

Karla Lizeth Villafuerte Ruiz

**Reducción del ruido
en el proceso de desdoblamiento
de fase espacial mediante
redes neuronales**

2026



Universidad Autónoma de Querétaro
Facultad de Ingeniería

**Reducción del ruido en el proceso de
desdoblamiento de fase espacial
mediante redes neuronales**

Tesis

Que como parte de los requisitos para obtener el
Grado de

**Maestra en Ciencias
en Inteligencia Artificial**

Presenta

Karla Lizeth Villafuerte Ruiz

Dirigido por:
Dr. Jesús Carlos Pedraza Ortega

Querétaro, Qro. a mayo de 2026

La presente obra está bajo la licencia:
<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.es>



CC BY-NC-ND 4.0 DEED

Atribución-NoComercial-SinDerivadas 4.0 Internacional

Usted es libre de:

Compartir — copiar y redistribuir el material en cualquier medio o formato

La licenciante no puede revocar estas libertades en tanto usted siga los términos de la licencia

Bajo los siguientes términos:



Atribución — Usted debe dar [crédito de manera adecuada](#), brindar un enlace a la licencia, e [indicar si se han realizado cambios](#). Puede hacerlo en cualquier forma razonable, pero no de forma tal que sugiera que usted o su uso tienen el apoyo de la licenciante.



NoComercial — Usted no puede hacer uso del material con [propósitos comerciales](#).



SinDerivadas — Si [remezcla, transforma o crea a partir](#) del material, no podrá distribuir el material modificado.

No hay restricciones adicionales — No puede aplicar términos legales ni [medidas tecnológicas](#) que restrinjan legalmente a otras a hacer cualquier uso permitido por la licencia.

Avisos:

No tiene que cumplir con la licencia para elementos del material en el dominio público o cuando su uso esté permitido por una [excepción o limitación](#) aplicable.

No se dan garantías. La licencia podría no darle todos los permisos que necesita para el uso que tenga previsto. Por ejemplo, otros derechos como [publicidad, privacidad, o derechos morales](#) pueden limitar la forma en que utilice el material.



Universidad Autónoma de Querétaro
Facultad de Ingeniería
Maestría en Ciencias (Inteligencia Artificial)

Reducción del ruido en el proceso de desdoblamiento de fase espacial mediante redes neuronales

TESIS

Que como parte de los requisitos para obtener el Grado de
Maestro en Ciencias (Inteligencia Artificial)

Presenta:

Karla Lizeth Villafuerte Ruiz

Dirigido por:

Dr. Jesús Carlos Pedraza Ortega

Dr. Jesús Carlos Pedraza Ortega

Presidente

MCIA. Osmar Antonio Espinosa Bernal

Secretario

Dr. Juan Manuel Ramos Arreguín

Vocal

MISD. Artemio Sotomayor Olmedo

Suplente

Dr. Saúl Tovar Arriaga

Suplente

Centro Universitario, Querétaro, Qro.

Fecha de Aprobación por el consejo (mayo 2026)

México

Agradecimientos

Quiero expresar mi más sincero agradecimiento a todas las personas que hicieron posible la culminación de esta etapa tan importante de mi formación académica.

A mis padres, por su amor incondicional, por su apoyo constante y por creer siempre en mí, incluso en los momentos más difíciles. A mis hermanos, por estar presentes, acompañarme y brindarme su apoyo en cada paso de este camino.

A mi asesor de tesis, el Dr. Jesús Carlos Pedraza, por su confianza, guía y acompañamiento durante el desarrollo de este trabajo, así como por motivarme a dar lo mejor de mí y creer en mis capacidades.

A mis maestros de la maestría, por compartir sus conocimientos, su experiencia y su tiempo, contribuyendo de manera significativa a mi formación académica y profesional.

A mis amigas y amigos, por su apoyo, compañía y comprensión durante esta etapa, haciendo más llevadero el camino y aportando ánimo en los momentos necesarios.

Finalmente, agradezco a la Secretaría de Ciencia, Humanidades, Tecnología e Innovación por el apoyo otorgado a través de la beca, la cual fue fundamental para poder dedicarme de manera plena al desarrollo de esta maestría.

Resumen

En esta tesis se presenta una evaluación integral del desdoblamiento de fase espacial mediante la comparación entre métodos tradicionales y modelos basados en aprendizaje profundo. Se trabajó con datos sintéticos y reales obtenidos a través de perfilometría por proyección de franjas (FPP), con el objetivo de reducir el ruido presente en los mapas de fase y disminuir el error en la reconstrucción tridimensional. Los métodos tradicionales, aunque ampliamente utilizados, mostraron limitaciones significativas en escenarios con discontinuidades, ruido estructurado y variaciones de iluminación, produciendo artefactos y pérdidas de información. En contraste, las arquitecturas basadas en redes neuronales convolucionales (CNN), particularmente U-Net y sus variantes, demostraron un desempeño superior tanto cuantitativa como cualitativamente. Los modelos alcanzaron mejoras notables en métricas como PSNR, SSIM e IMMSE, además de generar reconstrucciones visualmente más coherentes y robustas ante condiciones experimentales no ideales. Los resultados obtenidos confirman que el uso de aprendizaje profundo mejora de manera sustancial la calidad del desdoblamiento de fase espacial y representa una alternativa efectiva frente a los métodos clásicos. Finalmente, se identificaron limitaciones relacionadas con la disponibilidad de datos reales, la brecha entre entornos sintéticos y experimentales y los requerimientos computacionales, proponiendo líneas de trabajo futuro orientadas a ampliar los datos.

Palabras clave: Desdoblamiento de fase espacial, redes neuronales convolucionales, proyección de franjas, PSNR, SSIM, IMMSE.

Abstract

This thesis presents a comprehensive evaluation of spatial phase unwrapping through a comparative study between traditional algorithms and deep learning-based models. Both synthetic and real data acquired via Fringe Projection Profilometry (FPP) were used with the aim of reducing noise in phase maps and decreasing the error in three-dimensional reconstruction. Traditional phase unwrapping methods, although widely adopted, exhibited substantial limitations in scenarios involving structured noise, discontinuities, and illumination variability, often producing artifacts and loss of structural information. In contrast, convolutional neural network (CNN) architectures—particularly U-Net and its variants—demonstrated superior quantitative and qualitative performance. The proposed models achieved significant improvements in PSNR, SSIM, and IMMSE, while producing more coherent and robust reconstructions under non-ideal experimental conditions. The results confirm that deep learning significantly enhances the quality of spatial phase unwrapping and constitutes an effective alternative to classical approaches. Limitations related to the scarcity of real data, the gap between synthetic and experimental conditions, and computational demands were identified. Future work is proposed toward expanding real datasets.

Keywords: Unwrapping spatial phase, Convolutional neural networks, Fringe Projection Profilometry, PSNR, SSIM, IMMSE.

Abreviaturas y siglas

Aprendizaje profundo y arquitecturas

AI – Artificial Intelligence

CNN – Convolutional Neural Network

DL – Deep Learning

ML – Machine Learning

LSTM – Long Short-Term Memory

SQD-LSTM – Spatial Quadratic Depth Long Short-Term Memory

ResNet – Residual Network

Interferometría, fase y métodos de desdoblamiento

FPP – Fringe Projection Profilometry

GT – Ground Truth

FFT – Fast Fourier Transform

DFT – Discrete Fourier Transform

ROI – Region of Interest

STIEPU – Spatio-Temporal Iterative Error Propagation Unwrapping

PUMA – Phase Unwrapping Maximum A Posteriori

PEARLS – Phase Extraction And Reconstruction by Learning-based Strategies

Entrenamiento, optimización y redes

MSE – Mean Squared Error

MAE – Mean Absolute Error

RMSE – Root Mean Squared Error

PSNR – Peak Signal-to-Noise Ratio

SSIM – Structural Similarity Index Measure

IMMSE – Image Masked Mean Squared Error

LR – Learning Rate

BN – Batch Normalization

ReLU – Rectified Linear Unit

SGD – Stochastic Gradient Descent

Adam – Adaptive Moment Estimation

Computación, hardware y procesamiento

GPU – Graphics Processing Unit

CPU – Central Processing Unit

RAM – Random Access Memory

VRAM – Video Random Access Memory

NIfTI – Neuroimaging Informatics Technology Initiative

Datos y manejo de datasets

LW – Low Wrap

MW – Multi Wrap

DB – Database

OWS – Object-Wise Split

MOS – Mixed-Object Split

Índice

I.	Introducción	16
1.1	Planteamiento del problema.....	18
1.2	Justificación.....	19
II.	Hipótesis.....	23
III.	Objetivos	23
3.1.1	Objetivo principal.....	23
3.1.2	Objetivos específicos	23
IV.	Antecedentes	24
4.1	Desdoblamiento de fase	29
4.2	Métodos de desdoblamiento de fase espaciales tradicionales.....	31
4.2.1	Itoh (1982).....	32
4.2.2	Ghiglia & Romero (1994–1998).....	32
4.2.3	Costantini (1998).....	32
4.2.4	Herráez et al. (2002).....	33
4.2.5	PUMA (Bioucas-Dias & Valadão, 2007)	33
4.2.6	PEARLS (Bioucas-Dias et al., 2010).....	34
4.2.7	Zhao, Chen & Tan (Robust 2D, 2016)	34
4.2.8	Martínez-Carranza et al. (2012)	34
4.3	Ruido en el desenvolvimiento de fase.....	35
4.4	Perfilometría por cambio de fase	35
4.5	Phase shifting de n-step.....	37
4.6	Perfilometría de Fourier bidimensional	37
4.7	Mapa de fase	38

4.8	Inteligencia Artificial	39
4.8.1	Machine learning.....	40
4.8.2	Deep Learning.....	40
4.8.3	Redes neuronales.....	41
4.8.4	Redes neuronales convolucionales.....	42
4.8.4.1	U-net.....	43
4.8.4.2	Res-net.....	46
4.8.4.3	Multiresolution CNN	49
4.9	Métricas.....	50
4.9.1	Métricas IA.....	51
4.9.2	Métricas imágenes.....	52
4.9.2.1	Error Cuadrático Medio en Imágenes (IMMSE)	52
4.9.2.2	Relación Señal–Ruido de Pico (PSNR)	53
4.9.2.3	Índice de Similitud Estructural (SSIM)	53
4.10	Estado del arte.....	54
4.10.1	Métodos tradicionales	55
4.10.1	Estado del arte basado en Deep Learning para desdoblamiento de fase y reconstrucción 3D	57
V.	Metodología	60
5.1	Generación de bases de datos.....	61
5.1.1	Imágenes sintéticas.....	61
5.1.1.1	Generación de mapas de fase sintéticos	65
5.1.2	Imágenes reales	67
5.2	Establecimiento de métricas.....	68

5.2.1	Evaluación cuantitativa de los métodos tradicionales.....	69
5.2.2	Ajuste global de altura	69
5.2.3	Definición matemática	70
5.2.3.1	Error Cuadrático Medio en Imágenes (IMMSE)	70
5.2.3.2	Relación Señal-Ruido de Pico (PSNR).....	70
5.2.3.3	Índice de Similitud Estructural (SSIM)	70
5.2.4	Evaluación cuantitativa de los modelos de aprendizaje profundo	71
5.3	Implementación de métodos tradicionales.....	72
5.3.1	Preparación de los mapas de fase.....	72
5.3.2	Aplicación de los métodos de desdoblamiento de fase.....	73
5.3.3	Construcción de mapas de reconstrucción y normalización en la máscara	75
5.3.4	Organización de los resultados.....	76
5.4	Diseño y entrenamiento de la red neuronal convolucional	77
5.4.1	Arquitecturas	77
5.4.2	Función de pérdida.....	78
5.4.3	Datos de entrenamiento.....	78
5.4.4	Procedimiento de entrenamiento.....	79
5.4.5	Fine-tuning	80
5.4.5.1	Data augmentation especializado en mapas de fase.....	80
5.4.5.2	Entrenamiento del ajuste fino.....	81
5.5	Documentación y publicación.....	82
VI.	Resultados y discusión	84
6.1	Bases de datos	84
6.1.1	Objetos sintéticos	85

6.1.1.1	Conjuntos generados	85
6.1.1.2	Low-wrap vs. multi-wrap.....	86
6.1.2	Objetos reales	86
6.1.3	Relevancia de la construcción de bases	87
6.2	Métodos tradicionales	87
6.2.1	Objetos evaluados	89
6.2.2	Resultados cualitativos.....	89
6.2.3	Resultados cuantitativos.....	92
6.2.4	Desempeño según la geometría del objeto.....	95
6.2.5	Discusión del desempeño.....	95
6.3	Modelos de aprendizaje profundo.....	97
6.3.1	Introducción a los primeros experimentos y evolución del modelo	98
6.3.1.1	Primeras inferencias con el modelo pre-entrenado	98
6.3.1.2	Ciclo inicial de experimentos de fine-tuning	99
6.3.1.3	Experimentos fine-tuning con la base de datos actualizada.....	103
6.3.2	Comparación de los entrenamientos	107
6.3.3	Desempeño por arquitectura	109
6.3.3.1	U-Net con SQD-LSTM.....	110
6.3.3.2	U-Net Res-Net.....	114
6.3.3.3	Arquitectura Multiresolution.....	118
6.3.3.4	U-Net estándar	122
6.3.4	Evaluación en objetos sintéticos	126
6.3.4.1	Resultados cualitativos V1, V7, V12 y V17	127
6.3.4.1	Resultados cualitativos V6, V11.5, V12 y V17.....	129

6.3.4.2	Comparación entre el mejor y el peor entrenamiento	131
6.3.5	Inferencia en objetos reales	135
6.3.5.1	Desempeño del modelo V17 (U-Net estándar)	137
6.3.5.2	Desempeño del modelo V6 (U-Net con SQD-LSTM).....	138
6.3.5.3	Arquitecturas Res-Net y Multiresolution (V11.5, V12, V15, V16)	139
6.3.5.1	Comparación global entre modelos.....	139
6.3.6	Fine-tuning	143
6.3.6.1	Desajuste estadístico entre datos sintéticos y reales	143
6.3.6.2	Selección del modelo	145
6.3.6.3	Aumento de datos.....	145
6.3.6.4	Procedimiento de entrenamiento.....	146
6.3.6.5	Resultados y comportamiento del modelo	146
6.3.6.6	Evaluación del fine-tuning en objetos reales	147
6.4	CNN vs métodos tradicionales.....	149
6.4.1	Objetos sintéticos	149
6.4.2	Objetos reales	157
VII.	Conclusiones	161
VIII.	Bibliografía	165
IX.	Anexos	173
9.1	Registro del protocolo de investigación.....	174
9.2	Inglés.....	176
9.3	Retribución social.....	177
9.4	Artículo	178

Índice de figuras

Figura 1. Ejemplo de reconstrucción de objeto 3D con y sin ruido.....	20
Figura 2. Anatomía de un ojo.....	24
Figura 3. Diagrama del proceso de reconstrucción 3D.....	26
Figura 4. Ejemplo de flujo de trabajo.....	28
Figura 5. Sistema y diagrama de perfilometría por proyección de franjas.	30
Figura 6. Diagrama del proceso de perfilometría de fase.	37
Figura 7. Representación de objeto 3D y su mapa de fase.....	38
Figura 8. Diagrama de Inteligencia artificial y sus subcampos	39
Figura 9. Funciones de activación.....	41
Figura 10. Diagrama de una red neuronal.....	42
Figura 11. Arquitectura U-Net.	44
Figura 12. Arquitectura U-Net + LSTM.	45
Figura 13. Arquitectura UNetResNet.....	48
Figura 14. Arquitectura Multiresolution CNN.....	49
Figura 15. Diagrama de matriz de confusión.....	52
Figura 16. Conjunto representativo de las geometrías 3D.....	62
Figura 17. Diagrama de flujo del sistema automatizado.....	63
Figura 18. Conjunto de imágenes producidas Blender–Python.....	64
Figura 19. Conjunto de imágenes base correspondientes a un objeto sintético.....	64
Figura 20. Flujo general de un sistema de proyección de franjas	66
Figura 21. Ejemplos de proyección de franjas sobre objetos reales.	68
Figura 22. Mapas de fase	68
Figura 23. Objetos utilizados en métodos tradicionales.....	88
Figura 24. Reconstrucciones de los objetos 3, 5 y 6.....	91
Figura 25. Gráficas comparativas para los objetos 3, 5 y 6.....	92
Figura 26. Ejemplo de las primeras inferencias de los objetos sintéticos.....	99
Figura 27. Resultados visuales de los entrenamientos del fine-tuning.	100

