the reconstructions an 8% in terms of similarity using SHs with less than a half subapertures than the most used one.

From the second study made, it can be concluded that the number of pixels per subaperture of the SH favourably affects to the quality of the reconstruction made. As all the information received is used by the FCNs, the higher number of pixels have the subapertures, the better reconstructions will make the ANN. Having more pixels implies that the FCN receives more information of the atmosphere turbulence conditions at each moment. That fact becomes more important in worse atmospheric turbulence situations, where the biggest differences in the similarity of the reconstructions can be seen from the study. Anyway, in economic terms, it is more interesting to increase the number of pixels per subaperture than the subapertures of the Shack-Hartmann.

On the other hand, increasing the number of pixels implies more computational time for the reconstructions if that means growth the number of layers of the topology of the ANN. However, variations in computational time are not relevant as the maximum difference appreciated was a 4% and for all the cases ANNs have showed very competitive times of execution, being always lower than 4.6 ms.

Finally, the last research provides that the FCNs achieves better results in observations when the FOV is higher. The higher FOV in the observation, the more similar reconstructions made. In the situations studied through that subsection, the FCN got similar results between 6 and 10 arcseconds of FOV or even better with 10 arcseconds despite using in the last case simulations with less pixels per subaperture. In Table 7 is presented the configuration that has shown the best results from the cases studied.

The final subsection allows us to confirm the conclusions obtained in the previous research, as all the simulations of them were remade with 10 arcseconds of FOV for it. The trends observed of the influence of the number of subapertures and of the number of pixels per subaperture were also fulfilled in this subsection.

The  $r_0$  values used for testing the FCNs is something that should be noted, since all of them represents bad turbulence conditions, especially the cases with 8 cm as it is associated with extremely bad conditions. Seeing the trend of all the cases studied, the similarity increases with the  $r_0$  value, so it can be expected that the quality of the reconstructions improves under normal viewing conditions.

Future developments of this research will consist of apply the ANNs in an optical bench or a telescope, to verify that the conclusions drawn are still valid working with real data. Furthermore, the reconstructor could be developed for more complex AO systems, as the MCAO one since they are the most used currently in solar telescopes.

#### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### CRediT authorship contribution statement

Francisco García Riesgo: Methodology, Investigation, Software, Writing – original draft. Sergio Luis Suárez Gómez: Conceptualization, Data curation, Writing – review & editing. Jesús Daniel Santos Rodríguez: Methodology, Writing – review & editing. Carlos González Gutiérrez: Conceptualization, Software. Yolanda Martín Hernando: Software. Luz María Montoya Martínez: Visualization, Conceptualization, Writing – review & editing. Andrés Asensio Ramos: Methodology, Validation, Writing – review & editing. Manuel Collados Vera: Supervision, Writing – review & editing. Miguel Núñez Cagigal: Conceptualization, Writing – review & editing. Francisco Javier De Cos Juez: Visualization, Writing – review & editing, Supervision.

