Modelos de Aprendizaje Profundo para la Estimación y Corrección de Aberraciones del Frente de Onda en Sistemas Hartmann–Shack, Integrando Arquitecturas CNN, ResNet18 con Mecanismos de Atención y Vision Transformers

1st Alejandro Becerra
Departamento de Ingeniería de Sistemas
Universidad de Antioquia
Medellín, Colombia
alejandro.becerraa@udea.edu.co

***Resumen*—Este proyecto busca mejorar la precisión en la estimación de aberraciones ópticas a partir de imágenes Hartmann–Shack, mediante el uso de modelos de aprendizaje profundo. Se desarrolla un conjunto de modelos como lo son, dos configuraciones arquitectónicas del modelo ResNet18 y un modelo Vision Transformer, con el fin de comparar los resultados para determinar el modelo óptimo en cuanto a su desempeño de reducción del error cuadrático medio (RMSE) y tiempo de ejecución para los coeficientes de Zernike del 1 al 21. Los resultados obtenidos hasta el momento muestran una alta correlación entre los coeficientes reales y los predichos, una reducción progresiva del RMSE y una distribución de residuos centrada en cero, lo que evidencia modelos estables y generalizables, teniendo en cuenta también que mejoran los resultados en comparación con modelos que no se entrenaron teniendo en cuenta los coeficientes uno y dos de los polinomios de Zernike en la etapa de prueba con imágenes diferentes a las de entranamiento y validación (Testing).**

***Palabras clave*—Imágenes de Hartmann–Shack, Coeficientes de Zernike, Polinomios de Zernike, Red Neuronal Convolucional, ResNet18, Censado de frente de onda, Teorema del Límite Central, Vision Transformer, framework.**

I. Introducción

Durante las últimas décadas, diversas investigaciones han sentado las bases teóricas y experimentales para el estudio y modelado de aberraciones ópticas, así como para la aplicación de métodos computacionales avanzados en su análisis. Desde los primeros trabajos sobre redes neuronales artificiales, se ha reconocido su capacidad para aprender representaciones complejas a partir de datos visuales, emulando el comportamiento del cerebro humano mediante el ajuste iterativo de parámetros internos [4].

En paralelo, el campo de la óptica oftálmica ha experimentado un avance significativo gracias a la caracterización matemática de las aberraciones, entendidas como desviaciones en la propagación del frente de onda que afectan la calidad de las imágenes formadas por los sistemas ópticos, incluido el ojo humano [1].

La estimación precisa de aberraciones ópticas mediante imágenes Hartmann–Shack representa un componente esencial en la óptica adaptativa, permitiendo analizar las desviaciones del frente de onda reflejado por el ojo humano [1]. Tradicionalmente, estos análisis se han basado en el cálculo de los polinomios de Zernike, que describen de forma matemática las irregularidades del sistema visual [2]; sin embargo, los métodos clásicos presentan limitaciones ante la complejidad y el volumen de los datos. En este contexto, las redes neuronales convolucionales (CNN) [3] surgen como una alternativa eficaz para extraer patrones espaciales y predecir parámetros ópticos con alta precisión.

Así mismo, se postulan los modelos Vision transformers como prometedores en la precisión a la hora de evaluar imágenes como señalan Zhang et al. [9], el uso del modelo Vision Transformer permite abordar el sensado compresivo de forma más eficiente. Este modelo se basa en la arquitectura propuesta por Dosovitskiy et al. [10], quienes introdujeron el Vision Transformer como una alternativa a las CNN tradicionales para tareas de visión por computadora. Este modelo se entrenará y evaluará su comportamiento en comparación con otros modelos convolucionales residuales ResNet18 para predecir los coeficientes de Zernike a partir de imágenes Hartmann–Shack, incluyendo los coeficientes 1 y 2, cuya incorporación mejora la representación de los desplazamientos lineales de los píxeles en la etapa de prueba (Testing). Con ello, se busca optimizar la reconstrucción del frente de onda y avanzar hacia sistemas de diagnóstico óptico más precisos y automatizados, fortaleciendo la relación entre el análisis físico y el modelado mediante aprendizaje profundo.

II. Descripción del problema

A. Contexto del problema

Las aberraciones ópticas, según Vidal [5], son desviaciones en la propagación del frente de onda que afectan la calidad

de las imágenes formadas por los sistemas ópticos, incluido el ojo humano. Estas aberraciones pueden clasificarse en aberraciones de bajo orden (como el desenfoque y el astigmatismo) y de alto orden (como la coma y el trébol), las cuales cambian dinámicamente en el tiempo y presentan variaciones individuales entre sujetos. Para su caracterización, el sensor de frente de onda Hartmann–Shack (SHWFS, por sus siglas en inglés) se ha consolidado como una herramienta ampliamente utilizada en aplicaciones oftálmicas, gracias a su simplicidad, sensibilidad y versatilidad [1]. Este sistema se basa en una matriz de microlentes que descompone el haz incidente en múltiples subaperturas, permitiendo registrar la desviación local del frente de onda a través del desplazamiento de los puntos focales proyectados sobre un sensor ver figura 1.