Figura 28. Resultados del fine-tuning con GT normalizado globalmente y 50 mapas de fase.	101
Figura 29. Resultados del fine-tuning con normalización por imagen y 247 mapas de fase.	102
Figura 30. Mapas de fase low-wrap y GT generado mediante spline.	103
Figura 31. Resultados visuales de los entrenamientos del fine-tuning.	105
Figura 32. Gráficas de entrenamientos de U-Net con SQD-LSTM.	112
Figura 33. Predicciones de los modelos de la arquitectura U-Net con SQD-LSTM.	113
Figura 34. Gráficas de los entrenamientos de la arquitectura U-Net Res-Net.	116
Figura 35. Predicciones de los modelos de la arquitectura U-Net Res-Net.	117
Figura 36. Gráficas de los entrenamientos de la arquitectura Multiresolution.	120
Figura 37. Predicciones de los modelos de la arquitectura Multiresolution.	121
Figura 38. Gráficas de los entrenamientos de la arquitectura U-Net estándar.	124
Figura 39. Predicciones de los modelos de la arquitectura U-Net estándar.	125
Figura 40. Predicción de un objeto con los modelos V1, V7, V12 y V17.	127
Figura 41. Predicción de un objeto con los modelos V6, V11.5, V12 y V17.	129
Figura 42. Predicciones del desdoblamiento de fase entre el mejor vs peor entrenamiento.	132
Figura 43. Más ejemplos del modelo v11.	133
Figura 44. Ejemplo de objetos reales con franjas proyectadas y su mapa de fase.	136
Figura 45. Inferencias de todos los modelos entrenados para un objeto (sol).	136
Figura 46. Inferencias de todos los modelos entrenados para un objeto (corazones).	138
Figura 47. Inferencias de todos los modelos entrenados para un objeto (murciélago).	140
Figura 48. Inferencias arquitectura U-Net con SQDM para objetos reales.	142
Figura 49. Histogramas de los mapas de fase de objetos sintéticos.	144
Figura 50. Histogramas de los mapas de fase de objetos reales.	144
Figura 51. Mapas de fase con configuraciones de aumento de datos.	145
Figura 52. Ejemplo de los efectos aplicados para aumento de datos en los mapas de fase.	146
Figura 53. Curva de entrenamiento de fine-tuning con aumento de datos.	147
Figura 54. Inferencias V6 vs V6 con fine-tuning con aumento de datos.	148

Figura 55. Desdoblamiento de fase métodos tradicionales vs predicción CNN.	151
Figura 56. Desdoblamiento de fase métodos tradicionales vs predicción CNN mesh.....	152
Figura 57. Desdoblamiento de fase CCN vs Métodos tradicionales 3D.....	154
Figura 58. Desdoblamiento de objeto CNN vs métodos tradicionales 2D	155
Figura 59. Desdoblamiento de fase métodos tradicionales vs mejor modelo.	158
Figura 60. Desdoblamiento de fase métodos tradicionales vs mejor modelo.	158
Figura 61. Desdoblamiento de fase métodos tradicionales vs mejor modelo	159
Figura 62.Desdoblamiento de fase métodos tradicionales vs mejor modelo.	160

Índice de tablas

Tabla 1. Ejemplos de métodos tradicionales de fase.	56
Tabla 2. Estado del arte en DL para desdoblamiento de fase y reconstrucción 3D.	59
Tabla 3. Valores promedio Métodos tradicionales.	94
Tabla 4. Resultados de fine-tuning con el modelo pre-entrenado.	104
Tabla 5. Resultados comparativos de los entrenamientos con diferentes modelos.	108
Tabla 6. Comparación cuantitativa entre mejor método tradicional y modelo CNN.....	156

I. Introducción

El análisis y reconstrucción de información tridimensional a partir de imágenes es un componente fundamental en múltiples áreas científicas, industriales y tecnológicas, donde la necesidad de obtener mediciones precisas de la geometría de un objeto resulta esencial para tareas de inspección, control dimensional, caracterización de materiales y evaluación de fenómenos físicos. En particular, disciplinas como la interferometría, la metrología óptica y la perfilometría por proyección de franjas (Fringe Projection Profilometry, FPP) han demostrado ser herramientas indispensables para obtener representaciones detalladas de objetos mediante la modulación y análisis de patrones sinusoidales proyectados sobre su superficie (Su & Chen, 2004; Takeda, Ina, & Kobayashi, 1982). En este tipo de técnicas, la información de fase recuperada a partir de la deformación de las franjas juega un papel central, ya que codifica la topología del objeto de interés.

No obstante, la fase capturada mediante FPP se obtiene inicialmente en forma envuelta dentro del intervalo $[-\pi, \pi]$, lo cual introduce discontinuidades periódicas que no corresponden a características físicas del objeto, sino a una limitación matemática del proceso de adquisición. Para recuperar la geometría real, es necesario aplicar un procedimiento denominado desdoblamiento de fase (phase unwrapping), cuyo objetivo consiste en transformar la fase envuelta en un mapa continuo que represente con precisión la forma del objeto. Desde el trabajo pionero de Itoh (1982), el desdoblamiento de fase se ha consolidado como un paso crítico en prácticamente todas las aplicaciones ópticas basadas en análisis de franjas.

A pesar de su relevancia, el desdoblamiento de fase continúa enfrentando desafíos significativos. En condiciones reales, las imágenes suelen verse afectadas por ruido, variaciones de iluminación, sombras, texturas complejas, saturación, reflectancias no uniformes y geometrías con gradientes pronunciados. Estas condiciones dificultan la estabilidad de los algoritmos tradicionales y pueden causar errores locales que se propagan y deforman el mapa de fase absoluto. Métodos clásicos ampliamente utilizados —incluyendo enfoques basados en gradientes, rutas de fiabilidad, formulaciones variacionales y optimización global, han demostrado buen desempeño bajo escenarios controlados; sin

embargo, su funcionamiento se ve limitado ante condiciones experimentales adversas, lo que conduce a artefactos, pérdida de continuidad y distorsiones en reconstrucciones tridimensionales complejas (Ghiglia & Romero, 1994; Herráez, Burton, Lalor, & Gdeisat, 2002).

Debido a estas limitaciones, el desdoblamiento de fase se considera un problema inverso particularmente sensible y multidimensional, donde pequeñas variaciones en la calidad del mapa envuelto pueden producir diferencias significativas en la fase resultante. Esta complejidad ha motivado la búsqueda de métodos más robustos y adaptativos, capaces de manejar condiciones realistas sin requerir modelos estrictos o supuestos idealizados sobre la estructura del ruido o la geometría del objeto.

En este contexto, el aprendizaje profundo ha emergido como una alternativa poderosa para resolver problemas inversos asociados a la visión computacional y el procesamiento de imágenes. Las redes neuronales profundas, especialmente las redes neuronales convolucionales (CNN), han demostrado una notable capacidad para extraer características relevantes en presencia de ruido, capturar dependencias espaciales complejas y reconstruir señales físicas con alta fidelidad (LeCun, Bengio, & Hinton, 2015). Arquitecturas como las redes residuales (He, Zhang, Ren, & Sun, 2016) y U-Net (Ronneberger, Fischer, & Brox, 2015) han sido ampliamente utilizadas debido a su capacidad para preservar detalles locales mientras modelan patrones de alto nivel, lo que las convierte en candidatas prometedoras para mejorar la estabilidad y precisión del desdoblamiento de fase.

En años recientes, distintas investigaciones han comenzado a explorar el potencial del aprendizaje profundo aplicado al análisis de patrones de franjas y a la recuperación de fase, demostrando que las CNN pueden superar varias de las limitaciones de los métodos deterministas clásicos, especialmente al enfrentar ruido estructurado, discontinuidades o mapas multi-wrap (Feng et al., 2019; Perera & De Silva, 2020). Estos resultados indican que el modelado automático de relaciones espaciales complejas podría representar un avance significativo para la reconstrucción óptica tridimensional en aplicaciones reales.

Bajo este panorama, el presente trabajo propone el desarrollo, implementación y análisis comparativo de una red neuronal convolucional diseñada específicamente para mejorar el desdoblamiento de fase espacial en escenarios tanto sintéticos como reales. El enfoque se centra en evaluar la capacidad del aprendizaje profundo para reducir el ruido, preservar detalles estructurales, aumentar la continuidad del mapa de fase y disminuir el error acumulado durante el proceso de reconstrucción tridimensional en comparación con algoritmos tradicionales.

Para ello, se realizaron múltiples experimentos que abarcan métodos clásicos de desdoblamiento, arquitecturas profundas como U-Net, U-Net con SQD-LSTM, U-Net Res-Net y modelos Multiresolution, así como variaciones en los tipos de mapas de fase (low-wrap y multi-wrap), resolución espacial y tamaño del conjunto de entrenamiento. El análisis cuantitativo incluye métricas estándar como MSE, PSNR y SSIM, además de una métrica enmascarada específica para esta tesis (IM-MSE) que enfatiza regiones de interés. Complementariamente, se realizan evaluaciones cualitativas sobre predicciones sintéticas y sobre objetos reales adquiridos mediante FPP.

Esta tesis busca, en última instancia, demostrar la viabilidad del aprendizaje profundo como herramienta para el desdoblamiento de fase, identificar las condiciones que favorecen su desempeño, establecer comparaciones sistemáticas con métodos tradicionales y contribuir con evidencia experimental al avance de técnicas modernas de reconstrucción óptica. Los resultados aquí presentados tienen el potencial de fortalecer aplicaciones científicas y tecnológicas que dependen de mediciones tridimensionales precisas, aportando soluciones más robustas, eficientes y adaptables.

1.1 Planteamiento del problema

El principal problema es la presencia de ruido en las imágenes utilizadas en el desdoblamiento de fase espacial, que afecta a la precisión de los resultados. Este ruido introduce ambigüedades de fase que deben ser resueltas para obtener datos precisos y consistentes, lo cual es muy importante en este proceso, es esencial para obtener

mediciones confiables y extraer información significativa y no errónea de los datos.

Debido a la variedad de aplicaciones, es necesario desarrollar e implementar algoritmos de desdoblamiento de fase más confiables que consuman menos recursos computacionales. El éxito o el fracaso del desdoblamiento de fase afecta significativamente el rendimiento general de las aplicaciones (Olmedo et al., 2020).

Si bien los métodos tradicionales de procesamiento de señales se utilizan comúnmente para reducir el ruido en desdoblamiento el desdoblamiento de fase espacial, estas técnicas a menudo sufren desafíos significativos (Judge & Bryanston-Cross, 1994b).

Los avances recientes en el aprendizaje profundo prometen resolver estos problemas mediante el uso de las redes neuronales para aprender asignaciones complejas entre datos de entrada ruidosos y señales de salida limpias. Para dar respuesta a este desafío, que tiene como objetivo reducir el ruido en el sistema de clasificación espacial de imágenes reales y sintéticas, ha surgido la necesidad de desarrollar y utilizar técnicas de aprendizaje profundo. La comparación de estas técnicas con los métodos tradicionales de procesamiento de señales también se plantea como parte importante para resolver este problema, para determinar sus ventajas y que como consecuencia se lleve a una de reducción de ruido, aumento de precisión y eficiencia.

1.2 Justificación

El desdoblamiento de fase espacial (DFE) es el paso final en el proceso de extracción de fase en el procesamiento de imágenes. Esta técnica tiene muchos usos en múltiples campos científicos, como en la interferometría, la tomografía de coherencia óptica y la microscopía holográfica. Sin embargo, uno de los principales desafíos de este proceso es el ruido de las imágenes, que puede degradar el resultado de la reconstrucción y hacer que la imagen se vea mal, como se puede ver en la *Figura 4*. Es importante eliminar las ambigüedades de fase causadas por valores restringidos dentro de un rango específico, normalmente entre $-\pi$ y π . Tales limitaciones dan lugar a "envolturas" en los datos de fase, que deben resolverse para obtener información de fase precisa y completa (Abdul-Rahman et al., 2007).



Figura 1. Ejemplo de reconstrucción de objeto 3D con y sin ruido. La imagen ilustra la comparación entre un objeto 3D original (una esfera) y su reconstrucción simulando la técnica de proyección de franjas. La reconstrucción muestra franjas y ruido (picos), representando irregularidades en los datos de imagen. Imagen generada por ChatGPT de OpenAI.

La investigación se basa en la necesidad de reducir este ruido y mejorar la calidad de las imágenes en el proceso de desdoblamiento de fase espacial, la capacidad del aprendizaje profundo para resolver este problema utilizando redes neuronales muestra una nueva alternativa para mejorar la precisión y confiabilidades desarrollando técnicas más efectivas que superen las limitaciones de los métodos tradicionales.

El potencial del aprendizaje profundo para abordar este desafío, mediante el aprovechamiento de redes neuronales convolucionales, sugiere una nueva dirección prometedora para mejorar la exactitud del desdoblamiento de fase espacial. Las ventajas del aprendizaje profundo sobre las técnicas tradicionales de procesamiento de señales para resolver los desafíos de desdoblamiento de fase espacial son muchas, incluidas la capacidad de reducción de ruido, el desdoblamiento de fase eficaz y puntual además de que funciona en una variedad de entornos incluso en escenarios de desdoblamientos complejos. Se han propuesto modelos de aprendizaje profundo como SFNet como métodos competentes de desdoblamiento de fase espacial altamente eficientes y robustos, lo que demuestra la capacidad de las redes neuronales para mejorar los procesos de desdoblamiento (Zhang et al.,

2024). Además, el mapeo cuantitativo de susceptibilidad basado en el aprendizaje profundo ha demostrado la eficacia de UNET en el procesamiento directo de mapas de fase no envueltos para mejorar la calidad de los resultados(Hanspach et al., 2022).

El sólido enfoque de desdoblamiento de fase de aprendizaje profundo de un solo paso se ha mostrado prometedor para proporcionar resultados de extracción de fase precisos y eficientes (K. Wang et al., 2019). Además, el desdoblamiento de fase basado en el aprendizaje profundo para la tomografía holográfica ha empleado Attention U-Net con RB para la segmentación, lo que mostró avances en el aprendizaje profundo para aplicaciones de desdoblamiento de fases (Gontarz et al., 2023).

Esta investigación tiene una gran oportunidad para mejorar la calidad de las imágenes desdobladas, suministrando técnicas más eficaces de reducción de ruido que superen las limitaciones de los métodos tradicionales, promover el uso del aprendizaje profundo y, al mismo tiempo, explorar nuevas técnicas de redes neuronales en el campo, establecer una comparación sistemática entre las técnicas basadas en CNN y los métodos tradicionales para guiar futuras investigaciones y desarrollos tecnológicos, y contribuir a diversas aplicaciones mejorando herramientas esenciales en la ciencia y la tecnología, con impactos potenciales en la interferometría, la tomografía óptica y más. Por lo tanto, esta investigación no solo es relevante para la educación y el ámbito académico, sino que también puede tener un impacto significativo en varias aplicaciones prácticas, y proporcionar grandes beneficios para la industria y la investigación.

El principal problema es la presencia de ruido en las imágenes utilizadas en el desdoblamiento de fase espacial, que afecta a la precisión de los resultados. Este ruido introduce ambigüedades de fase que deben ser resueltas para obtener datos precisos y consistentes, lo cual es muy importante en este proceso, es esencial para obtener mediciones confiables y extraer información significativa y no errónea de los datos.

Debido a la variedad de aplicaciones, es necesario desarrollar e implementar algoritmos de desdoblamiento de fase más confiables que consuman menos recursos computacionales. El

éxito o el fracaso del desdoblamiento de fase afecta significativamente el rendimiento general de las aplicaciones (Olmedo et al., s/f).