#### References

- [1] Tyson R. Principles of adaptive optics. CRC Press; 2010.
- [2] Collados M, et al. European solar telescope: progress status. *Astron Nachr* 2010;331(6):609–14.
- [3] Scharmer GB, Detoni PM, Lofdahl MG, Shand M. In: Keil SL, Avakyan SV, editors. Innovative Telescopes and Instrumentation for Solar Astrophysics, 4853. Bellingham, WA, USA: Society of Photo Optical; 2003. p. 370–81.
- [4] Scharmer GB, Bjelkejo K, Korhonen TK, Lindberg B, Petterson B. In: Keil SL, Avakyan SV, editors. Innovative Telescopes and Instrumentation for Solar Astrophysics, 4853. Bellingham, WA, USA: Society of Photo Optical; 2003. p. 341–51.
- [5] Mirowski PW, LeCun Y, Madhavan D, Kuzniecky R. Comparing SVM and convolutional networks for epileptic seizure prediction from intracranial EEG. In: Proceedings of the IEEE workshop on machine learning for signal processing; 2008. p. 244–9.
- [6] Nagi J, et al. Max-pooling convolutional neural networks for vision-based hand gesture recognition. In: Proceedings of the IEEE international conference on signal and image processing applications (ICSIPA); 2011. p. 342–7.
- [7] Osborn J, et al. Open-loop tomography with artificial neural networks on CANARY: on-sky results. *Mon Not R Astron Soc* 2014;441(3):2508–14.
- [8] Gómez SLS, Gutiérrez CG, Rodríguez JDS, Rodríguez MLS, Lasheras FS, de Cos Juez FJ. Analysis the performance of a tomographic reconstructor with different neural networks frameworks. *Adv Intell Syst Comput* 2017;557:1051–60.
- [9] Osborn J, et al. First on-sky results of a neural network based tomographic reconstructor: carmen on Canary. *Adaptive Optics Systems IV*, 9148. SPIE; 2014. 91484M p. 1541–6.
- [10] Zilberman A, Golbraikh E, Kopeika NS. Propagation of electromagnetic waves in Kolmogorov and non-Kolmogorov atmospheric turbulence: three-layer altitude model. *Appl Opt* 2008;47(34):6385–91.
- [11] Rimmele TR. Solar adaptive optics. *Adapt Opt Syst Technol* 2000;4007:218–32.
- [12] Rimmele TR, Marino J. Solar adaptive optics. *Living Rev Sol Phys* 2011;8(1):2.
- [13] Ellerbroek BL. First-order performance evaluation of adaptive-optics systems for atmospheric-turbulence compensation in extended-field-of-view astronomical telescopes. *JOSA A* 1994;11(2):783–805.
- [14] Roggemann MC. Optical performance of fully and partially compensated adaptive optics systems using least-squares and minimum variance phase reconstructors. *Comput Electr Eng* 1992;18(6):451–66.
- [15] Riesgo FG, Gómez SLS, Lasheras FS, Gutiérrez CG, San Cristóbal CP, de Cos Juez FJ. Convolutional CARMEN: tomographic Reconstruction for Night Observation. In:

- Pérez García H, Sánchez González I, Castejón Lamas M, Quintián Pardo H, Corchado Rodríguez E, editors. Proceedings of the international conference on hybrid artificial intelligence systems. Lecture Notes in Computer Science, 11734. Springer; 2019. p. 335–45.
- [16] Berkefeld T. Solar adaptive optics. In: Kneer Franz, Puschmann Klaus, Witman Axel, editors. Modern solar facilities-advanced solar science. Proceedings of a Workshop Held at Göttingen, September 27-29. Universitätsverlag Göttingen; 2007. p. 107–13.
- [17] Knutson PA, Omer-Peterson M, Dainty C. Extended object wavefront sensing based on the correlation spectrum phase. *Opt Express* 2005;13(23):9527–36.
- [18] Gardner MW, Dorling SR. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmos Environ* 1998;32(14):2627–36.
- [19] Goodfellow I, Bengio Y, Courville A, Bengio Y. Deep learning, 1. MIT Press Cambridge; 2016.
- [20] García Riesgo F, et al. Early fully-convolutional approach to wavefront imaging on solar adaptive optics simulations. In: Proceedings of the international conference on hybrid artificial intelligence systems; 2020. p. 674–85.
- [21] Shelhamer E, Long J, Darrell T. Fully Convolutional Networks for Semantic Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 2017;39(4):640–51.
- [22] Baden AG, Bharmal NA, Jenkins D, Morris TJ, Osborn J, Peng J, et al. The Durham Adaptive Optics Simulation Platform (DASP): Current status. *SoftwareX* 2018;7:63–9.
- [23] Abadi M, et al. TensorFlow: a system for large-scale machine learning. *OSDI* 2016;16:265–83.
- [24] Gulli A, Pal S. Deep learning with keras. Packt Publishing Ltd; 2017.
- [25] Berkefeld T, Schmidt D, Soltan D, von der Lühe O, Heidecke F. The GREGOR adaptive optics system. *Astron Nachr* Nov 2012;333(9):863–71.