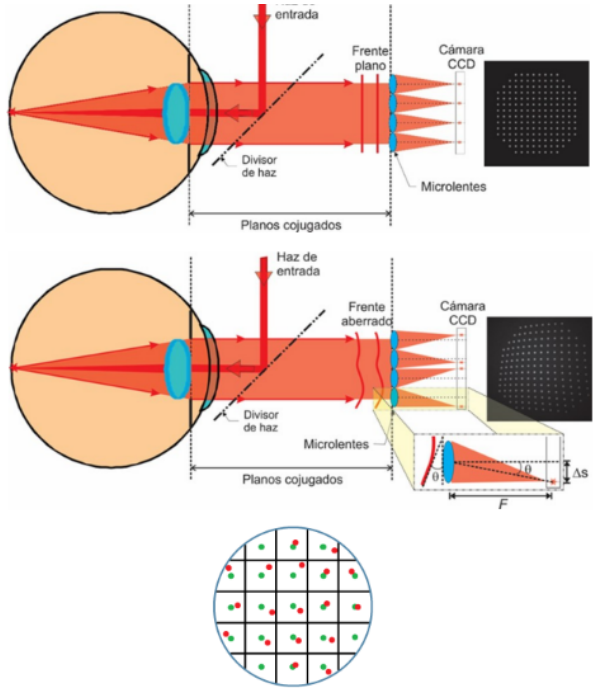


Fig. 1: Comparación entre imágenes de Hartmann-Shack ideal y aberrada

Sin embargo, como señalan Zhang *et al.* [8], existen limitaciones en la medición cuando se presentan aberraciones de gran amplitud o tamaños pupilares amplios, pues el rango dinámico del SHWFS resulta insuficiente para capturar con precisión las desviaciones extremas. Los métodos tradicionales de reconstrucción del frente de onda como los algoritmos modales y zonales fallan en estos casos debido al desplazamiento excesivo de los puntos fuera de su subapertura de referencia, generando errores significativos en la reconstrucción ver figura 2 [2] .

Para abordar esta problemática, se han propuesto métodos de expansión del rango dinámico basados en extrapolación, correlación gaussiana y optimización iterativa, pero estos suelen depender de condiciones iniciales precisas o de la visibilidad total de los puntos de referencia. Recientemente,

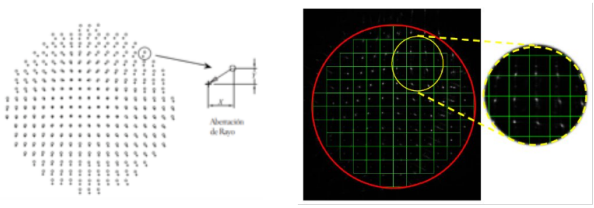


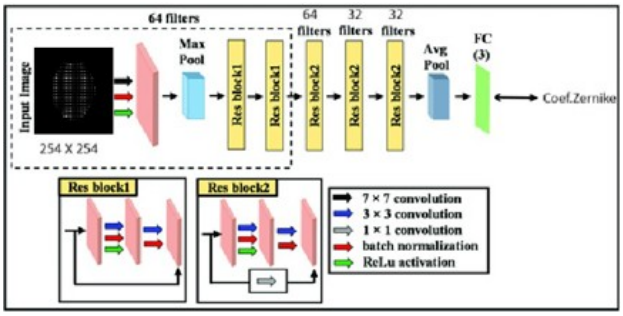
Fig. 2: Método tradicional calculo de aberraciones

enfoques basados en aprendizaje profundo han demostrado su capacidad para aprender directamente la correspondencia entre los patrones Hartmann–Shack y los mapas de fase del frente de onda, evitando los errores de asignación subapertura y mejorando simultáneamente la precisión y el rango dinámico [3].

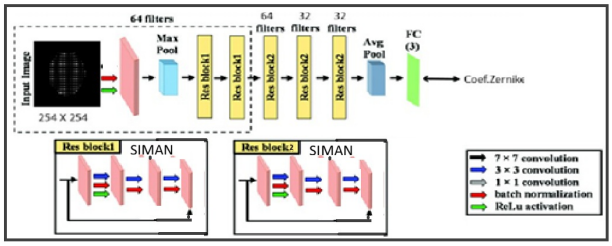
Así mismo, se evalúa los modelos Vision Transformer la cuál la forma de emplearse es que las imágenes se segmentan en numerosos patches. Los cuales se incorporan al modelo del Transformador, lo que permite capturar relaciones inter-patches tanto locales como globales. Un componente crucial del Transformador, el mecanismo de atención, permite al modelo asignar distintos grados de importancia a diferentes patches, centrándose en áreas con mayor información. Esta arquitectura ha demostrado un rendimiento excepcional, superando incluso a las redes neuronales convolucionales tradicionales al aplicarse a conjuntos de datos de imágenes a gran escala [10]. Si bien la aplicación original no se centraba en el procesamiento de datos para imágenes de Hartmann Shack, su principio de diseño abre nuevas posibilidades para aplicaciones innovadoras, como las exploradas en nuestro estudio.