Si bien los métodos tradicionales de procesamiento de señales se utilizan comúnmente para reducir el ruido en desdoblamientoel desdoblamiento de fase espacial, estas técnicas a menudo sufren desafíos significativos (Judge & Bryanston-Cross, 1994a).

Los avances recientes en el aprendizaje profundo prometen resolver estos problemas mediante el uso de las redes neuronales para aprender asignaciones complejas entre datos de entrada ruidosos y señales de salida limpias. Para dar respuesta a este desafío, que tiene como objetivo reducir el ruido en el sistema de clasificación espacial de imágenes reales y sintéticas, ha surgido la necesidad de desarrollar y utilizar técnicas de aprendizaje profundo.

La comparación de estas técnicas con los métodos tradicionales de procesamiento de señales también se plantea como parte importante para resolver este problema, para determinar sus ventajas y que como consecuencia se lleve a una de reducción de ruido, aumento de precisión y eficiencia.

II. Hipótesis

Mediante la implementación de una CNN en el proceso de desdoblamiento de fase espacial se reduce el ruido de la fase y con ello el error dentro del proceso de reconstrucción tridimensional en comparación contra métodos tradicionales.

III. Objetivos

3.1.1 Objetivo principal

Desarrollar e implementar una CNN para mejorar el desdoblamiento de fase espacial en objetos sintéticos y reales, para la reducción de ruido en el mapa de fase y del error en el proceso de reconstrucción tridimensional, así como llevar a cabo una comparación con métodos tradicionales con el fin de evaluar su rendimiento.

3.1.2 Objetivos específicos

- i. Implementar y evaluar métodos tradicionales de desdoblamiento de fase espacial en objetos sintéticos y reales.
- ii. Diseñar y entrenar una arquitectura de CNN para el desdoblamiento de fase espacial.
- iii. Implementar la CNN para la reconstrucción 3D, de objetos sintéticos y reales.
- iv. Establecer métricas cuantitativas para evaluar la reducción de ruido y error del desdoblamiento de fase, como por ejemplo Error Cuadrático Medio (MSE), Peak Signal-to-Noise Ratio (PSNR) y Structural Similarity Index (SSIM).
- v. Documentar y publicar los resultados obtenidos en algún congreso internacional.

IV. Antecedentes

Los seres humanos captan las imágenes y observan el mundo a través del proceso de la visión, donde utiliza un sistema en conjunto de ojos-cerebro donde se procesa la luz a través de la retina, ver la *Figura 1* para una mejor comprensión de la anatomía humana. Es en la retina, una parte del ojo, en la que las células fotorreceptoras convierten la luz en energía eléctrica. Estas señales se envían al cerebro a través del nervio óptico y se procesan en la corteza visual para interpretar la luz recibida y darle sentido. Este complejo proceso permite a las personas ver detalles, colores y movimientos, y percibir el entorno de forma rápida y precisa (Kandel et al., 2021; Purves & others, 2004).

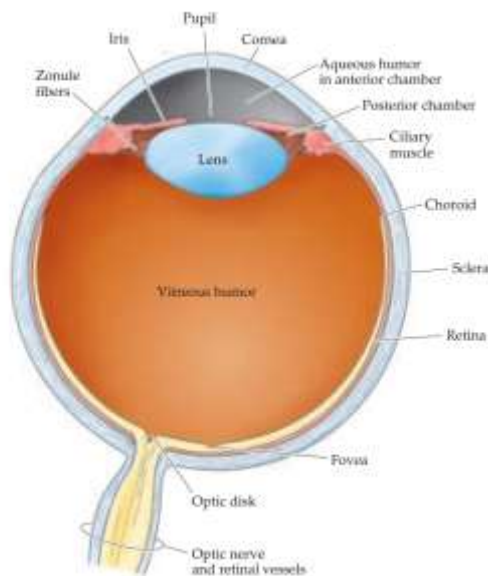


Figura 2. Anatomía de un ojo.

Visión humana, anatomía del ojo humano. Recuperado de (Purves & others, 2004)

Existen diferentes áreas de la inteligencia artificial como el procesamiento del lenguaje natural, la robótica, el aprendizaje automático y la visión por computadora. La visión por computadora se refiere a algoritmos desarrollados con sistemas diseñados para ayudar a las computadoras a interpretar y comprender imágenes y videos digitales. Nació de la necesidad

de desarrollar tareas que requieran visión humana, como control industrial, robots autónomos y análisis de imágenes médicas (Lecun et al., 2015).

Comprender la visión humana es importante para el desarrollo de la visión por computadora porque muchos algoritmos intentan imitar los mecanismos biológicos del ojo y el cerebro. Por ejemplo, las redes neuronales convolucionales (CNN) inspiradas en estructuras y funciones que son característica del humano son muy útiles en el reconocimiento de imágenes (Krizhevsky et al., s/f).

En el campo de la visión por computadora, la reconstrucción 3D es el proceso de capturar la forma tridimensional de un objeto o escena a partir de una imagen bidimensional. Se basa en técnicas como la visión estereoscópica, la fotogrametría y el uso de sensores de profundidad. La reconstrucción 3D permite aplicaciones en campos como la realidad aumentada, la fabricación digital, la arqueología y la medicina (Hartley & Zisserman, 2004; Seitz et al., 2006) .

Un aspecto clave de la reconstrucción 3D es la capacidad de interpretar con precisión los datos de imagen para generar un modelo tridimensional exacto. Aquí es donde entra en juego el desdoblamiento de fase espacial, conocido como phase unwrapping en inglés. Muchas técnicas de reconstrucción 3D y métodos avanzados de procesamiento de imágenes, como la interferometría y la holografía, se basan en medir con precisión la fase de la luz o las ondas reflejadas por un objeto.

El desdoblamiento de fase espacial, phase unwrapping en inglés, es un proceso crucial en el procesamiento de imágenes que busca transformar mediciones de fase envueltas en representaciones continuas. Este procedimiento es esencial en aplicaciones como la interferometría SAR (Synthetic Aperture Radar), holografía digital y tomografía de coherencia óptica, donde la precisión en la reconstrucción de la fase es vital para la interpretación de datos (Su & Chen, 2004), en la *Figura 2* podemos observar un ejemplo del proceso de reconstrucción 3D donde se visualiza en qué paso se encuentra el proceso de desdoblamiento de fase.



Figura 3. Diagrama del proceso de reconstrucción 3D.

Ejemplo en el contexto del procesamiento de imágenes, destacando las principales etapas desde la captura de imágenes hasta la reconstrucción del modelo 3D.

Para superar los problemas y desafíos que surgen en el proceso de desdoblamiento de fase, se incluyeron técnicas tradicionales como la integración directa, métodos basados en gradientes y algoritmos de red de corte, principalmente. Estos métodos fueron eficientes en la mayoría de los casos, pero presentan ciertas limitaciones, especialmente en escenarios con características específicas como el ruido elevado, sombras, geometrías complejas, o discontinuidades en los datos. Este tipo de obstáculos o limitaciones han sido estudiados, se han presentado revisiones exhaustivas sobre estos primeros métodos tradicionales cubriendo sus limitaciones en el caso de ruido (Vidal et al., 1999). Métodos como el algoritmo de mínimos cuadrados ponderados y el algoritmo de gradiente conjugado han sido utilizados para mitigar algunos de estos problemas, pero aún enfrentan desafíos importantes en situaciones de mucho ruido.

Las técnicas de desdoblamiento de fase han mejorado enormemente con el tiempo, a través de varios métodos y algoritmos implementados para resolver los problemas asociados con el desenvolvimiento de datos de fase con precisión. El algoritmo de desdoblamiento de fase 3D propuesto por Cusack y Papadakis, uno de los algoritmos iniciales, fue diseñado específicamente para desenvolver de manera eficiente los volúmenes de fase de resonancia magnética (Abdul-Rahman et al., 2007). Los principios anteriores marcaron el comienzo de una serie de desarrollos en las técnicas de desenvolvimiento de fase.

El objetivo principal de las técnicas de desenvolvimiento de fase es eliminar las ambigüedades del módulo 2π contenida en los mapas de fase envueltos (Estrada et al., 2011). Estas distorsiones ocurren de vez en cuando dada su naturaleza periódica, y el desenvolvimiento es importante obtener más información sobre la pieza que se está

investigando. Por ejemplo, se ha propuesto un enfoque que utiliza la conectividad espacial y la estimación del ruido para dirigir la ruta de desdoblamiento. Este método garantiza que las áreas menos afectadas por el ruido se desenvuelvan primero, lo que ayuda a mitigar la frecuencia de repetición y mejora el gráfico de salida (Liu et al., 2013).

Entre otras propuestas se incluyen métodos que utilizan rutas en el dominio espacial y ajustes lineales en el dominio temporal. (Li et al., 2014) Los algoritmos de crecimiento de regiones también son comúnmente empleados, donde el desdoblamiento comienza en regiones de fase estables y progresa en función de la coherencia de fase (Robinson et al., 2014). Estos métodos tradicionales han sido vitales para aplicaciones como el monitoreo de deslizamientos de tierra y las mediciones de formas en 3D.

Otras ideas son métodos que utilizan métodos espaciales y ejes de tiempo. El desdoblamiento espacial aprovecha la alta correlación de fase espacial y es útil en aplicaciones como el monitoreo de deslizamientos de tierra y mediciones 3D. Sin embargo, puede ser computacionalmente intensivo y desafiante en objetos complejos. Por otro lado, el desdoblamiento temporal se basa en la evolución suave de la fase a lo largo del tiempo, siendo ventajoso en escenarios donde el desdoblamiento espacial es complejo o computacionalmente costoso (Kurian et al., 2024).

Para desarrollar técnicas modernas para resolver estos problemas ha contribuido mucho el avance de la tecnología y la aparición de técnicas de aprendizaje profundo (Deep Learning), donde se han perfeccionado enfoques más avanzados para el desdoblamiento de fase. Se ha demostrado que las redes neuronales convolucionales (CNN) mejoran la precisión y eficiencia de la segmentación residual en comparación con los métodos tradicionales.

Las CNN han mostrado un rendimiento excelente en tareas como el reconocimiento de imágenes, el reconocimiento de objetos y la segmentación de imágenes (Y. Sun et al., 2020a). Pueden utilizar filtros entrenados para aprender características de objetos complejos y combinar características según la entrada, lo que da como resultado una jerarquía de características cada vez más abstractas (Ji et al., 2013). La profundidad de CNN es una función importante para estimar funciones continuas con alta precisión, lo que la convierte

en un aproximador de función universal cuando la red es lo suficientemente profunda (Zhou, 2020). Además, las arquitecturas CNN tienen un gran impacto en su rendimiento, siendo el diseño de la arquitectura de red un factor crítico para lograr una alta precisión en las tareas (Y. Sun et al., 2020b). Se han realizado varios avances en las arquitecturas CNN, incluido el uso de redes convolucionales 3D para el aprendizaje de características espaciotemporales (Tran et al., 2015), y el desarrollo de redes de segmentación profunda para la detección de tareas (ZHOU et al., 2019).

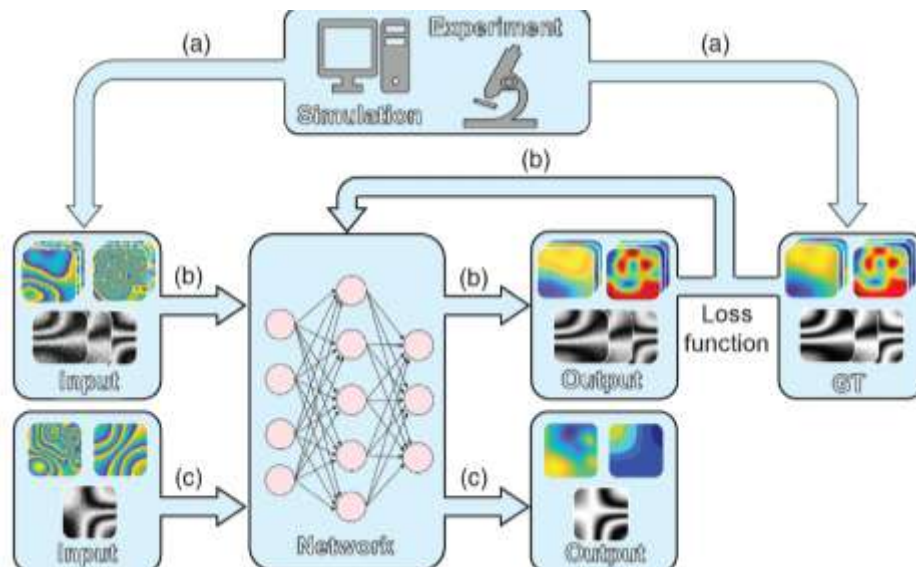


Figura 4. Ejemplo de flujo de trabajo.

Flujo de trabajo integrado para la simulación y experimentación en el contexto de redes neuronales aplicadas al procesamiento de imágenes. Recuperado de (K. Wang et al., 2022)

Una revisión comparativa examina métodos de desdoblamiento de fase basados en aprendizaje profundo y métodos tradicionales, demuestra el éxito de las CNN en términos de precisión y eficiencia y proporciona un análisis detallado de diferentes métodos de desdoblamiento de fase espacial, demostrando que los métodos basados en aprendizaje profundo, como desempaquetado de red generativa profunda (dRG) y desempaquetado de codificación de ondas profundas (dWC), son superiores a los métodos tradicionales en términos de precisión y capacidad para manejar problemas como el ruido, las discontinuidades y el aliasing, la Figura 3 corresponde al flujo de trabajo utilizado en esta investigación comparativa (K. Wang et al., 2022)

Los modelos mixtos y los Transformers también lograron buenos resultados. En un estudio reciente (Z. Zhao et al., 2022), se propuso un algoritmo robusto que combina el ajuste polinómico de Zernike con la red Swin-Transformer, lo que demuestra el poder del aprendizaje profundo para mejorar la precisión y confiabilidad del desdoblamiento de fase. Además, en el ámbito geológico, se empleó una CNN con una arquitectura codificador-decodificador para la detección automatizada de deformaciones en la superficie volcánica (J. Sun et al., 2020). Este enfoque demuestra la capacidad del aprendizaje profundo para suprimir el ruido y extraer información valiosa de datos geográficos complejos.

Se han desarrollado herramientas especiales de aprendizaje profundo como PEENet y Hformer para su uso en aplicaciones como la perfilometría de proyección de franjas y la medición interferométrica. Estos modelos utilizan redes neuronales profundas para mejorar la precisión del desdoblamiento de fase en entornos complejos. En una reciente investigación (Zhu et al., 2021), se utilizaron redes neuronales profundas para volver a plantear el desdoblamiento de fase como un problema de segmentación semántica, lo que ilustra lo versátil que puede ser el aprendizaje profundo.

4.1 Desdoblamiento de fase

Uno de los principales conceptos de esta investigación es el desdoblamiento de fase, el cual debemos entender a fondo para abordar la principal motivación del trabajo. El desdoblamiento de fase es un proceso utilizado en métodos como la Perfilometría de Fourier (FTP), la Perfilometría de cambio de Fase (PSP) o técnicas de modelado 3D similares que se basan en la proyección de patrones de franjas sobre un objeto, ver *Figura 5*, representando uno de los últimos pasos en estas metodologías.

Los patrones de franjas en las imágenes aportan y contienen características importantes sobre la información del objeto como algunas de sus propiedades físicas. Cuando se proyecta un patrón de líneas sobre un objeto y se captura su forma, la información 3D del objeto se codifica en la fase de las franjas. La información de la altura está envuelta, lo que significa que la fase se mide solo dentro de un rango pequeño con limitaciones, en una fase de discontinuidades típicamente entre $-\pi$ y π o de 0 a 2π . Por lo tanto, esta imagen debe ser

procesada y no puede ser utilizada hasta que se remueven las discontinuidades, el proceso mediante el cual esta envoltura se elimina es el desdoblamiento de fase, permitiendo que la fase medida se represente como una función continua y sin discontinuidades a lo largo del campo observado (Olmedo et al., s/f; Servin et al., 2014).