## 12 INFORME DEL FACTOR DE IMPACTO

---

### Primer artículo

Overview and Choice of Artificial Intelligence Approaches for Night-Time Adaptive Optics Reconstruction

Mathematics, 2021

Factor de impacto: 2.592

21/333 Mathematics. Q1

### Segundo artículo

Fully Convolutional Approaches for Numerical Approximation of Turbulent Phases in Solar Adaptive Optics

Mathematics, 2021

Factor de impacto: 2.592

21/333 Mathematics. Q1

### Tercer artículo

Comparative study of Shack-Hartmann configurations for atmospheric turbulence reconstructions in solar adaptive optics

Optics and Lasers in Engineering, 2022

Factor de impacto 5.666

18/101 Optics. Q1



## 13 REFERENCIAS

---

- [1] T. R. Rimmele and J. Marino, "Solar adaptive optics," *Living Rev. Sol. Phys.*, vol. 8, no. 1, p. 2, 2011.
- [2] R. Tyson, *Principles of adaptive optics*. CRC press, 2010.
- [3] D. Rouan, "Adaptive Optics," in *Encyclopedia of Astrobiology*, R. Amils, M. Gargaud, J. Cernicharo Quintanilla, H. J. Cleaves, W. M. Irvine, D. Pinti, and M. Viso, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014, pp. 1–2.
- [4] A. Zilberman, E. Golbraikh, and N. S. Kopeika, "Propagation of electromagnetic waves in Kolmogorov and non-Kolmogorov atmospheric turbulence: three-layer altitude model," *Appl. Opt.*, vol. 47, no. 34, p. 6385, Dec. 2008.
- [5] P. Hickson, "Fundamentals of Atmospheric and Adaptive Optics," The University of British Columbia, 2008.
- [6] S. M. Salim, R. Buccolieri, A. Chan, and S. Di Sabatino, "Numerical simulation of atmospheric pollutant dispersion in an urban street canyon: Comparison between RANS and LES," *J. Wind Eng. Ind. Aerodyn.*, vol. 99, no. 2–3, pp. 103–113, 2011.
- [7] U. Frisch and A. Kolmogorov, *Turbulence: the legacy of AN Kolmogorov*. .
- [8] F. M. Hoblit, "Gust loads on aircraft: concepts and applications," 1988.
- [9] T. R. Rimmele, "Solar adaptive optics," in *Adaptive Optical Systems Technology*, 2000, vol. 4007, pp. 218–232.
- [10] E. Golbraikh, H. Branover, N. S. Kopeika, and A. Zilberman, "Non-Kolmogorov atmospheric turbulence and optical signal propagation," *Nonlinear Process. Geophys.*, vol. 13, no. 3, pp. 297–301, 2006.
- [11] J. W. Hardy, *Adaptive optics for astronomical telescopes*, vol. 16. Oxford University Press on Demand, 1998.
- [12] R. K. Tyson, *Introduction to adaptive optics*, vol. 41. SPIE press, 2000.
- [13] T. Rimmele and J. Marino, "Solar Adaptive Optics," *Living Rev. Sol. Phys.*, vol. 2, 2011.
- [14] P. Hickson, "Atmospheric and adaptive optics," *Astron. Astrophys. Rev.*, vol. 22, no. 1, pp. 1–38, Oct. 2014.
- [15] K. Atkinson and W. Han, "Numerical solution of fredholm integral equations of the second kind," in *Theoretical Numerical Analysis*, Springer, 2009, pp. 473–549.
- [16] M. Johansson, "The hilbert transform," *Math. Master's Thesis. Växjö Univ. Suecia. Dispon. en internet [http://w3.msi.vxu.se/exarb/mj\\_ex.pdf](http://w3.msi.vxu.se/exarb/mj_ex.pdf)*, Consult. el, vol. 19, 1999.
- [17] R. K. Tyson and P. L. Wizinowich, "Principles of Adaptive Optics," *Phys. Today*, vol. 45, no. 2, pp. 100–100, Feb. 1992.
- [18] S. Klassen, "The photoelectric effect: Reconstructing the story for the physics classroom," *Science & Education*, vol. 20, no. 7. Springer, pp. 719–731, 2011.
- [19] C. Wu, J. Ko, and C. C. Davis, "Determining the phase and amplitude distortion of a wavefront using a plenoptic sensor," *J. Opt. Soc. Am. A*, vol. 32, no. 5, p. 964, May 2015.
- [20] L. Seifert, H. J. Tiziani, and W. Osten, "Wavefront reconstruction with the adaptive Shack-