En este proyecto, se emplea un enfoque de aprendizaje profundo utilizando una red neuronal convolucional (CNN) basada en la arquitectura *ResNet18*. Utiliza bloques residuales para mejorar el flujo de gradientes y evitar el problema de desaparición del gradiente en redes profundas. Cada bloque residual consta de dos capas convolucionales con un tamaño de kernel de (3,3), seguidas de y función de activación. El primer conjunto de bloques comienza con 64 canales de entrada y reduce a 32 canales. Posteriormente, las capas aumentan progresivamente la profundidad de la red, pasando por 32, 64 canales en distintos bloques, aplicando stride 2 en algunos bloques para reducir la resolución espacial de la imagen. Esto permite capturar características de alto nivel mientras mantiene la eficiencia computacional. Al final, la arquitectura sigue la idea de *ResNet18*, donde los bloques residuales facilitan el aprendizaje profundo al permitir conexiones de identidad que preservan la información a través de la red [3], también se añade una capa *SimAM*, la cual representa la distancia entre un pixel y la media de todos los pixeles de la imagen el cual introduce un mecanismo de atención auto–iterativa que modula dinámicamente los mapas de características para resaltar información relevante y suprimir señales redundantes [11] ver figura 3.

En la figura 3 se aprecia la Arquitectura usada para entrenar



(a) Arquitectura ResNet18 convencional



(b) Arquitectura ResNet18 con SimAM

Fig. 3: Se presenta dos arquitecturas basadas en ResNet18. En la Figura (a) se muestra la ResNet18 convencional, compuesta por una capa convolucional inicial seguida de cuatro etapas de bloques residuales con convoluciones 3×3, normalización por lotes y activaciones *Silu*, que es una activación mas suavizada de la *Relu* acompañadas de conexiones de atajo que facilitan el flujo del gradiente y permiten entrenar modelos profundos de forma estable. En contraste, la Figura (b) ilustra la *ResNet18* modificada con el módulo *SimAM*. Este módulo se inserta dentro de la misma estructura residual sin alterar su organización general, pero aportando una mayor capacidad para capturar dependencias espaciales complejas y mejorar la discriminación de patrones en tareas de visión por computadora.

unos de los modelos evaluados, que en base es parecida a la arquitectura entregada por *Torchvision* (libreria del *framework Pytorch*) para ser entrenada y evaluada la cual se usó también se uso una arquitectura propia para evaluar resultados ver imagen 3 (b). También sobre la arquitectura *Vision Transformer* se creó en base a lo que existe en la literatura con cambios de parámetros como se puede observar en la figura 4.

Antes del entrenamiento de la red neuronal, se construyó una base de datos de imágenes del sistema óptico a evaluar. En este caso, el sistema estudiado corresponde a un ojo artificial diseñado para simular las diferentes aberraciones presentes en un ojo humano real. Las imágenes se generaron empleando el sensor Hartmann–Shack, el cual permite analizar cómo se comportan diferentes regiones del ojo al hacer pasar un haz de luz a través de cada microlente. Dependiendo del ángulo de desviación de la luz reflejada, se determinan los coeficientes de Zernike, que a su vez caracterizan matemáticamente el tipo y la magnitud de la aberración presente [2]. Estos polinomios, definidos por Mahajan, constituyen una base ortogonal que describe las aberraciones en pupilas circulares y se utilizan

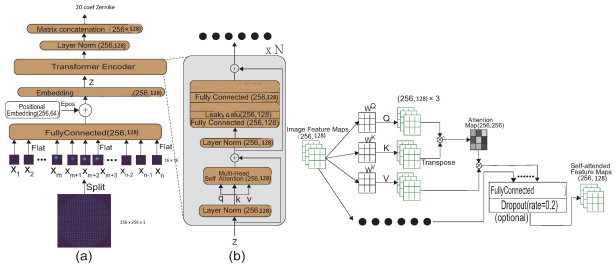


Fig. 4: Arquitectura Vision transformer. En lugar de procesar la imagen completa, ViT la divide en pequeños patches (como cuadritos), convierte cada patch en un vector numérico y los trata como una secuencia de tokens, igual que palabras en un texto. Luego agrega un class token, suma positional embeddings para conservar la información espacial y pasa toda la secuencia por varias capas Transformer que aprenden cómo los patches se relacionan entre sí. Finalmente, el token de clase se usa para predecir la categoría de la imagen. Esta estructura permite que ViT capture relaciones globales en la imagen de manera eficiente y con gran capacidad de generalización.

ampliamente en la representación y corrección de errores ópticos.