El desdoblamiento de fase permite interpretar la información codificada en los patrones de franjas proyectadas sobre el objeto, lo que genera importante información que tiene diferentes aplicaciones, entre ellas el de reconstruir modelos tridimensionales mediante las características extraídas de los objetos. Los principales desafíos son el ruido, las discontinuidades y ambigüedades en la fase envuelta. (Huntley & Saldner, 1997)

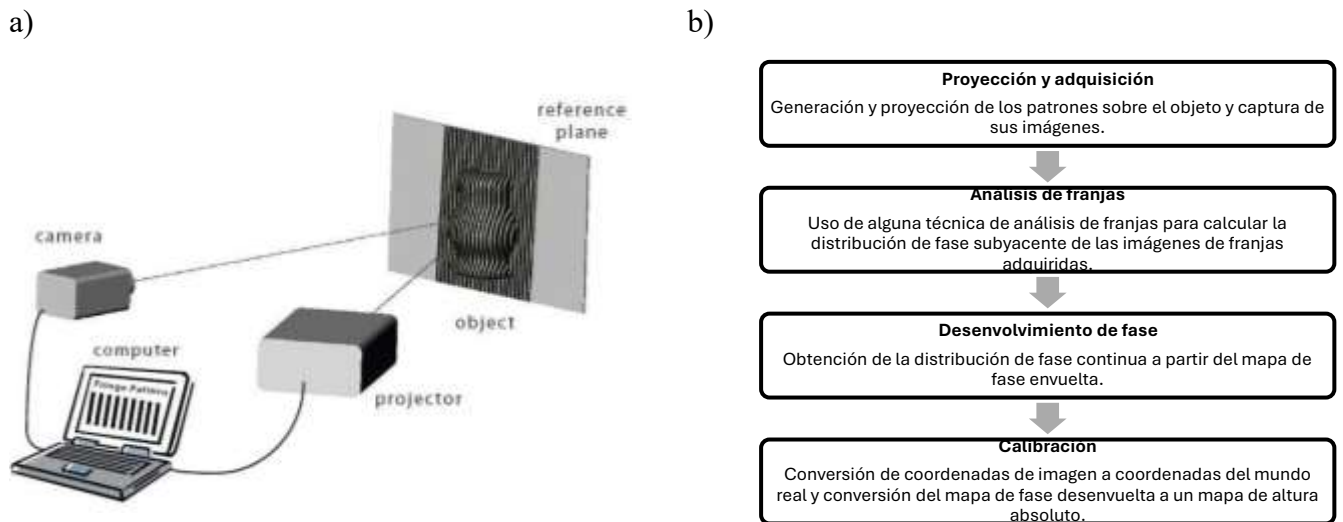


Figura 5. Sistema y diagrama de perfilometría por proyección de franjas.

a) Sistema de perfilometría por proyección de franjas b) El diagrama muestra el proceso secuencial de un sistema de perfilometría por proyección de franjas. Recuperado de (Gorthi & Rastogi, 2010b)

Hay dos tipos de desenvolvimiento de fase según su naturaleza o aplicación: el desenvolvimiento de fase temporal y el desenvolvimiento de fase espacial. Su principal diferencia consiste en la dimensión y el contexto en el que se aplican. El primero consiste en resolver discontinuidades de fase en señales bidimensionales que varían en el tiempo, asegurando que la fase de la señal se interprete de manera continua en el dominio temporal. Esto es crucial en el procesamiento de señales de audio y video, así como en comunicaciones de fase. El segundo consiste en resolver discontinuidades de fase en imágenes y señales

bidimensionales o tridimensionales, evaluando las relaciones espaciales entre múltiples puntos. Este enfoque es fundamental en aplicaciones como la topografía, las imágenes médicas (MRI, CT) y la interferometría óptica y de radar (SAR).

En esta investigación nos enfocaremos en el desenvolvimiento de fase espacial.

4.2 Métodos de desdoblamiento de fase espaciales tradicionales

Los métodos tradicionales de desdoblamiento de fase se desarrollaron antes del uso de técnicas basadas en aprendizaje profundo y se fundamentan en principios matemáticos deterministas, tales como diferencias de fase, integración de gradientes, ecuaciones de Poisson y optimización discreta. Estos algoritmos operan directamente sobre la fase envuelta y recuperan la fase continua sin recurrir a modelos entrenables ni redes neuronales, apoyándose únicamente en supuestos físicos y estructura matemática del problema.

En la literatura, estos métodos suelen agruparse en dos categorías principales:

- **Métodos locales**, que desenvuelven la fase siguiendo una trayectoria punto a punto, guiándose por la continuidad local o por medidas de fiabilidad. Estos métodos son sensibles al ruido acumulado, pero pueden preservar discontinuidades reales de la fase.
- **Métodos globales**, que formulan el desdoblamiento como un problema de optimización o integración global, buscando minimizar un error energético en toda la imagen. Tienden a ser más robustos al ruido y evitan la propagación de errores locales, aunque pueden suavizar discontinuidades fuertes.

Considerando la amplia variedad de algoritmos existentes en esta tesis se seleccionaron e implementaron métodos representativos de ambas categorías, incluyendo los enfoques de Itoh, Ghiglia y Romero, Costantini, Herráez, PUMA, PEARLS, Zhao y Martínez-Carranza. Estos algoritmos permiten analizar comparativamente el comportamiento de los principales enfoques tradicionales antes de introducir los métodos basados en redes neuronales profundas.

4.2.1 Itoh (1982)

El algoritmo de Itoh constituye la base teórica del desdoblamiento de fase espacial. Se fundamenta en la comparación entre la diferencia de fase real y su diferencia envuelta, estableciendo que la fase puede recuperarse correctamente siempre que el gradiente real sea menor que π . El método integra directamente los gradientes envueltos aplicando el operador de envoltura $W(\cdot)$, lo que permite reconstruir la fase continua en problemas unidimensionales o con variaciones suaves. Sin embargo, en imágenes interferométricas reales, donde existen discontinuidades, ruido elevado o aliasing, la condición de Itoh suele violarse, generando saltos erróneos que se propagan en la reconstrucción. Aun así, este criterio constituye un fundamento esencial sobre el cual se han construido numerosos algoritmos posteriores (Itoh, 1982).

4.2.2 Ghiglia & Romero (1994–1998)

Ghiglia y Romero desarrollaron uno de los métodos más influyentes para desdoblamiento de fase 2D, basado en un modelo de mínimos cuadrados. Su formulación establece que el desdoblamiento puede resolverse como una ecuación de Poisson aplicada a los gradientes envueltos. En el caso no ponderado, la solución se obtiene eficientemente mediante la Transformada Discreta del Coseno (DCT), lo que permite un cálculo rápido y estable en mallas grandes. Para el caso ponderado, se introduce una matriz de pesos para controlar regiones ruidosas, requiriendo métodos iterativos como la iteración de Picard o el gradiente conjugado preconditionado. Este enfoque ofrece alta robustez y estabilidad, siendo especialmente adecuado en datos densos y sin grandes discontinuidades (Ghiglia & Romero, 1994).

4.2.3 Costantini (1998)

Costantini propuso un método de desdoblamiento formulado como un problema global de optimización entera, donde los errores de $\pm 2\pi$ se modelan explícitamente como variables discretas. El algoritmo traduce el problema en una red de flujo de costo mínimo, lo que

permite identificar inconsistencias globales y resolverlas sin dispersar errores. A diferencia de los enfoques locales, este método es capaz de preservar discontinuidades reales y manejar regiones ruidosas con mayor coherencia global. Su principal ventaja radica en su robustez, aunque la formulación implica un mayor costo computacional que los métodos basados en mínimos cuadrados (Costantini, 1998).

4.2.4 Herráez et al. (2002)

El método de Herráez pertenece a la familia de los quality-guided algorithms. Introduce una métrica de fiabilidad basada en segundas diferencias locales de la fase envuelta, permitiendo priorizar el desdoblamiento en las regiones más confiables. En lugar de seguir trayectorias continuas, el algoritmo ordena las aristas según su fiabilidad, limitando así la propagación de errores desde zonas ruidosas hacia regiones limpias. Esto lo hace especialmente adecuado para imágenes experimentales con ruido moderado o discontinuidades suaves, y lo convierte en uno de los métodos más utilizados en procesamiento óptico (Herráez et al., 2002).

4.2.5 PUMA (Bioucas-Dias & Valadão, 2007)

PUMA (Phase Unwrapping Max-flow/Min-cut) reformula el desdoblamiento como una minimización de energía en un grafo. Para potenciales convexos, obtiene soluciones exactas mediante cortes de grafo, constituyendo un método global robusto que preserva contornos y bordes. En el caso de potenciales no convexos, típicos en escenarios con discontinuidades fuertes, utiliza un esquema *majorize–minimize* que aproxima la solución de manera estable. Es uno de los métodos más confiables en presencia de ruido y discontinuidades, aunque su implementación puede ser más compleja y su costo computacional mayor que métodos directos (J. M. Bioucas-Dias & Valadão, 2007).

4.2.6 PEARLS (Bioucas-Dias et al., 2010)

El método PEARLS combina suavizado adaptativo y desdoblamiento. En la primera etapa, aplica filtrado local basado en aproximación polinómica y selección de ventana mediante el criterio ICI, lo que permite reducir ruido de forma adaptativa preservando estructuras relevantes. En la segunda etapa, utiliza PUMA para realizar el desdoblamiento sobre un mapa de fase menos ruidoso. Este enfoque en dos fases ha demostrado un alto rendimiento en condiciones adversas, aunque su complejidad computacional es mayor debido al filtrado adaptativo previo (J. Bioucas-Dias et al., 2008).

4.2.7 Zhao, Chen & Tan (Robust 2D, 2016)

Zhao y colaboradores desarrollaron un método robusto para desdoblamiento de fase empleando dos distribuciones de fase con distinta precisión. El mapa de baja resolución se utiliza para determinar las órdenes de franja, mientras que el mapa de alta resolución se corrige punto a punto utilizando esa información. Este enfoque se caracteriza por su simplicidad y por su capacidad para resolver discontinuidades, aunque su rendimiento depende de la calidad relativa entre ambas fases y puede ser sensible al ruido en el mapa grueso (H. Zhao et al., 1994).

4.2.8 Martínez-Carranza et al. (2012)

Este método utiliza la ecuación de transporte de intensidad (TIE) para obtener la fase absoluta sin necesidad de propagación de ondas. A partir de la fase envuelta, el algoritmo estima la derivada axial de la intensidad empleando una aproximación basada en Fresnel, y posteriormente recupera la fase absoluta resolviendo el Laplaciano mediante transformada de Fourier con condiciones de contorno de Neumann. Además, incorpora un esquema iterativo que mejora la estabilidad y exactitud en presencia de ruido o aliasing. Es un método independiente de ruta y relativamente eficiente, aunque su versión de un solo paso puede ser menos precisa que los enfoques de mínimos cuadrados tradicionales (Martinez-Carranza et al., 2017).

4.3 Ruido en el desenvolvimiento de fase

El ruido es una interferencia no deseada que afecta la calidad de las mediciones de fase. Este puede crear errores significativos en el proceso de desenvolvimiento de fase por lo que es imprescindible conocer los diferentes tipos de ruido y cómo abordarlos.

Tipos de ruido:

- Ruido Gaussiano: Este tipo de ruido sigue una distribución normal y es común en muchas aplicaciones porque se origina a partir de procesos térmicos o electrónicos.
- Ruido impulsivo: Caracterizado por picos repentinos, este ruido puede deberse a interferencias eléctricas o errores de transmisión.
- Ruido de cuantificación: Creado por la digitalización de señales analógicas, e incluye errores debido a la resolución baja del sistema de medición.
- Ruido debido a la geometría de los objetos: Este ruido es creado por la complejidad geométrica de los objetos, pudiendo provocar ocultamiento de las franjas proyectadas, o bien sombras que se generan cuando el objeto obstruye la proyección de las franjas.

4.4 Perfilometría por cambio de fase

La perfilometría de fase es un método de metrología óptica, alta tecnología, para medir las propiedades de la superficie con alta precisión. Se basa en la proyección de un patrón de franjas (principalmente senoidales) en la superficie de interés y el análisis de las diferencias de fase de estas franjas debido a la forma de la superficie, ver pasos del proceso en la *Figura 6*.

Proyección del patrón de franjas:

- Se proyecta un patrón de franjas senoidales sobre la superficie a medir.
- Las franjas proyectadas tendrán una variación de intensidad $I(x, y)$ que puede describirse como se muestra en la Ecuación (1):

$$I(x, y) = I_0(x, y) + I_m(x, y) \cos\left(2\pi \frac{x}{p} + \phi(x, y)\right) \quad \text{Ecuación (1)}$$

donde $I_0(x, y)$ es la intensidad promedio, $I_m(x, y)$ es la modulación de intensidad, p es el período de las franjas, y $\phi(x, y)$ es la fase.

- Captura de imágenes: Se capturan varias imágenes de la superficie con diferentes desplazamientos de fase del patrón de franjas (normalmente se utilizan 3 o más imágenes).

Para cada imagen, k -ésima imagen, con desplazamiento de fase δk , se pueden describir cómo se observa en la Ecuación (2):

$$I_k(x, y) = I_0(x, y) + I_m(x, y) \cos\left(2\pi \frac{x}{p} + \phi(x, y) + \delta k\right) \quad \text{Ecuación (2)}$$

- Cálculo de la fase: Utilizando las imágenes capturadas con diferentes desplazamientos de fase, se puede calcular la fase $\phi(x, y)$ mediante métodos de demodulación de fase. Un método común es el algoritmo de cuatro fases, que usa cuatro imágenes con desplazamientos de fase de $0, \pi/2, \pi, \text{ y } 3\pi/2$:

La fase $\phi(x, y)$ se obtiene como se muestra en la Ecuación (3):

$$\phi(x, y) = \arctang\left(\frac{I_4(x, y) - I_2(x, y)}{I_1(x, y) - I_3(x, y)}\right) \quad \text{Ecuación (3)}$$

- Reconstrucción de la topografía: La fase calculada $\phi(x, y)$ está relacionada con la altura de la superficie $z(x, y)$. Esta relación puede depender de la geometría del sistema y debe calibrarse.

$$z(x, y) = f(\phi(x, y)) \quad \text{Ecuación (4)}$$

- Finalmente, se obtiene el mapa de la topografía $z(x, y)$ de la superficie como se observa en la Ecuación (4).

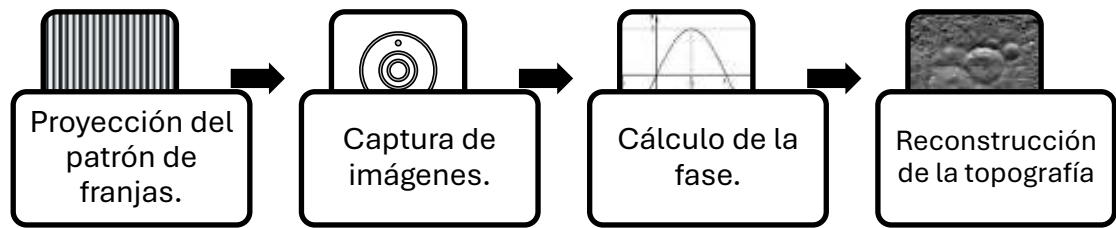


Figura 6. Diagrama del proceso de perfilometría de fase.

4.5 Phase shifting de n-step

El phase shifting de n-step, cambio de fase de n-pasos en español, es una técnica que utiliza múltiples imágenes con diferentes desplazamientos de fase para calcular la fase de la luz reflejada por un objeto. En la mayoría de los casos, se utilizan al menos tres o más desplazamientos (n-step) para lograr la mayor precisión. Las imágenes capturadas con diferentes desplazamientos o configuraciones de fase se combinan matemáticamente para completar la fase en cada punto del objeto. Se utiliza ampliamente en interferometría y perfilometría por cambio de fase debido a su capacidad para proporcionar mediciones precisas, sólidas y robustas (Feng et al., 2019).