- Hartmann sensor," *Opt. Commun.*, vol. 245, no. 1–6, pp. 255–269, 2005.
- [21] B. C. Platt and R. Shack, "History and principles of Shack-Hartmann wavefront sensing," *J. Refract. Surg.*, vol. 17, no. 5, pp. S573--S577, 2001.
- [22] F. Roddier, *Adaptive optics in astronomy*. Cambridge university press, 1999.
- [23] D. Guzmán, F. J. de Cos Juez, F. S. Lasheras, R. Myers, and L. Young, "Deformable mirror model for open-loop adaptive optics using multivariate adaptive regression splines," *Opt. Express*, vol. 18, no. 7, pp. 6492–6505, 2010.
- [24] P.-Y. Madec, "Overview of deformable mirror technologies for adaptive optics and astronomy," no. March, pp. 844705-844705–18, 2012.
- [25] P.-Y. Madec, "Overview of deformable mirror technologies for adaptive optics and astronomy," in *Adaptive Optics Systems III*, 2012, vol. 8447, pp. 22–39.
- [26] E. J. Fernández and P. Artal, "Membrane deformable mirror for adaptive optics: performance limits in visual optics," *Opt. Express*, vol. 11, no. 9, pp. 1056–1069, 2003.
- [27] G. Dai, "Modal wave-front reconstruction with Zernike polynomials and Karhunen–Loève functions," *J. Opt. Soc. Am. A*, vol. 13, no. 6, p. 1218, 2008.
- [28] S. Hu, B. Xu, X. Zhang, J. Hou, J. Wu, and W. Jiang, "Double-deformable-mirror adaptive optics system for phase compensation," *Appl. Opt.*, vol. 45, no. 12, pp. 2638–2642, 2006.
- [29] R. J. Noll, "Zernike polynomials and atmospheric turbulence," *JOsA*, vol. 66, no. 3, pp. 207–211, 1976.
- [30] V. Lakshminarayanan and A. Fleck, "Zernike polynomials: a guide," *J. Mod. Opt.*, vol. 58, no. 7, pp. 545–561, 2011.
- [31] A. Basden *et al.*, "DASP the Durham Adaptive optics Simulation Platform: Modelling and simulation of adaptive optics systems," *SoftwareX*, 2018.
- [32] A. G. Basden *et al.*, "The Durham Adaptive Optics Simulation Platform (DASP): Current status," *SoftwareX*, vol. 7, pp. 63–69, 2018.
- [33] J. N. Kok, E. J. Boers, W. A. Kosters, P. der Putten, and M. Poel, "Artificial intelligence: definition, trends, techniques, and cases," *Artif. Intell.*, vol. 1, pp. 270–299, 2009.
- [34] M. A. Nielsen, *Neural Networks and Deep Learning*. 2015.
- [35] S. Theodoridis, *Neural Networks and Deep Learning*. 2015.
- [36] Keras Special Interest Group, "Keras Layer Activation Functions," 2021. [Online]. Available: <https://keras.io/api/layers/activations/>.
- [37] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.
- [38] R. Rojas, "The backpropagation algorithm," in *Neural networks*, Springer, 1996, pp. 149–182.
- [39] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*, vol. 1. MIT press Cambridge, 2016.
- [40] M. Minsky and S. A. Papert, *Perceptrons, Reissue of the 1988 Expanded Edition with a new foreword by Léon Bottou: An Introduction to Computational Geometry*. MIT press, 2017.