Una vez obtenidas las imágenes Hartmann–Shack, la redes neuronales convolucionales (*ResNet18*) y el modelo *Vision Transformer* se entrenan para reconocer las características morfológicas del patrón y predecir los coeficientes de Zernike asociados. De esta forma, la CNN actúa como un mapeo de extremo a extremo entre los Hartmannogramas y los frentes de onda reconstruidos, eliminando la necesidad de métodos modales o zonales clásicos y aumentando la robustez del proceso de medición. El uso de este enfoque basado en aprendizaje profundo representa una integración efectiva entre la óptica física y la inteligencia artificial, contribuyendo al desarrollo de sistemas oftálmicos más precisos y automatizados para la evaluación de aberraciones oculares.

B. Composición de la base de datos

El conjunto de datos empleado proviene de una base de imágenes sintéticas de un ojo artificial, generadas mediante la simulación del patrón Hartmann–Shack para diferentes combinaciones de coeficientes de Zernike en el rango [-1, 1].

- **Tipo de datos:** Imágenes (matrices 2D) en formato .png acompañadas de un vector de 21 coeficientes reales en formato .txt.
- **Tamaño:** 34.000 imágenes generadas (tamaño en disco de aproximadamente 4 GB).
- **Dimensión de entrada:** 1280×1024 píxeles por imagen.
- **Etiquetas:** 21 valores reales asociados a los coeficientes de Zernike.
- **Distribución:** Entrenamiento 80%, Validación 10%, Prueba 10%.

III. Métricas de desempeño

A. Métricas de regresión

- RMSE (Root Mean Squared Error): utilizada como función de pérdida principal.
- Coeficiente de correlación (R): superior a 0.98 en la mayoría de los coeficientes.
- Distribución de residuos: simétrica y centrada en cero.
- Diagramas de caja: muestran estabilidad y ausencia de valores atípicos significativos.

B. Métricas de desempeño del negocio

- Precisión óptica: mejora en la caracterización del frente de onda con errores inferiores al 1%.
- Fiabilidad del modelo: estabilidad ante variaciones en los coeficientes lineales (1 y 2).
- Generalización: validación cruzada con coeficientes no vistos durante el entrenamiento.

IV. Referencias y resultados previos

Estudios pioneros, como el de Tabernero, profundizaron en el origen de estas aberraciones y su relación con el diseño de lentes intraoculares, aportando un marco experimental sólido para su medición y corrección [7].

Posteriormente, Torres y Ruiz demostraron la utilidad del análisis del frente de onda mediante el sensor Hartmann–Shack, aplicándolo a ojos con queratocono para identificar aberraciones de alto orden, lo que permitió correlacionar patrones de desviación del haz de luz con irregularidades corneales y estableció un precedente clave para trabajos de reconstrucción del frente de onda [1].

En el ámbito matemático, Mahajan formalizó el uso de los polinomios de Zernike como herramienta para describir las aberraciones ópticas en sistemas con pupilas circulares, definiendo una base ortogonal que facilita la cuantificación y representación de cada tipo de aberración [2].

Complementariamente, Comastri *et al.* extendieron este análisis mediante transformaciones de coeficientes de Zernike en escenarios donde la pupila se desplaza o contrae transversalmente, aportando una visión más completa de la dinámica del sistema óptico [6].

Con el advenimiento de la inteligencia artificial moderna, la introducción de redes neuronales convolucionales (CNN) —arquitecturas particularmente eficientes para extraer características espaciales y jerárquicas en imágenes— ha abierto nuevas posibilidades para el análisis automatizado de frentes de onda. Estas redes permiten detectar patrones complejos en imágenes Hartmann–Shack y relacionarlos con los coeficientes de Zernike, ofreciendo una alternativa más rápida, precisa y generalizable frente a los métodos clásicos de reconstrucción óptica [3].

Se usa los modelos *Vision Transformer* para estimar el los coeficientes de *Zernike* para la aberración de frente de onda, Si bien la aplicación original del Transformador de Visión no se centraba en el procesamiento de datos *SHWFS*, su principio de diseño abre nuevas posibilidades para aplicaciones innovadoras, como las exploradas en nuestro estudio [9].

V. Resultados

Entrenamiento: Durante la etapa de entrenamiento se ejecutaron 50 épocas utilizando un tamaño de lote (*batch size*) de 32 imágenes y empleando 4 núcleos de procesamiento para optimizar el uso de la GPU NVIDIA T4 disponible en las plataformas de Google Colab y Kaggle. En cada época, el modelo ajustó sus parámetros a partir del conjunto de entrenamiento mientras se registraba la pérdida correspondiente con el fin de evaluar la convergencia y la estabilidad del proceso de optimización. Esta configuración permitió un equilibrio adecuado entre velocidad de entrenamiento y estabilidad numérica en la estimación de los coeficientes.

Validación: Al finalizar cada época se procedió a validar el modelo empleando un conjunto independiente de imágenes almacenado específicamente para esta tarea. Utilizando también un *batch size* de 32 y la misma configuración de cómputo, se evaluó el desempeño medio del modelo frente a datos no utilizados durante el ajuste. El modelo final seleccionado correspondió a aquel que presentó la menor pérdida de validación dentro de las 50 épocas, garantizando así una elección centrada en la mejor capacidad de generalización antes de incurrir en sobreajuste.