4.6 Perfilometría de Fourier bidimensional

La perfilometría de Fourier bidimensional es una técnica ampliamente utilizada para la recuperación de la fase a partir de patrones de franjas, introducida formalmente por Takeda, Ina y Kobayashi (1982) para el análisis interferométrico mediante el método de la transformada de Fourier. En este enfoque, el patrón de franjas se analiza en el dominio espectral a través de la Transformada de Fourier 2D, donde se separan la componente continua y las bandas laterales asociadas a la frecuencia portadora del sistema (Takeda et al., 1982; Takeda & Mutoh, 1983).

A partir del espectro obtenido, se selecciona y filtra la banda lateral correspondiente a la frecuencia portadora f_0 utilizando un filtro paso banda adecuado, tal como es común en aplicaciones de proyección de franjas (Kema, 2007). Esta etapa elimina ruido y componentes no deseadas del espectro, conservando únicamente la información modulada por la fase del objeto. Una vez aislada la portadora, se aplica la Transformada Inversa de Fourier, lo que permite recuperar el mapa de fase envuelta limitado al intervalo $[-\pi, \pi]$.

4.7 Mapa de fase

Un mapa de fase es una representación visual y gráfica de los datos de fase obtenidos mediante técnicas interferométricas o de perfilometría, ver *Figura 7*. Un mapa de fase muestra la distribución espacial de la fase en una superficie o volumen. Estos mapas se crean a partir de las imágenes tomadas con desplazamientos de fase aplicando algoritmos específicos para calcular la fase en cada punto. Los mapas de fase envueltos necesitan ser desdoblados para obtener una fase continua que represente el campo geométrico real de la superficie (Malacara et al., 2005).

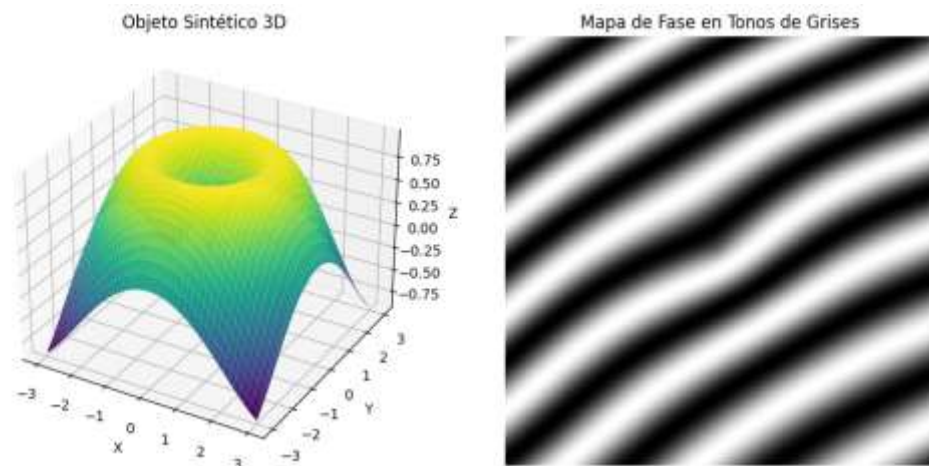
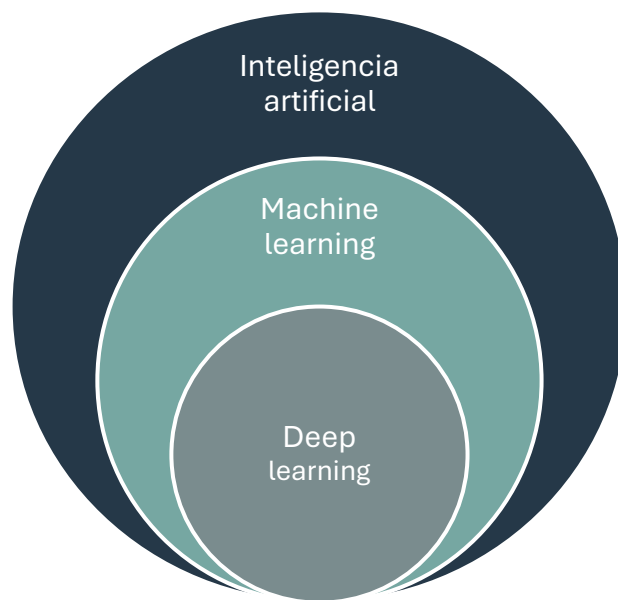


Figura 7. Representación de objeto 3D y su mapa de fase
El mapa de fase en la figura muestra la distribución espacial de la fase proyectada sobre un objeto sintético 3D utilizando un patrón de franjas en tonos de grises, creado con código en Python. Las variaciones en los tonos de gris representan los cambios de fase causados por la geometría del objeto. Las franjas más claras y oscuras indican diferentes niveles de fase, permitiendo observar cómo varía la fase en la superficie del objeto.

4.8 Inteligencia Artificial

La inteligencia artificial es un campo que busca emular los procesos de la inteligencia humana por medio de máquinas, especialmente en los sistemas informáticos, estos procesos incluyen el descubrimiento de información y reglas a partir de más datos, es decir el aprendizaje, también el uso de reglas para llegar a una conclusión, como un razonamiento y la autocorrección.

La IA es un campo innovador, que ha tomado fama y se ha vuelto popular por sus diferentes tipos de aplicaciones comerciales como son los asistentes virtuales, los sistemas de recomendación utilizados en plataformas de streaming y mercado electrónico, los vehículos autónomos o innovaciones como el diagnóstico médico asistido por IA. Este tiene diferentes subcampos como aprendizaje máquina, aprendizaje profundo, podemos observar estos campos en la *Figura 8*.



*Figura 8. Diagrama de Inteligencia artificial y sus subcampos
Recuperada de (Chollet, 2021).*

4.8.1 Machine learning

Machine Learning se traduce como aprendizaje automático, aunque se utiliza más su término en inglés para esta disciplina, este es una rama derivada de la IA y la informática, como podemos verlo en la figura 8. El aprendizaje automático se centra en el uso de datos y el desarrollo de algoritmos que ayudan a las computadoras a "aprender" características de los datos. Por lo general, estos se entrenan utilizando grandes cantidades de datos, lo que permite que la máquina aprenda. El campo ha evolucionado a lo largo del tiempo, desde técnicas formales clásicas como las vectoriales y probabilísticas hasta las más actuales que emplean el aprendizaje automático profundo (Díaz et al., 2022).

Existen diferentes tipos de Aprendizaje, como son el supervisado, no supervisado, un híbrido entre ambos, el semi-supervisado o el aprendizaje por refuerzo. Estos son clasificados de esta manera según la forma en que se los datos se le presenten al modelo, ya sea etiquetados o no, así como su interacción con el entorno.

4.8.2 Deep Learning

El aprendizaje profundo, Deep Learning en inglés, es un subcampo del aprendizaje automático, utiliza redes neuronales artificiales con múltiples capas (de ahí el término "profundo") que pueden aprender de grandes cantidades de datos. Las redes que son utilizadas o creadas son capaces de modelar y extraer características complejas de los datos, una forma de "aprendizaje", estas múltiples capas de neuronas interpretan la información en secuencia, y cada capa aprende, es decir recibe y procesa datos, de la capa anterior (Jabir & Falih, 2021) Las redes neuronales profundas han ganado gran popularidad porque permiten modelar abstracciones de alto nivel de datos complejos (Schmidhuber, 2015). Sus aplicaciones incluyen reconocimiento de imágenes y voz, traducción automática, juegos de estrategia y simulaciones.

4.8.3 Redes neuronales

Una red neuronal es un programa informático o modelo de aprendizaje automático que se inspira en la estructura del cerebro humano y toma decisiones de forma similar. Consisten en capas de neuronas artificiales interconectadas y el peso de cada conexión se ajusta durante el proceso de entrenamiento. Recibe señales de entrada x_i de n neuronas $x_1, x_2, x_3 \dots x_n$, cada una con su respectivo peso sináptico w_{ji} . La entrada total a la neurona se calcula como una suma ponderada de las entradas. Aquí, el peso de cada sinapsis representa la importancia relativa de la entrada correspondiente:

$$y_j^{(in)} = \sum_{i=1}^n w_{ji} x_i \quad \text{Ecuación (5)}$$

La neurona se activa si la entrada total sobrepasa cierto umbral, para ello se aplica una función de activación (Izaurieta & Saavedra, s/f), en la Figura 9 podemos ver dos ejemplos de funciones de activación.

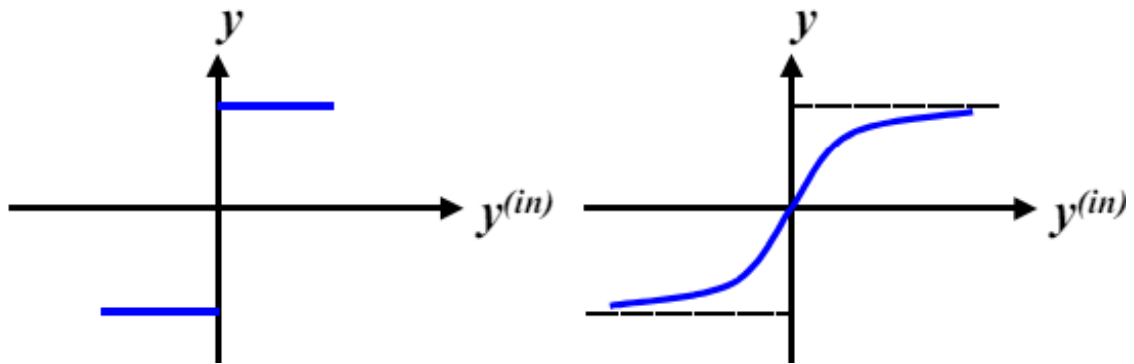


Figura 9. Funciones de activación Escalón y sigmoidea. Recuperado de (Izaurieta & Saavedra, s/f).

Su estructura consta de tres partes importantes: capa de entrada, capas ocultas y capa de salida. Los datos que llegan a la red se les denomina la parte de la capa de entrada, en las capas ocultas se extrae la información y se realizan cálculos a partir de los datos de entrada y finalmente la capa de salida donde se proporciona los resultados, puede ser en forma de clasificación, regresión u otra tarea específica, se llega a esta conclusión mediante la

interpretación y cálculos procesados por las capas ocultas. El diagrama de la Figura 10 representa una red neuronal con cada una de las partes esenciales de una red neuronal.

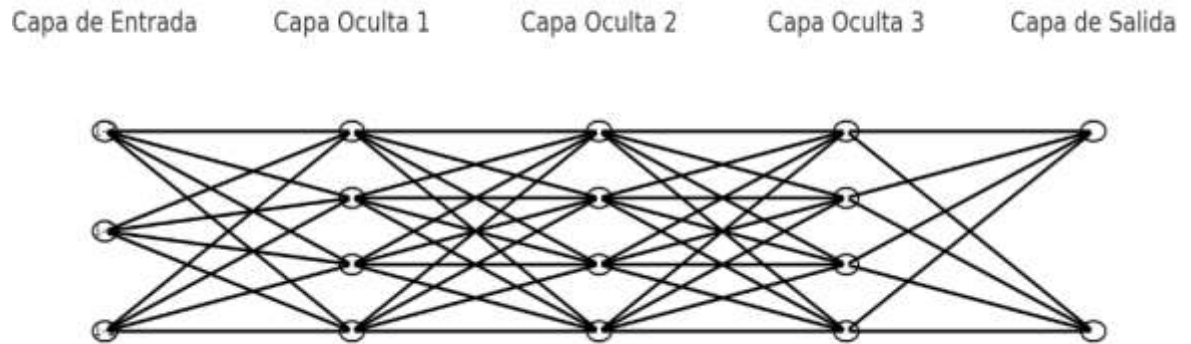


Figura 10. Diagrama de una red neuronal
Cada círculo representa una neurona, y las líneas entre los círculos representan las conexiones (pesos) entre las neuronas de diferentes capas. Las etiquetas en la parte superior indican las diferentes capas de la red.

4.8.4 Redes neuronales convolucionales

Las Redes Neuronales Convolucionales (CNN) son importantes en el campo del aprendizaje profundo, en especial en el análisis de imágenes y aplicaciones de visión por computadora. Estas redes utilizan variadas capas de filtros convolucionales para extraer características de los datos de entrada, típicamente imágenes, que luego se conectan a un perceptrón multicapa para tareas de clasificación (de Santos García et al., 2022).

Su funcionamiento se basa en la aplicación de múltiples capas de filtros convolucionales a una entrada, generalmente imágenes, para extraer características relevantes. Estas capas adaptativas son responsables de identificar características en diferentes partes de la imagen, lo que permite a la red aprender más y más problemas a medida que atraviesa las capas (Curioso & Brunette, 2020; Jiménez-Murillo, 2023).

Las CNN utilizan operaciones como convolución, activación no lineal (como ReLU), agrupación (pooling) y capas totalmente conectadas para transformar y extraer características del proceso. La convolución consiste en deslizar un filtro sobre la imagen de entrada para producir mapas de características que resaltan ciertas características visuales. La activación no lineal introduce no linealidades en la red, permitiendo aprender patrones más complejos. La fusión reduce la dimensionalidad del mapa de características y preserva la información relevante. Finalmente, una capa completamente conectada combina las características extraídas para realizar la clasificación final (Niño-Carmona et al., 2021; Quiroga, s/f).

4.8.4.1 U-net

La U-Net es una de las arquitecturas más influyentes en el procesamiento de imágenes biomédicas y constituye una referencia fundamental en modelos de tipo encoder-decoder. Fue propuesta por Ronneberger, Fischer y Brox en 2015 durante la conferencia MICCAI, donde obtuvo resultados sobresalientes en el desafío ISBI de segmentación celular. Su nombre proviene de su estructura en forma de “U”, la cual puede observarse en la Figura 11, donde se muestra la arquitectura original presentada por los autores (Ronneberger et al., 2015).

La parte descendente de la red (encoder) se compone de bloques sucesivos de convoluciones 3×3 con activaciones ReLU, seguidos de operaciones de max pooling 2×2 . Este proceso reduce progresivamente la resolución espacial mientras incrementa la profundidad de los mapas de características, permitiendo capturar representaciones semánticas de alto nivel.

En la parte ascendente (decoder), la red recupera la resolución original mediante capas de upsampling o convoluciones transpuestas 2×2 . El elemento distintivo de la U-Net es la incorporación de conexiones de salto (skip connections) entre el encoder y el decoder en cada nivel, permitiendo concatenar características de alta resolución con las representaciones profundas reconstruidas. Estas conexiones son esenciales para preservar detalles espaciales finos que se pierden usualmente durante el downsampling, mejorando así la precisión en la localización y definición de bordes.

Aunque originalmente se diseñó para segmentación, su flexibilidad estructural ha permitido adaptarla a un amplio conjunto de tareas, incluyendo reconstrucción, eliminación de ruido, estimación de mapas continuos y superresolución. Posteriores variantes han extendido sus capacidades incorporando normalización por lotes, funciones de activación alternativas, convoluciones dilatadas o bloques residuales, así como extensiones a 3D para trabajar con volúmenes médicos (Çiçek et al., 2016; Isensee et al., 2021).