- [41] A. Zhang, Z. C. Lipton, M. Li, and A. J. Smola, "Dive into deep learning," *arXiv Prepr. arXiv2106.11342*, 2021.
- [42] R. A. Horn, "The hadamard product," in *Proc. Symp. Appl. Math*, 1990, vol. 40, pp. 87–169.
- [43] K. O'Shea and R. Nash, "An introduction to convolutional neural networks," *arXiv Prepr. arXiv1511.08458*, 2015.
- [44] J. Wu, "Introduction to convolutional neural networks," *Natl. Key Lab Nov. Softw. Technol. Nanjing Univ. China*, vol. 5, no. 23, p. 495, 2017.
- [45] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.
- [46] J. Long, E. Shelhamer, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation."
- [47] J. Brownlee, *Deep learning with Python: develop deep learning models on Theano and TensorFlow using Keras*. Machine Learning Mastery, 2016.
- [48] J. Osborn *et al.*, "Using artificial neural networks for open-loop tomography," *Opt. Express*, vol. 20, no. 3, p. 2420, Jan. 2012.
- [49] J. Osborn *et al.*, "Open-loop tomography with artificial neural networks on CANARY: On-sky results," *Mon. Not. R. Astron. Soc.*, vol. 441, no. 3, pp. 2508–2514, 2014.
- [50] F. García Riesgo, S. L. Suárez Gómez, J. D. Santos, E. Diez Alonso, and F. Sánchez Lasheras, "Overview and Choice of Artificial Intelligence Approaches for Night-Time Adaptive Optics Reconstruction," *Mathematics*, vol. 9, no. 11, p. 1220, 2021.
- [51] "AOTools," 2019. [Online]. Available: <https://aotools.readthedocs.io/en/v1.0/index.html>. [Accessed: 30-Mar-2022].
- [52] F. G. Riesgo, S. L. S. Gómez, E. D. Alonso, C. González-Gutiérrez, and J. D. Santos, "Fully Convolutional Approaches for Numerical Approximation of Turbulent Phases in Solar Adaptive Optics," *Mathematics*, vol. 9, no. 14, 2021.
- [53] E. Sidick, J. J. Green, R. M. Morgan, C. M. Ohara, and D. C. Redding, "Adaptive cross-correlation algorithm for extended scene Shack-Hartmann wavefront sensing.," *Opt. Lett.*, vol. 33, no. 3, pp. 213–5, 2008.
- [54] F. G. Riesgo *et al.*, "Early Fully-Convolutional Approach to Wavefront Imaging on Solar Adaptive Optics Simulations," in *International Conference on Hybrid Artificial Intelligence Systems*, 2020, pp. 674–685.
- [55] C. González-Gutiérrez, M. L. Sánchez-Rodríguez, J. L. Calvo-Rolle, and F. J. de Cos Juez, "Multi-GPU Development of a Neural Networks Based Reconstructor for Adaptive Optics," *Complexity*, vol. 2018, no. Intelligent Control Approaches for Modeling and Control of Complex Systems, pp. 1–9, 2018.
- [56] F. G. Riesgo *et al.*, "Comparative study of Shack-Hartmann configurations for atmospheric turbulence reconstructions in solar adaptive optics," *Opt. Lasers Eng.*, vol. 158, p. 107157, 2022.
- [57] T. Berkefeld, D. Schmidt, D. Soltau, O. von der Lühe, and F. Heidecke, "The GREGOR adaptive optics system," *Astron. Nachrichten*, vol. 333, no. 9, pp. 863–871, Nov. 2012.

- [58] F. García Riesgo, S. L. Suárez Gómez, F. Sánchez Lasheras, C. González Gutiérrez, C. Peñalver San Cristóbal, and F. J. de Cos Juez, “Convolutional CARMEN: Tomographic Reconstruction for Night Observation,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2019, vol. 11734 LNAI, pp. 335–345.
- [59] F. G. Riesgo, S. L. S. Gómez, F. S. Lasheras, C. G. Gutiérrez, C. P. San Cristóbal, and F. J. de Cos Juez, “Convolutional CARMEN: Tomographic Reconstruction for Night Observation,” in *International Conference on Hybrid Artificial Intelligence Systems*, 2019, pp. 335–345.
- [60] F. García Riesgo *et al.*, “Early Fully-Convolutional Approach to Wavefront Imaging on Solar Adaptive Optics Simulations,” in *International Conference on Hybrid Artificial Intelligence Systems*, 2020, pp. 674–685.