Test: La etapa de prueba (*test*) se llevó a cabo utilizando imágenes completamente nuevas para el modelo, siguiendo un procedimiento análogo al de validación y manteniendo el mismo esquema computacional. Esta fase permite estimar el rendimiento real del modelo frente a datos desconocidos y, por lo tanto, representa la métrica más confiable para evaluar su capacidad de generalización. Los resultados obtenidos proporcionan una perspectiva clara sobre la eficacia final de cada arquitectura considerada.

A continuación se presentan los resultados obtenidos con los tres modelos evaluados: la *ResNet18* estándar de *Torchvision*, una *ResNet18* modificada con el módulo *SimAM* y, finalmente, el *Vision Transformer* implementado. En general, los tres modelos muestran un desempeño sólido, pero con diferencias importantes que evidencian cómo cada arquitectura captura la información de manera distinta. La *ResNet18* base ofrece resultados estables y coherentes, mientras que la versión con *SimAM* presenta mejoras en la atención espacial y obtiene un rendimiento un poco superior en el tiempo de ejecución. Por su parte, el *Vision Transformer* demuestra una excelente capacidad para aprender relaciones globales dentro de las imágenes, alcanzando resultados competitivos y revelando un comportamiento especialmente prometedor dada la naturaleza del conjunto de datos, teniendo en cuenta que este último fue el de peor desempeño de los tres.

TABLE I: Comparación de métricas de desempeño entre los modelos evaluados.

Modelo	MSE	RMSE	MAE	R^2	Tiempo/época (min)
ResNet18 (torchvision)	0.0039	0.0623	0.0403	0.9956	02:59
ResNet18 + SimAM	0.0041	0.0644	0.0493	0.9907	01:46
Vision Transformer	0.0100	0.1000	0.0707	0.9800	02:10

La Tabla I presenta una comparación cuantitativa del desempeño entre los tres modelos evaluados: *ResNet18* estándar

de *torchvision*, *ResNet18* con el módulo de atención *SimAM* y el *Vision Transformer*. Se incluyen métricas de error (*MSE*, *RMSE* y *MAE*), el coeficiente de determinación (R^2) y el tiempo promedio por época. Los resultados muestran que la *ResNet18* convencional obtiene el mejor desempeño global en términos de error y capacidad explicativa, seguida de la variante con *SimAM*, que presenta ligeras pérdidas de precisión pero una reducción significativa en el tiempo de entrenamiento. Por su parte, el *Vision Transformer* presenta los errores más altos y un R^2 menor en comparación con las arquitecturas convolucionales, lo que sugiere un menor ajuste al problema bajo las condiciones experimentales evaluadas.

Las Figuras 5 (a), (b) y (c) presentan las curvas de pérdida de entrenamiento (línea azul) y validación (línea roja punteada) para los modelos ResNet18, ResNet18 con SimAM y Vision Transformer, respectivamente. En los tres casos se observa un comportamiento estable y coherente entre ambas curvas, lo que indica que ninguno de los modelos presenta sobreajuste, dado que la pérdida de validación no diverge ni se separa significativamente de la pérdida de entrenamiento conforme avanza el proceso de optimización. Del mismo modo, no se evidencia subajuste, ya que las pérdidas disminuyen adecuadamente y alcanzan valores consistentes con un aprendizaje efectivo. Es importante resaltar que, en el caso del Vision Transformer, aun sin aplicar técnicas explícitas de regularización como *dropout*, el modelo mantiene un equilibrio adecuado entre las curvas de entrenamiento y validación, demostrando una correcta generalización sin indicios de sobreajuste.

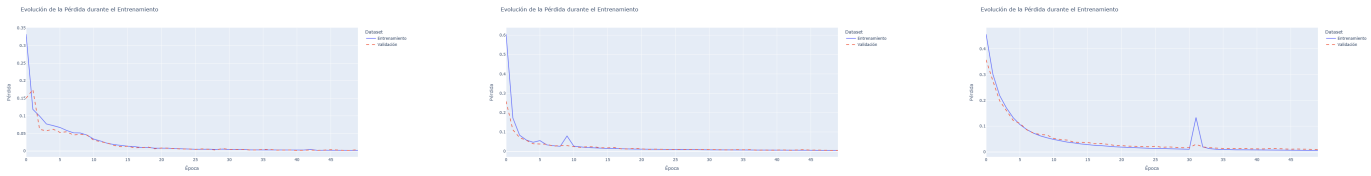
La Figura 6 presenta el error acumulado de los coeficientes de Zernike para los tres modelos evaluados. Al analizar la forma de las distribuciones obtenidas, se observa que todas presentan una estructura aproximadamente gaussiana. De acuerdo con el Teorema del Límite Central, cuando un error total proviene de la suma de múltiples fuentes independientes y pequeñas, es natural que su distribución tienda hacia una forma normal. Por lo tanto, la simetría y el perfil suavemente acampanado de las curvas indican que el error presente en los tres modelos corresponde principalmente a error aleatorio inherente al proceso de estimación, sin evidencia de sesgos sistemáticos o desviaciones significativas que sugieran problemas de sobreajuste o subajuste. Esta coherencia en la forma de las distribuciones respalda la estabilidad estadística del proceso de predicción realizado por los modelos.