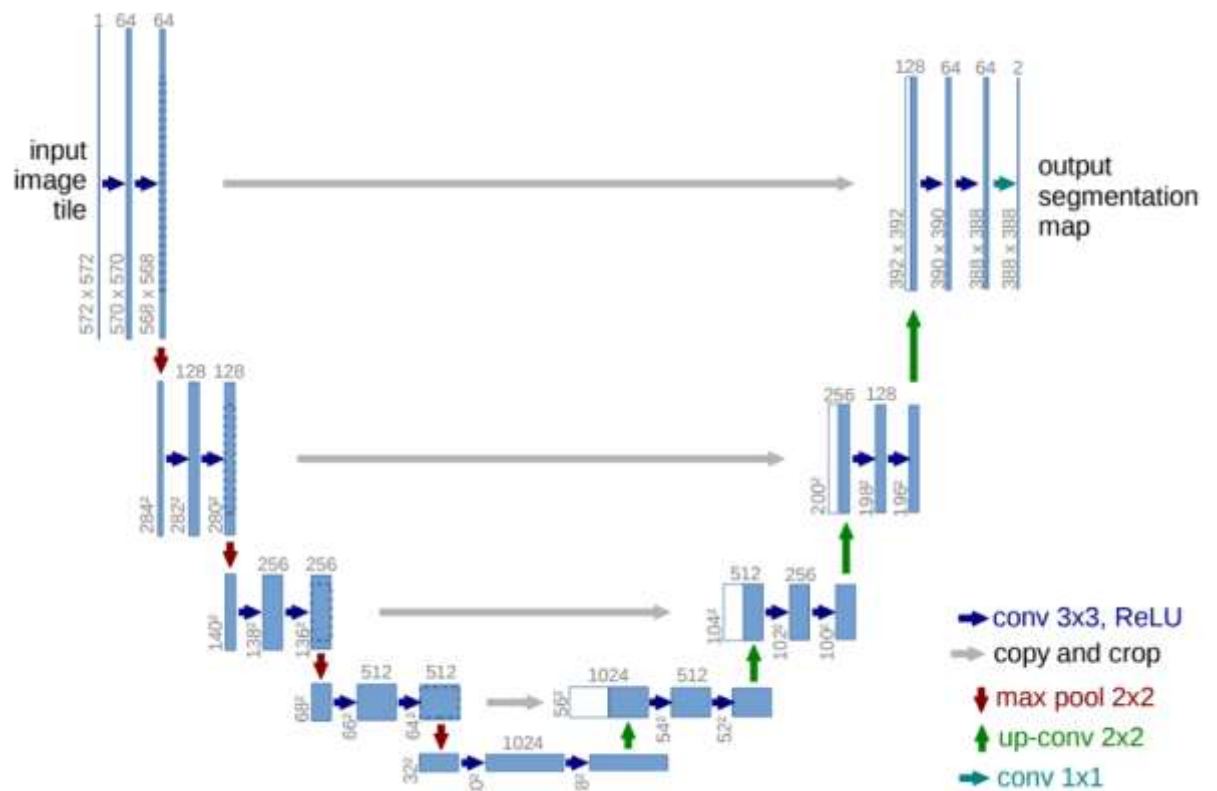


Figura 11. Arquitectura U-Net.
Recuperado de (Ronneberger et al., 2015).

Gracias a su capacidad para combinar información global y local, la U-Net se ha convertido en un modelo base ampliamente utilizado en visión computacional, especialmente en escenarios donde es crucial mantener la coherencia espacial de la salida, como es el caso de la reconstrucción de mapas de fase en técnicas interferométricas y de proyección de franjas.

U-net con bloques LST-SQDM

La U-Net con módulo Spatial Quad-Directional LSTM (SQD-LSTM) es una variante arquitectónica de la U-Net clásica, diseñada para integrar procesamiento convolucional jerárquico con un mecanismo de memoria espacial en cuatro direcciones. Esta arquitectura fue propuesta por Perera y De Silva (2020) como una extensión del esquema encoder-decoder de la U-Net, y se muestra en la Figura 12.

Estructuralmente, mantiene el diseño simétrico de codificador-decodificador basado en convoluciones 3×3 , normalización por lotes y activación ReLU, junto con skip connections que conectan los niveles correspondientes del encoder y decoder. El encoder reduce gradualmente la resolución mediante *average pooling*, mientras aumenta la profundidad de los mapas de características; el decoder invierte este proceso mediante convoluciones transpuestas 3×3 con *stride 2* para recuperar la resolución original.

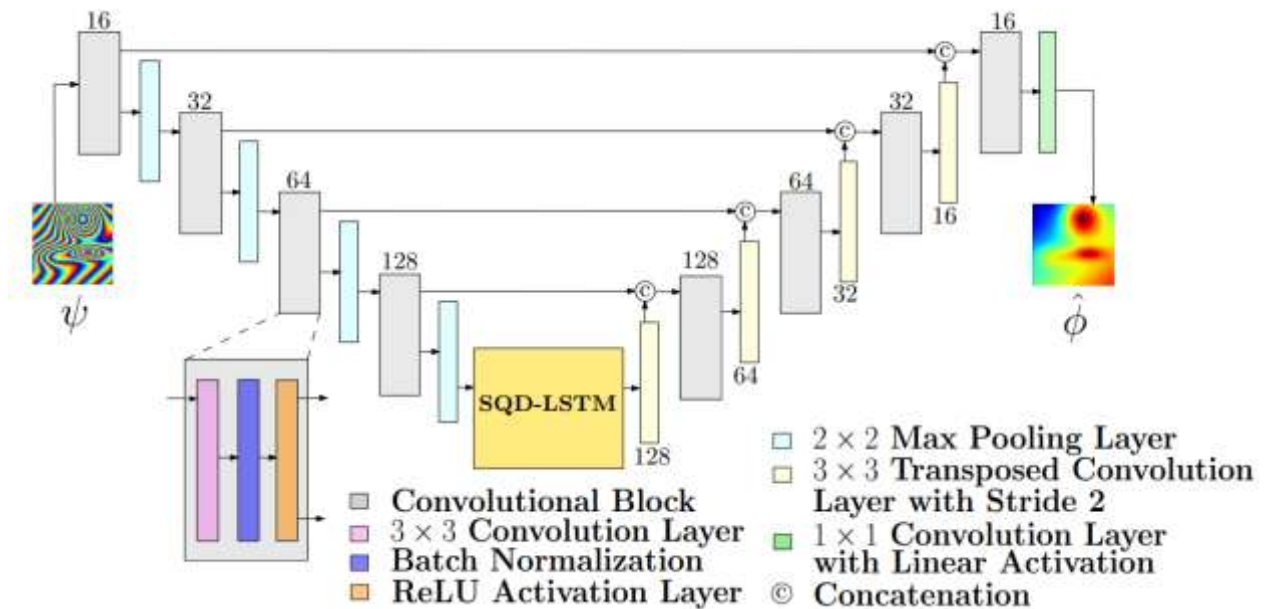


Figura 12. Arquitectura U-Net + LSTM.
Recuperado de (Perera & De Silva, 2020).

El elemento distintivo de esta arquitectura es el módulo SQD-LSTM, ubicado en el cuello de botella entre el encoder y el decoder. Este módulo transforma el mapa de características profundo en cuatro secuencias espaciales: izquierda→derecha, derecha→izquierda, arriba→abajo y abajo→arriba. Cada secuencia es procesada por un LSTM independiente, lo que permite capturar dependencias espaciales de largo alcance en las cuatro direcciones. Las salidas de los LSTM se reorganizan y concatenan, y posteriormente se refinan mediante capas convolucionales adicionales, generando un mapa de características enriquecido que se transmite al decoder (Perera & De Silva, 2020).

Finalmente, la arquitectura produce su salida mediante una convolución 1×1 , típica de las variantes de U-Net empleadas para regresión. Esta combinación del flujo jerárquico convolucional con un módulo de memoria direccional cuatri-espacial permite que la red preserve tanto información local como dependencias espaciales amplias, superando las limitaciones del procesamiento puramente convolucional.

4.8.4.2 *Res-net*

Las Redes Residuales (ResNet) fueron introducidas por (He et al., 2016) como una solución al problema de degradación en redes profundas, donde aumentar la profundidad no necesariamente mejora el desempeño debido a dificultades de optimización y desaparición del gradiente. La idea central consiste en aprender no directamente un mapeo $H(x)$, sino una función residual $F(x) = H(x) - x$, de manera que la salida del bloque adopta la forma:

$$y = F(x) + x$$

donde x es la entrada del bloque y $F(x)$ es la salida de una secuencia de capas convolucionales. Esta suma se implementa mediante una conexión de atajo (*shortcut* o *skip connection*) que copia la entrada del bloque y la suma a la salida de las convoluciones. En la práctica, un bloque residual típico está formado por dos convoluciones 3×3 (a menudo seguidas de normalización por lotes y activación ReLU) más la conexión de atajo, que puede ser identidad o una convolución 1×1 cuando cambia el número de canales.

Este esquema residual facilita la propagación del gradiente a través de muchas capas, permite entrenar redes mucho más profundas y favorece la reutilización de características, lo que ha convertido a ResNet en una arquitectura base en tareas de clasificación de imágenes, segmentación semántica, detección, restauración y reducción de ruido. La incorporación de **bloques residuales** dentro de arquitecturas encoder–decoder, como U-Net, se ha explorado ampliamente en la literatura (por ejemplo,(He et al., 2016)) para mejorar la capacidad de representación manteniendo una estructura multiescala.

U-net Res-net

La arquitectura U-Net–ResNet utilizada en esta tesis corresponde a una variante híbrida de U-Net en la que los bloques convolucionales tradicionales son reemplazados por bloques residuales tipo ResNet. Este diseño combina la capacidad multiescala del esquema encoder–decoder de U-Net con la estabilidad de entrenamiento y el flujo de gradiente mejorado que aportan las conexiones de atajo de ResNet. La estructura completa empleada en esta tesis se ilustra en la Figura 13, donde se muestra el recorrido de la imagen de fase ruidosa hasta obtener la fase restaurada.

En el encoder, la red aplica sucesivamente bloques residuales con 64, 128 y 256 filtros, cada uno seguido de una operación de *max pooling* 2×2 que reduce la resolución espacial mientras aumenta la profundidad de los mapas de características. El cuello de botella está conformado por un bloque residual con 512 filtros, encargado de capturar las representaciones de mayor nivel y globalidad.

En el decoder, la resolución espacial se recupera mediante convoluciones transpuestas (*UpConv* 2×2). Tras cada operación de *upsampling*, las características se concatenan con las del nivel correspondiente del encoder mediante *skip connections* largas, preservando detalles locales esenciales para una reconstrucción precisa. Cada nivel del decoder vuelve a procesar estas características a través de un bloque residual, lo que permite refinar progresivamente la información recuperada.

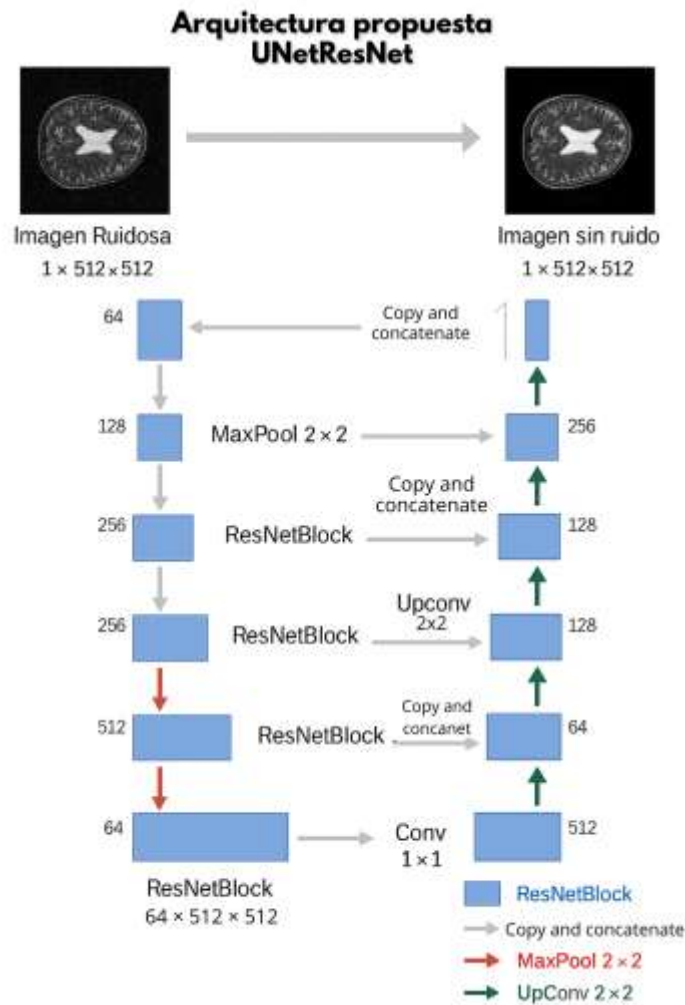


Figura 13. Arquitectura UNetResNet

Finalmente, una convolución 1×1 genera el mapa de salida, correspondiente al valor de fase desenvuelta o imagen sin ruido, dependiendo de la tarea. La combinación de skip connections largas (U-Net) y skip connections cortas internas (ResNet) dota al modelo de una capacidad expresiva superior frente a arquitecturas puramente convolucionales, siendo especialmente efectiva para problemas de restauración, reducción de ruido y reconstrucción de fase.

Esta arquitectura, basada en el diagrama mostrado en la Figura 13, fue implementada específicamente en esta tesis para integrar los beneficios del aprendizaje profundo multiescala con las propiedades de estabilidad y aprendizaje residual características de ResNet.

4.8.4.3 Multiresolution CNN

La red multiresolución propuesta por (Y. Sun et al., 2018) es una arquitectura convolucional diseñada para procesar imágenes a través de múltiples escalas de forma paralela. Aunque no deriva de U-Net ni de ResNet, comparte el principio general de las CNN de extraer características mediante convoluciones, pero incorpora un esquema piramidal completamente entrenable.

La idea central del modelo es que una imagen puede contener patrones complejos distribuidos a diferentes niveles de detalle; por ello, la red descompone la entrada en varias resoluciones mediante convoluciones con stride 2, generando una pirámide de características no lineal. Cada nivel de esta pirámide se procesa mediante una secuencia de capas convolucionales que se encargan de extraer y transformar la información correspondiente a esa banda de frecuencia.

Posteriormente, las ramas de menor resolución se amplían nuevamente utilizando capas de deconvolución hasta recuperar el tamaño original. Finalmente, cada rama genera un mapa de características de tres canales, y la salida final de la red se obtiene a partir de la suma de todos estos mapas. De esta manera, la arquitectura integra la información procesada en distintas escalas sin emplear skip connections, lo que la diferencia de modelos como U-Net.

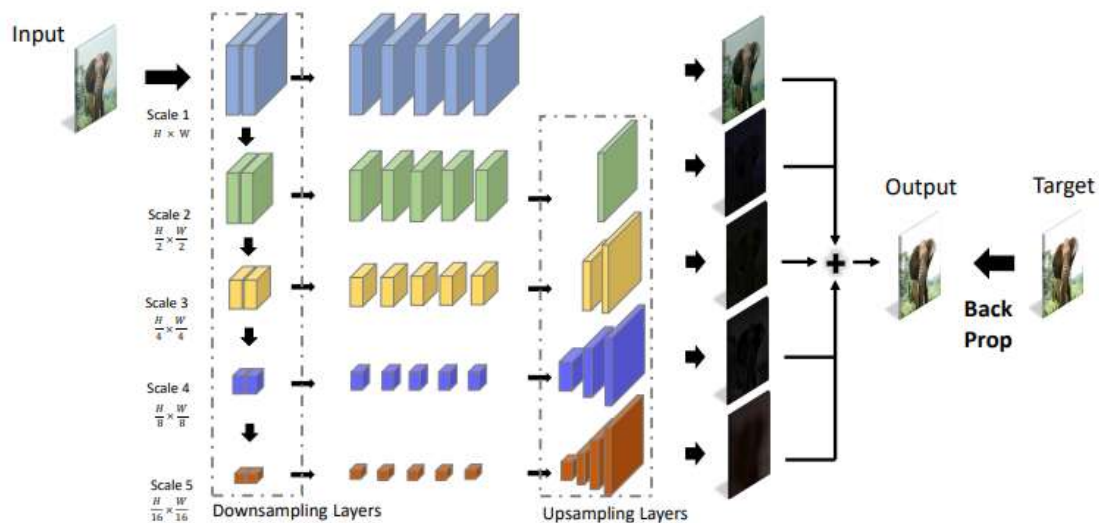


Figura 14. Arquitectura Multiresolution CNN. Arquitectura propuesta y recuperado de (Y. Sun et al., 2018).

La Figura 14 ilustra el funcionamiento multiescala de la red, iniciando con la generación de una pirámide de características mediante convoluciones con stride 2 (bloque de downsampling). Cada nivel de resolución se procesa de manera independiente a través de una secuencia de capas convolucionales que extraen información específica de esa banda de frecuencia. Posteriormente, los niveles inferiores se restauran mediante capas de deconvolución (upsampling) que recuperan progresivamente la resolución original. Cada rama produce un mapa RGB que representa su contribución a la eliminación del artefacto en su escala correspondiente, y la salida final se obtiene mediante la suma de todas las ramas. Este diseño permite integrar información estructural distribuida a diferentes escalas sin emplear skip connections, lo cual lo distingue de arquitecturas como U-Net.

Aunque esta red fue desarrollada inicialmente para la restauración de artefactos de muaré en fotografías de pantallas digitales, su estructura es aplicable a tareas donde los patrones no deseados aparecen distribuidos en diversas bandas de frecuencia, como ocurre en problemas de reducción de ruido complejo o en la corrección de distorsiones con componentes multiescala.

4.9 Métricas

Aunque este trabajo se enmarca en el ámbito de la inteligencia artificial, su objetivo principal es la reconstrucción tridimensional mediante desdoblamiento de fase, por lo que resulta necesario distinguir entre dos tipos de métricas. Por un lado, se emplean métricas propias de IA que permiten evaluar el aprendizaje de los modelos durante el entrenamiento (como la función de pérdida y su convergencia). Por otro lado, la calidad final de la reconstrucción debe cuantificarse mediante métricas de imágenes, ya que estas describen de manera objetiva el desempeño del modelo sobre los mapas de fase y altura reconstruidos. Esta separación permite evaluar tanto el comportamiento interno del modelo como la fidelidad visual y cuantitativa del resultado físico que se desea recuperar.

4.9.1 Métricas IA

Para poder evaluar el desempeño de los modelos de Inteligencia existen varias métricas, que miden el rendimiento del modelo y qué tan efectivamente realizó la tarea correspondiente, entre ellas las principales son:

- Precisión: Porcentaje de predicciones correctas respecto al número total de predicciones.
- Recall: El resultado de verdaderos positivos sobre la suma de verdaderos positivos y falsos negativos.
- F1 Score: Media armónica de precisión y recall.
- ROC-AUC: Área bajo la curva ROC, mide el rendimiento de un modelo binario.
- Matriz de confusión: Es una forma de condensar la información en forma de tabla, donde se puede representar los resultados de los resultados de predicción. Falsos positivos, falsos negativos, verdaderos negativos y verdaderos positivos. Como se puede observar en la Figura 15. Diagrama de matriz de confusión.
- Loss: La loss (o función de pérdida) cuantifica el error cometido por el modelo durante el proceso de entrenamiento. Representa qué tan lejos se encuentra la predicción del modelo respecto al valor real. Su objetivo es ser minimizada, ya que una pérdida menor indica un mejor ajuste del modelo a los datos. La loss se actualiza en cada iteración utilizando un método de optimización (como Adam, SGD o RMSProp), y es el principal indicador del aprendizaje interno del modelo durante el entrenamiento.
- Val_loss: La val_loss (validation loss) corresponde al error calculado sobre el conjunto de validación, el cual contiene datos no vistos durante el entrenamiento. Esta métrica permite evaluar la capacidad de generalización del modelo. Valores de val_loss cercanos a la loss de entrenamiento indican un modelo bien ajustado; mientras que diferencias significativas entre ambos valores pueden revelar sobreajuste (overfitting) subajuste (underfitting).

		Predicción	
		Positivos	Negativos
observación	Positivos	Verdaderos positivos (VP)	Falsos negativos (FN)
	Negativos	Falsos positivos (FP)	Verdaderos negativos (VN)

Figura 15. Diagrama de matriz de confusión
Ejemplo de los datos que incluyen matriz de confusión.

4.9.2 Métricas imágenes

Para evaluar objetivamente la calidad de los mapas de fase desenvueltos y comparar el rendimiento entre distintos métodos, se emplearon tres métricas cuantitativas ampliamente utilizadas en procesamiento de imágenes: IMMSE, PSNR y SSIM. Estas métricas permiten analizar la similitud entre el mapa de fase reconstruido y el ground truth, considerando tanto diferencias puntuales como variaciones estructurales percibidas. Todas ellas han sido reportadas como herramientas estándar en visión por computadora y restauración de imágenes.

4.9.2.1 Error Cuadrático Medio en Imágenes (IMMSE)

El IMMSE (*Image Mean Square Error*) cuantifica la diferencia promedio entre la imagen original $I(x, y)$ y la imagen procesada $K(x, y)$. En el contexto de esta tesis, se utiliza para medir el error entre el mapa de fase real (ground truth) y el mapa desenvuelto estimado.

$$IMMSE = \frac{1}{mn} \sum_{x=1}^m \sum_{y=1}^n [I(x, y) - K(x, y)]^2$$

donde:

- $I(x, y)$: imagen original,
- $K(x, y)$: imagen reconstruida,

- m : número de renglones,
- n : número de columnas.

Valores menores indican una mayor similitud entre ambas imágenes. Esta métrica es especialmente útil para cuantificar errores locales que pueden ser críticos en aplicaciones de reconstrucción de fase (Köl, 2016).

4.9.2.2 *Relación Señal–Ruido de Pico (PSNR)*

El PSNR es una métrica derivada del IMMSE que evalúa la relación entre la señal máxima posible en la imagen y el ruido presente en la reconstrucción. Se expresa como:

$$PSNR = 10 \cdot \log_{10} \left(\frac{Max^2}{IMMSE} \right)$$

donde Max es el valor máximo posible del píxel.

Un PSNR alto implica que la imagen reconstruida es más cercana a la original. Valores entre 30 y 50 dB se consideran típicos de imágenes de alta calidad 19. En esta tesis, el PSNR se emplea para evaluar qué tan bien cada método preserva la información de fase bajo distintos niveles de ruido (Köl, 2016).

4.9.2.3 *Índice de Similitud Estructural (SSIM)*

El SSIM cuantifica la similitud estructural entre dos imágenes, considerando cambios en luminancia, contraste y estructura:

$$SSIM(x, y) = l(x, y)^\alpha c(x, y)^\beta s(x, y)^\gamma$$

donde:

- $l(x, y)$: luminancia,
- $c(x, y)$: contraste,
- $s(x, y)$: similitud estructural.

El SSIM fue diseñado como una métrica perceptual, ya que modela la degradación visual en términos cercanos a la percepción humana. Sus valores oscilan entre -1 (sin similitud) y 1 (imágenes idénticas).

En el contexto de mapas de fase, SSIM es especialmente relevante porque evalúa la preservación de estructuras, bordes y contornos, elementos esenciales durante el desdoblamiento de fase (Z. Wang et al., 2004).

4.10 Estado del arte

En esta sección se presenta el estado del arte relativo al desdoblamiento de fase y la reconstrucción tridimensional mediante técnicas de proyección de franjas. Dado que el trabajo propuesto combina elementos clásicos de profilometría con modelos de aprendizaje profundo, el análisis se organiza en dos bloques complementarios. En primer lugar, en

Tabla 1 se resumen los métodos tradicionales, donde se incluyen algoritmos guiados por confiabilidad y enfoques locales/globales aplicados a FTP e interferometría, así como un estudio comparativo reciente que integra variantes dRG, dWC y dDN.

En segundo lugar, la

Tabla 2 sintetiza los aportes basados en redes neuronales profundas, incluyendo arquitecturas tipo U-Net residual, modelos multi-resolución, enfoques de unwrapping en un solo paso y marcos unificados temporales, así como trabajos recientes que emplean variantes híbridas (U-Net + Transformers, GANs, etc.). En esta tabla se destacan las arquitecturas empleadas, el tamaño y naturaleza de los conjuntos de datos, las métricas de desempeño y los algoritmos tradicionales con los que se realiza la comparación. Esta organización permite situar claramente la contribución de este trabajo dentro de la evolución del campo, desde los métodos deterministas clásicos hasta las soluciones actuales basadas en aprendizaje profundo.

4.10.1 Métodos tradicionales

Diversos métodos clásicos han sido desarrollados para el desdoblamiento de fase en sistemas de perfilometría, interferometría o patrones de franjas. Su y Chen (2004) propusieron un algoritmo guiado por confiabilidad que utiliza parámetros como la modulación, el SNR y la frecuencia espacial para determinar el camino de desdoblamiento, evitando la propagación de errores en zonas de sombra o discontinuidades. El método fue evaluado sobre patrones de perfilometría, FTP, interferometría y fotoelasticidad, utilizando objetos reales como cráneos, mandíbulas y piezas industriales, con evaluación cualitativa mediante reconstrucción visual y progresión de regiones desenvueltas.

Posteriormente, Sotomayor-Olmedo et al. (2010) analizaron un esquema de desdoblamiento local (residue balancing) y uno global (mínimo-norma) aplicado a un método FTP modificado. Sus experimentos incluyeron formas sintéticas (mariposa, mano, gato) y objetos reales con franjas proyectadas.

Finalmente, Wang et al. (2022) presentaron una comparación moderna de tres enfoques de desdoblamiento: dRG (regresión directa), dWC (conteo de envoltura) y dDN (denoising de CAF) basados en modelos profundos como Res-UNet y GANs. Trabajaron con datos simulados (RME, GFS, ZPS) y 421 objetos reales (llamas, hoyos, ranuras). Los errores RMSE obtenidos fueron: 0.0005 para dRG ideal, 0.0285 para dRG con ruido y 0.0219 para dWC en escenarios discontinuos. La evaluación incluyó RMSE, PFS y PIP en cinco condiciones: ideal, ruidoso, discontinuo, aliasing y mixto.

Artículo (Cita IEEE)	Modelos / Algoritmos	Datos utilizados	Resultados clave	Evaluación
X. Su y W. Chen, 2004Opt. Lasers Eng., vol. 42, pp. 245–261.	Algoritmo de desdoblamiento guiado por confiabilidad (modulación, frecuencia espacial, SNR, etc.).	Fringe patterns en: perfilometría de fase, FTP, interferometría y fotoelasticidad. Objetos reales como cráneo, mandíbula, piezas industriales.	Camino de desdoblamiento desde regiones de alta confiabilidad. Evita propagación de errores en áreas de sombra o discontinuidades.	Comparación visual de reconstrucción, progresión de áreas desenvueltas, precisión cualitativa en zonas complejas.
A. Sotomayor-Olmedo et al., 2010Proc. IEEE	- Local (residue balancing)- Global (mínimo-norma).Aplicado en un método FTP modificado.	Simulaciones: mariposa, gato, mano (formas sintéticas).- Objeto real con patrón de franjas proyectado.	Error global menor con método global:- Global: 3.4–3.9%- Local: 4.5–4.7%	Error porcentual calculado comparando las reconstrucciones con el ground truth en objetos sintéticos. Comparación visual de malla reconstruida.
Y. Wang et al., 2022Adv. Photonics Nexus, vol. 1, no. 1	- dRG (regresión directa)- dWC (conteo de envoltura)- dDN (denoising de CAF) Todos basados en Deep Learning con Res-UNet, GAN, etc.	Simulados: RME, GFS, ZPS (20k train / 2k test)- Reales: 421 objetos físicos (llamas, hoyos, ranuras).	RMSE mínimo:- dRG ideal: 0.0005- dRG ruido: 0.0285- dWC discontinuo: 0.0219- Mejor generalización con RME datasets.	RMSE, PFS (proporción de muestras fallidas), PIP (proporción de píxeles incorrectos). Evaluación en cinco escenarios: ideal, ruido, discontinuo, aliasing, mixto.

Tabla 1. Ejemplos de métodos tradicionales de fase.

4.10.1 Estado del arte basado en Deep Learning para desdoblamiento de fase y reconstrucción 3D

El uso de redes neuronales ha introducido mejoras significativas en el manejo de ruido, aliasing y complejidad geométrica.

Uno de los trabajos más relevantes es el modelo Anti-Noise-Designed Residual Phase Unwrapping Network, una arquitectura Residual U-Net diseñada para interferometría speckle. El método incorpora bloques residuales que mejoran la gestión del ruido y logra valores de SSIM superiores a 0.98, con tiempos de desdoblamiento cercanos a 1 segundo para imágenes de distintos tamaños. El entrenamiento utilizó 40,000 imágenes simuladas (256×256 px) con ruido gaussiano, sal-pimienta y multiplicativo. La comparación se realizó contra métodos tradicionales de camino espiral y confiabilidad.

El trabajo One-step robust deep learning phase unwrapping (DLPU) emplea una arquitectura U-Net combinada con bloques ResNet. Alcanzó SSIM > 0.98 y valores inferiores de RMSE bajo ruido y aliasing, utilizando 37,500 imágenes de fase con ruido agregado. Sus resultados superaron a mínimos cuadrados (LS) y métodos guiados por calidad (QG).

El modelo Unifying temporal phase unwrapping framework utiliza MultiResUNetFOA-Net, una red con módulos MultiRes, conexiones residuales ResPath y variantes especializadas para multifrecuencia, multi-longitud de onda y número-teórico. El método redujo el error de desenrollado (ruido 3σ) de 9.26% a 4.22% y alcanzó velocidades de 12 fps, empleando patrones proyectados reales (1–64 frecuencias) capturados con cámara industrial y proyector DLP.

Otros aportes recientes incluyen:

- Wang et al. (2022): revisión comparativa de modelos Res-UNet unificados para dRG, dWC y dDN, probados en RMSE y datos experimentales reales con ruido.
- Jingbei Wang et al. (2022): CNNs y redes densas para Near-Infrared FPP; evaluado con 800 imágenes reales y métricas MAE/RMS.

- Nguyen et al. (2023): SIDO (Single-Input Dual-Output) basado en UNet para reconstrucción 3D; 2048 muestras reales; métricas rE, t90, t125.
- Bujgosqui et al. (2024): arquitectura híbrida CNN-Transformer para holografía digital; SSIM de 0.9978.
- Zheng et al. (2024): SFNet con Transformer jerárquico y decodificador All-MLP; RMSE 0.079 y PSNR 157.217 dB.
- Jiang et al. (2025): GAN-PhaseNet para dispersión severa y objetos translúcidos; MAE y SSIM superiores a métodos tradicionales.

Año / Autor	Método / Arquitectura	Enfoque	Dataset y tamaño	Resultados (métricas)	Características / Objetos
<i>Anti-Noise Residual PUNet</i>	Residual U-Net con bloque residual	dRG	40,000 imágenes simuladas (256×256) con ruido gaussiano, sal-pimienta y multiplicativo	SSIM > 0.98; tiempo ≈ 1 s por imagen	Interferometría speckle digital; comparación con camino espiral y confiabilidad
<i>One-step robust DLPU</i>	U-Net + bloques ResNet	dWC / dRG	37,500 imágenes con ruido agregado	SSIM > 0.98; RMSE menor que LS y QG	Fase con ruido y aliasing; condiciones simuladas de interferometría
<i>Unifying Temporal PU (MultiResUNetFOA-Net)</i>	MultiRes + ResPath + DL-MF / DL-MW / DL-NT	dWC temporal	6 frecuencias proyectadas (1–64) sobre superficies reales	Error reducido de 9.26% → 4.22%; 12 fps	Objetos reales; cámara Basler + proyector DLP
2022 – Xiaojing Wang et al.	Res-UNet comparativo (dRG, dWC, dDN)	Comparativo general	20,000 RME + 2,000 ZPS + 421 RDR reales	RMSEm 0.0285; RMSsed 0.1148	Fases simuladas y reales (ruinas, pozos)
2022 – Jingbei	CNN1 + Residual Block	dDN	800 imágenes reales	MAE ideal 0.0278–	Near-Infrared FPP; piezas cerámicas y

Wang et al.	+ CNN2 (4-path CNN)		NIR (70/25/5 split)	0.0332; RMS ~80 μ m	modelos 3D
2023 – Nguyen et al.	SIDO (Single-Input Dual-Output UNet)	dWC	2048 muestras reales (WBUST dataset)	rE: 0.004; RMS: 0.385; t125: 9.5%	Superficies cerámicas industriales e
2024 – Bujgosqui et al.	Hybrid UNET + Vision Transformer	dRG	Sintético + DHM experimental	SSIM 0.9978; RMS 0.1391 rad	Holografía digital; células epiteliales
2024 – Zheng Zhang et al.	SFNet (Transformer jerárquico + All-MLP)	dWC	38,000 train + 2,000 test	RMSE: 0.079; PSNR: 157.217; SSIM: 0.9994	Objetos reales (contorno facial)
2025 – Jiang et al.	GAN-PhaseNet (GAN + SFNet)	dWC	1,200 train + 928 test	MAE 0.1314 mm; RMS: 0.2369; PSNR: 13.5463	Objetos translúcidos y cerámicas dispersivas

Tabla 2. Estado del arte en DL para desdoblamiento de fase y reconstrucción 3D.