La Figura 7 muestra el desempeño individual de los coeficientes de Zernike para los tres modelos evaluados, se observa un comportamiento coherente con la teoría estadística. En particular, el modelo ResNet18 por defecto de *torchvision* presenta un rendimiento global adecuado; sin embargo, el coeficiente 1 exhibe una ligera inconsistencia en comparación con los demás, lo cual sugiere una variabilidad particular en dicho componente al analizarlo de manera individual. Aun así, esta variación no afecta significativamente el desempeño general del modelo. Por otra parte, tanto en la ResNet18 con SimAM como en el Vision Transformer, los coeficientes presentan un comportamiento más uniforme. En los tres modelos, los errores individuales muestran una distribución compatible con el Teorema del Límite Central, lo que indica que las desviaciones observadas corresponden principalmente a error aleatorio acumulado por múltiples fuentes pequeñas e independientes, en lugar de sesgos sistemáticos. Esto confirma que, pese a las diferencias entre arquitecturas, los modelos mantienen estabilidad estadística y una convergencia adecuada en la estimación de los coeficientes.

La Figura 8 muestra el diagrama de cajas correspondiente a la distribución de errores individuales para los coeficientes de Zernike en los tres modelos evaluados. Se observa que la ResNet18 por defecto de *torchvision* presenta una mayor variabilidad en comparación con las otras arquitecturas, reflejada en cajas más amplias y medianas que tienden a alejarse ligeramente de cero, lo que indica que los coeficientes estimados presentan un ajuste menos centrado respecto al valor ideal. En contraste, la ResNet18 con SimAM exhibe una reducción notable en la dispersión y medianas más próximas a cero, mostrando un comportamiento más estable y uniforme. Finalmente, aunque el Vision Transformer es el modelo con menor desempeño cuantitativo en términos de las métricas globales, sus coeficientes presentan la menor varianza entre los tres modelos, con distribuciones más compactas y simétricas. Esta baja variabilidad resulta especialmente favorable para aplicaciones relacionadas con la evaluación de aberraciones del frente de onda, donde la consistencia y estabilidad entre coeficientes es tan importante como la magnitud del error promedio.

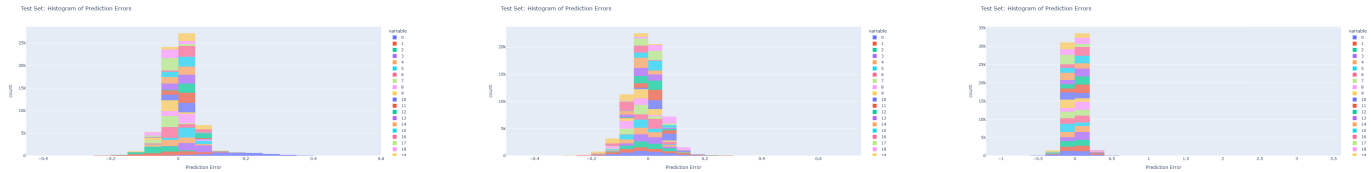
Los resultados obtenidos en este estudio permiten comparar de manera integral el comportamiento de las tres arquitecturas evaluadas: ResNet18 por defecto, ResNet18 con SimAM y el Vision Transformer. A nivel global, las métricas de desempeño indican que la ResNet18 convencional alcanza los mejores valores en términos de error promedio, seguida de la versión con SimAM, que aunque presenta una ligera disminución en la precisión, logra tiempos de entrenamiento más reducidos y un comportamiento más uniforme en la predicción de los coeficientes. El Vision Transformer, si bien obtiene los errores globales más altos, demuestra importantes ventajas en términos de estabilidad, evitando tanto sobreajuste como subajuste incluso sin técnicas explícitas de regularización. Además, sus curvas de pérdida presentan un comportamiento consistente entre entrenamiento y validación, lo que confirma su capacidad para generalizar adecuadamente bajo las condiciones de experimentación utilizadas.

Por otra parte, el análisis estadístico de los coeficientes, mediante curvas de error acumulado y diagramas de caja, revela que los tres modelos producen errores compatibles con una distribución aproximadamente normal, en concordancia con el Teorema del Límite Central, lo que sugiere que las desviaciones observadas provienen mayormente de ruido aleatorio y no de sesgos sistemáticos. Aun así, se identificó que la ResNet18 por defecto presenta mayor variabilidad en ciertos coeficientes, especialmente el coeficiente 1, aunque sin afectar su desempeño global. En contraste, la ResNet18 con SimAM reduce la dispersión de forma notable y el Vision Transformer destaca por ser el modelo con menor varianza



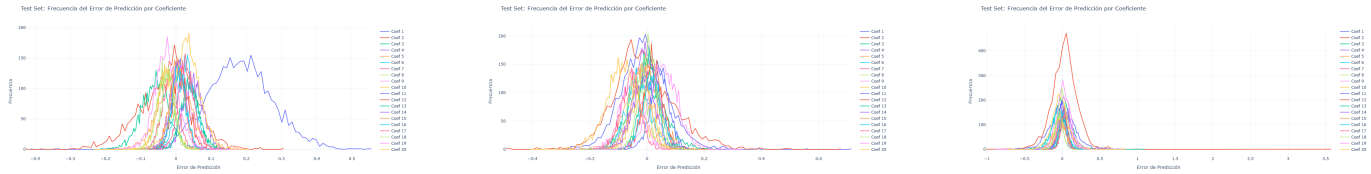
(a) *ResNet18 torchvision* (b) *ResNet18 con SimAM* (c) *Vision Transformer*

Fig. 5: Comparación de la pérdida de los tres modelos analizados.



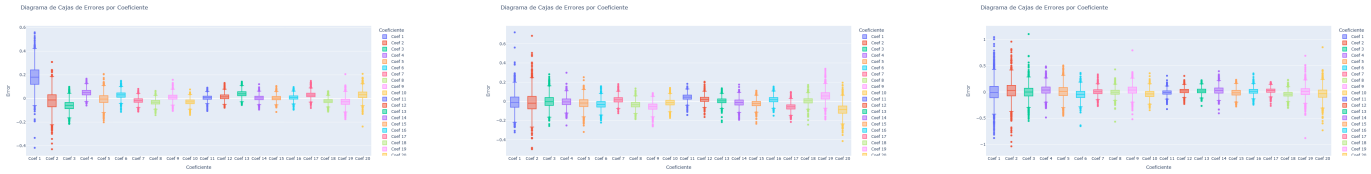
(a) *ResNet18 torchvision* (b) *ResNet18 con SimAM* (c) *Vision Transformer*

Fig. 6: Comparación error acumulado de los tres modelos analizados.



(a) *ResNet18 torchvision* (b) *ResNet18 con SimAM* (c) *Vision Transformer*

Fig. 7: Comparación error con su componente individual de los tres modelos analizados.



(a) *ResNet18 torchvision* (b) *ResNet18 con SimAM* (c) *Vision Transformer*

Fig. 8: Comparación del diagrama de cajas para los tres modelos analizados.

en los coeficientes, una propiedad especialmente deseable para aplicaciones que requieren alta estabilidad en la evaluación de aberraciones del frente de onda. En conjunto, estos hallazgos evidencian la robustez de las tres arquitecturas y permiten comprender sus fortalezas específicas en el contexto del análisis de coeficientes de Zernike.

VI. Discusión

Los resultados obtenidos en este trabajo permiten analizar de manera integral el comportamiento de las arquitecturas ResNet18, ResNet18 con SimAM y Vision Transformer en la estimación de coeficientes de Zernike a partir de imágenes sintéticas de un sensor Shack–Hartmann. A diferencia de estudios previos en la literatura, como el presentado en [9], donde se evalúa el desempeño bajo condiciones de iluminación críticas, diferentes magnitudes estelares y múltiples valores de R_0, los experimentos realizados en este laboratorio se desarrollaron bajo un esquema controlado de generación de datos, dando como resultado errores considerablemente menores. En el artículo citado, las configuraciones basadas en redes convolucionales muestran un desempeño inferior al de los modelos Vision Transformer, especialmente al analizarse la corrección de fase y las distribuciones de energía del PSF bajo condiciones severas de turbulencia y baja iluminación. Por el contrario, en este trabajo la ResNet18 por defecto muestra la mejor métrica global, seguida de su versión con SimAM, mientras que el Vision Transformer obtiene un error mayor pero destaca por su estabilidad y baja varianza en los coeficientes estimados, lo cual resulta altamente favorable para aplicaciones ópticas.

Otro aspecto relevante es que, al evaluar los coeficientes

de manera individual, los tres modelos presentan errores compatibles con ruido aleatorio según el Teorema del Límite Central, sugiriendo ausencia de sesgos sistemáticos. Aunque la ResNet18 convencional muestra una inconsistencia particular en el coeficiente 1, el resto de coeficientes mantienen estabilidad estadística. Asimismo, los diagramas de caja muestran que la ResNet18 por defecto presenta mayor dispersión, mientras que la ResNet18 con SimAM reduce significativamente dicha variabilidad y el Vision Transformer ofrece el comportamiento más compacto. Esto contrasta parcialmente con los resultados de la literatura, donde el Vision Transformer domina en escenarios de turbulencia elevada y condiciones fotométricas adversas. La diferencia sugiere que, en condiciones controladas como las de este laboratorio, las arquitecturas convolucionales pueden aprovechar mejor la regularidad espacial del problema, mientras que el Transformer revela su mayor potencial en situaciones de iluminación complejas, como también lo señala [9].

VII. Conclusiones

En conjunto, los resultados experimentales demuestran que las tres arquitecturas evaluadas son capaces de estimar de manera precisa los coeficientes de Zernike provenientes de un sensor Shack–Hartmann sintético, sin signos de sobreajuste o subajuste, incluso en el caso del Vision Transformer, que no empleó técnicas explícitas de regularización. La ResNet18 por defecto alcanza el mejor rendimiento global, mientras que la versión con SimAM ofrece un equilibrio notable entre precisión, estabilidad y tiempo de entrenamiento reducido. El Vision Transformer, aunque exhibe un error mayor en las métricas agregadas, produce la menor varianza entre coeficientes, un resultado especialmente relevante para sistemas de corrección de aberraciones donde la consistencia es tan importante como la precisión puntual.

Así también, Se concluye que los modelos empleados en este laboratorio operan en un régimen distinto al de las pruebas bajo condiciones extremas de iluminación realizados en [9], mostrando resultados prometedores. Como trabajo futuro, se propone mejorar la fase de generación de datos incorporando la propagación de la luz en el modelo físico, con el fin de reproducir las deformaciones características de las imágenes reales de sensores Shack–Hartmann. Esto permitirá entrenar modelos más robustos y transferibles a condiciones experimentales reales, así como evaluar el desempeño de Transformers y CNN en escenarios más comparables a los expuestos en la literatura.

Finalmente, es importante resaltar la incorporación de los coeficientes 1 y 2 dentro del proceso de estimación, aun cuando estos no representan información asociada directamente a aberraciones ópticas. Dichos coeficientes corresponden a los términos de desplazamiento en los polinomios del frente de onda plana, y por tanto, no serían de interés físico para el análisis de aberraciones. Sin embargo, su inclusión se vuelve necesaria debido a que estos desplazamientos afectan de manera directa la estimación del modelo. Durante las pruebas iniciales, cuando el modelo entrenado recibía imágenes que únicamente presentaban desplazamientos en los ejes “x” y “y”, pero cuya estructura era idéntica a las imágenes del conjunto de entrenamiento, se observó un incremento notable en el error cuadrático medio. Al agregar estos coeficientes, los modelos adquieren consistencia frente a variaciones de traslación y generalizan adecuadamente dichos componentes en imágenes futuras. No obstante, se evidenció que el modelo ResNet18 por defecto de `torchvision` presenta un mayor error de estimación específicamente en estos dos coeficientes en comparación con los otros modelos evaluados, tal como se refleja en los errores individuales. Este ajuste contribuyó a disminuir el error global en relación con lo reportado en la literatura, fortaleciendo la robustez y la validez del enfoque propuesto.

Referencias

[1] K. Torres y N. Ruiz, “Aberraciones de alto orden en ojos con queratocono, medidas mediante análisis de frente de onda Hartmann–Shack,” *Revista Mexicana de Oftalmología*, vol. 83, no. 2, pp. 100–105, 2009.

[2] V. N. Mahajan, “Zernike circle polynomials and optical aberrations of systems with circular pupils,” *Applied Optics*, vol. 20, no. 13, pp. 2091–2098, 1981.

[3] E. Todt and B. A. Krinski, *Convolutional Neural Network – CNN*, 30 Nov. 2019. [En línea]. Disponible: https://www.inf.ufpr.br/todt/IAaplicada/CNN_Presentation.pdf

[4] R. Salas, *Redes neuronales artificiales*, Universidad de Valparaíso, Departamento de Computación, vol. 1, no. 1, pp. 1–7, 2004.

[5] R. Vidal Olarte, *Entendiendo e interpretando las aberraciones ópticas / Understanding and Interpreting Optical Aberrations*, 2011.

[6] S. A. Comastri, K. Bastida, G. Martin, y A. Bianchetti, *Aberrometrías oculares y de otros sistemas ópticos: transformación de coeficientes Zernike al contraer y desplazar transversalmente la pupila*, Informe N.º 208, Facultad de Ingeniería, Universidad de Belgrano, 2008.

[7] J. Tabernero, *Estudios de las fuentes de aberraciones en el ojo humano. Aplicaciones en lentes intraoculares*, Tesis doctoral, Universidad de Murcia, Murcia, España, 2007.

[8] H. Zhang, J. Zhao, H. Chen, Z. Zhang, C. Yin, y S. Wang, “Large-Dynamic-Range Ocular Aberration Measurement Based on Deep Learning with a Shack–Hartmann Wavefront Sensor,” *Sensors*, vol. 24, no. 9, p. 2728, 2024. DOI: 10.3390/s24092728.

[9] Zhang, Q.; Zuo, H.; Cui, X.; Yuan, X.; Hu, T. Automatic Compressive Sensing of Shack–Hartmann Sensors Based on the Vision Transformer. Photonics 2024, 11, 998. https://doi.org/10.3390/photonics11110998

[10] Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image is Worth 16 × 16 Words: Transformers for Image Recognition at Scale. In Proceedings of the International Conference on Learning Representations, Virtual, 3–7 May 2021.

[11] Yang, Y., Yu, H., Gao, S. (2021). SimAM: A simple, parameter-free attention module for convolutional neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 2). https://proceedings.mlr.press/v139/yang21o/yang21o.pdf