V. Metodología

El presente capítulo describe de manera integral y coherente todas las etapas que conforman la metodología empleada en esta investigación, desde la construcción y curación de los datos hasta la implementación, entrenamiento y evaluación de los algoritmos utilizados para el desdoblamiento de fase. En primer lugar, se detalla el proceso de generación de las bases de datos, tanto sintéticas como reales, las cuales constituyen el insumo fundamental para el análisis comparativo entre métodos tradicionales y modelos basados en aprendizaje profundo.

Para los objetos sintéticos se desarrolló un pipeline en Blender y Python capaz de producir proyecciones de franjas, máscaras binarias, mapas de altura y fases envueltas mediante los métodos de tres pasos y de Transformada de Fourier (FT). Este pipeline permitió controlar parámetros geométricos, condiciones de iluminación, ruido aditivo y variaciones en la orientación del proyector, garantizando conjuntos suficientemente diversos para el entrenamiento supervisado. En paralelo, se adquirieron objetos reales bajo el mismo esquema interferométrico experimental, obteniendo mapas low-wrap y multi-wrap sujetos a fenómenos inherentes al sistema físico como ruido estructurado, sombras parciales, superficies no lambertianas y fluctuaciones en la intensidad proyectada.

Posteriormente, se establecieron tres métricas cuantitativas que permiten evaluar rigurosamente la calidad de la fase desenvuelta y comparar el desempeño entre métodos, arquitecturas y configuraciones. Estas métricas, MSE, PSNR y SSIM, constituyen la base objetiva del análisis posterior. De igual manera, se implementaron diversos métodos tradicionales de desdoblamiento espacial, entre ellos los algoritmos de Itoh, Ghiglia-Romero, Zhao, Constantini, Herraéz, PUMA, PEARLS y Martínez-Carranza, que funcionaron como línea base y referencia comparativa frente a los modelos entrenados.

La metodología incluye también el diseño, implementación y entrenamiento de cuatro arquitecturas de redes neuronales profundas: una U-Net estándar, una U-Net con bloques residuales, una arquitectura multirresolución inspirada en pirámides espaciales y una CNN con módulo SQD-LSTM para el modelado secuencial del gradiente de fase. Cada modelo se

entrenó empleando una función de pérdida compuesta (MSE + L1) y bajo distintos esquemas, resoluciones y particiones del dataset, incorporando estrategias de fine-tuning, aumento de datos y validación cruzada con el objetivo de mejorar la generalización hacia el dominio real.

Finalmente, se describe la estructura experimental que organizó todo el flujo de trabajo: estandarización y normalización de datos, protocolos de entrenamiento reproducibles, registro sistemático de configuraciones, control de versiones y criterios homogéneos para la evaluación comparativa. Esta metodología proporciona el soporte técnico necesario para interpretar los resultados presentados en el capítulo siguiente y asegura la trazabilidad completa del proceso.

5.1 Generación de bases de datos

Esta sección se divide en dos partes principales: generación de imágenes sintéticas y adquisición de imágenes de objetos reales, detallando para cada una el procedimiento experimental, los parámetros controlados y los pasos de preprocesamiento requeridos.

5.1.1 Imágenes sintéticas

La base de datos sintética se generó mediante un sistema automatizado desarrollado en Blender (versión 4.3.2) y Python, diseñado para simular un sistema real de proyección de franjas sinusoidales. La generación de imágenes se realizó a partir de modelos tridimensionales obtenidos de la plataforma TurboSquid, seleccionados bajo criterios de variabilidad geométrica y continuidad superficial. Se descartaron modelos con cavidades internas, debido a que el material tipo spline empleado para renderizar el ground truth se dibuja desde el interior de la geometría y requiere superficies continuas para representar correctamente la altura. En la

Figura 18 se muestran ejemplos representativos de las geometrías 3D que conforman la base de datos de objetos sintéticos.

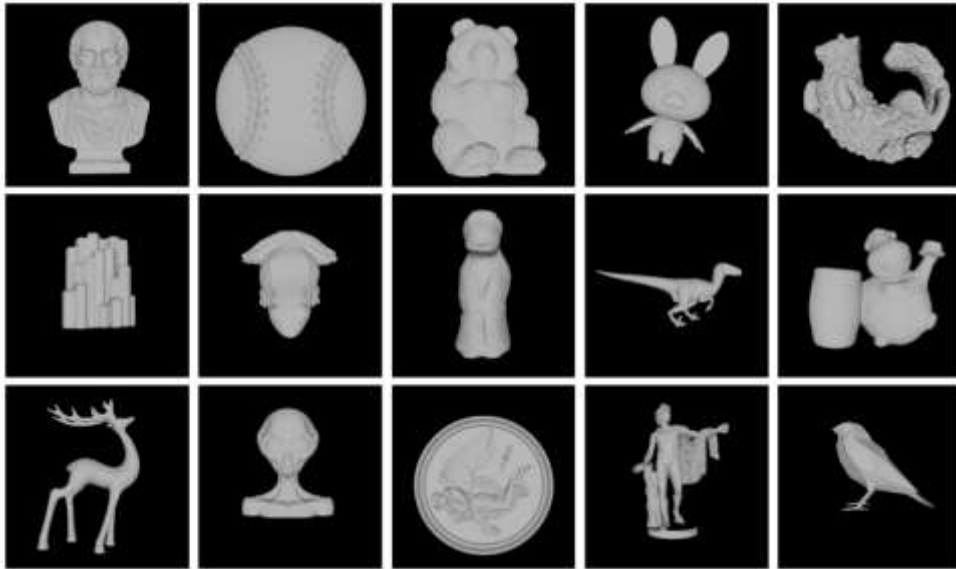


Figura 16. Conjunto representativo de las geometrías 3D. Ejemplos de objetos 3D sintéticos utilizados en el dataset, diseñados para cubrir diferentes perfiles de altura, discontinuidades y topologías relevantes para el desdoblamiento de fase.

La automatización se implementó mediante un script en Python integrado en Blender, el cual ejecuta la secuencia completa de renderización por objeto. El flujo general del proceso se resume en el diagrama de la Figura 17.

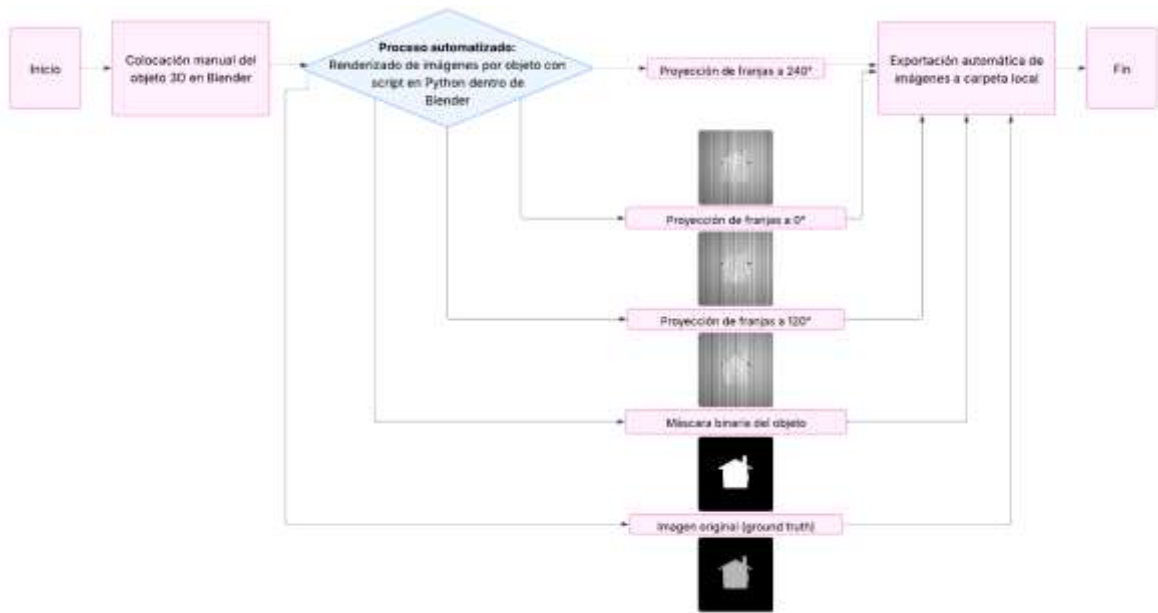


Figura 17. Diagrama de flujo del sistema automatizado.
El diagrama representa los pasos para de la generación de imágenes sintéticas.

Para cada objeto se genera un conjunto de siete imágenes base,

Figura 18:

- i. Proyecciones de franjas sinusoidales con desviaciones de fase de 240° , 0° y 120° .
- ii. Máscara binaria del objeto (mask).
- iii. Imagen original renderizada sin franjas (org).
- iv. Mapas de ground truth (GT-SP y GT-CA), obtenidos asignando dos materiales tipo spline distintos para representar la distribución de alturas superficiales.

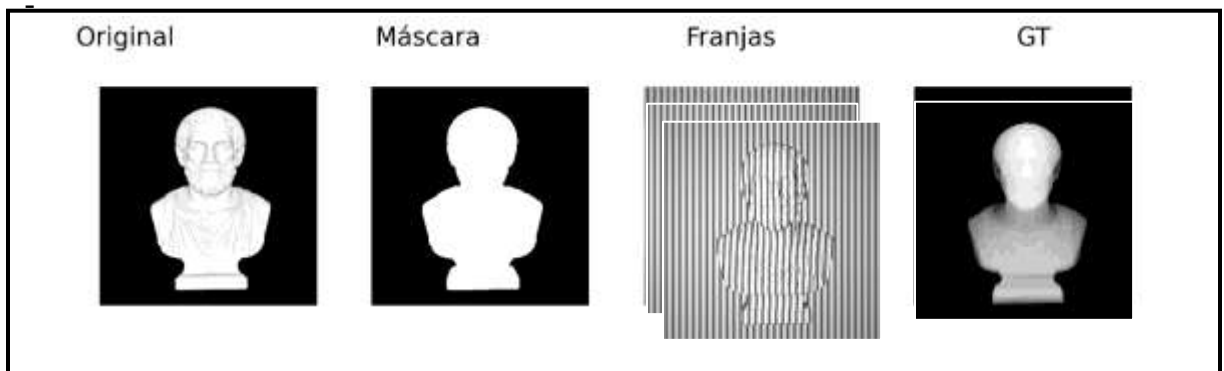


Figura 18. Conjunto de imágenes producidas Blender–Python.

En orden de izquierda a derecha el pipeline genera: Imagen original sin franjas (*org*), máscara binaria (*mask*), proyección sinusoidal correspondiente cada fase de 0° – 120° – 240° (*p1*, *p2* y *p3*), y mapas de ground truth (*GT*) *SP* y *CA* utilizado como referencia para la reconstrucción.

Las imágenes se generaron en resolución 512×512 , escala de grises a 16 bits, y sin aplicar ningún tipo de normalización ni adición de ruido sintético. Cada objeto se rotó 9 veces en incrementos de 40° , con el fin de incrementar la variabilidad geométrica y angular del conjunto. El tiempo de renderización por objeto osciló entre 5 y 40 minutos, dependiendo de la complejidad de la escena y del modelo 3D.

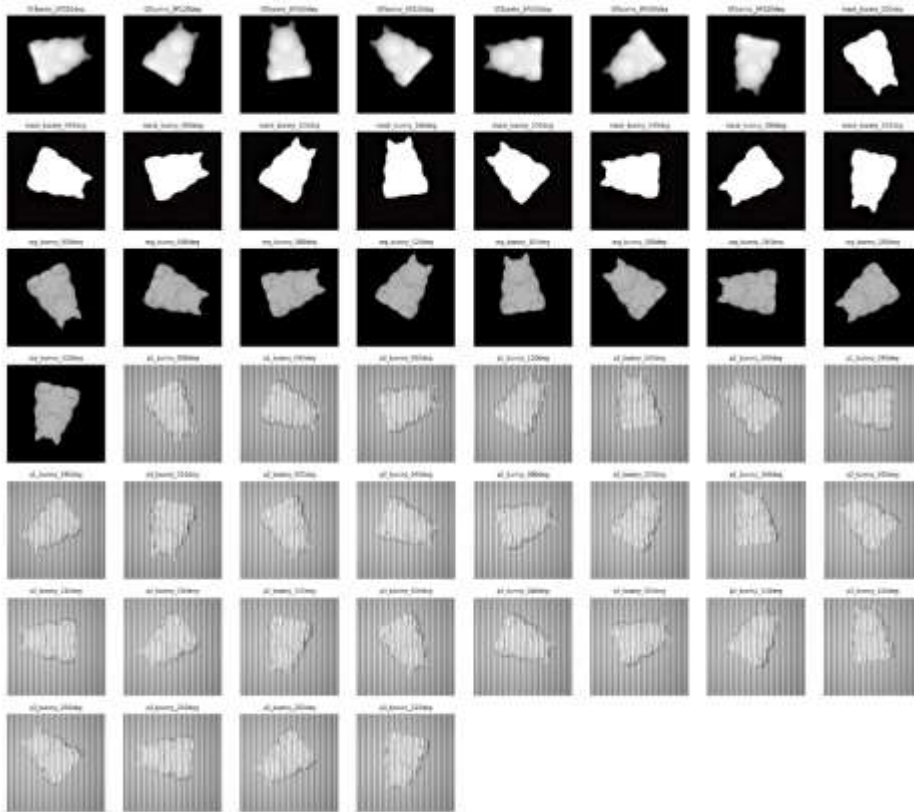


Figura 19. Conjunto de imágenes base correspondientes a un objeto sintético.

Es el conjunto completo bajo distintas rotaciones angulares, incluyendo *org*, *mask*, proyecciones sinusoidales y mapas de ground truth.

5.1.1.1 *Generación de mapas de fase sintéticos*

A partir de las imágenes con franjas (p1, p2, p3), se obtuvieron los mapas de fase envuelta mediante dos técnicas de detección de fase:

- **Método de tres pasos (3-step)** según el algoritmo de Malacara, el cual produce mapas de fase **multi-wrap** con múltiples brincos (≈ 25 saltos de 2π).
- **Perfilometría de Fourier bidimensional (FT)**, donde se aplicó la Transformada de Fourier para extraer el espectro de amplitud, seleccionar la frecuencia portadora f_0 y, tras la Transformada Inversa, obtener el mapa de fase envuelta **low-wrap**.
- En el caso del **multi-wrap por FT**, se omitió la selección del pico portador, lo que conserva múltiples envolventes en la fase recuperada.

Estas técnicas dieron origen a bases de datos independientes, cada una correspondiente a distintos experimentos:

- **BD1:** 99 mapas de fase multi-wrap (3-step) donde todavía no se contaba con el GT con tipo SP.
- **BD2:** 528 mapas de fase multi-wrap (3-step).
- **BD3:** 999 mapas de fase low-wrap (FT).
- **BD4:** 1593 mapas de fase low-wrap (FT).
- **BD5:** 1593 mapas de fase multi-wrap (FT).

Cada base de datos está compuesta por los siete elementos por objeto (p1, p2, p3, mask, org, GT-SP, GT-CA) más el mapa de fase envuelta correspondiente.

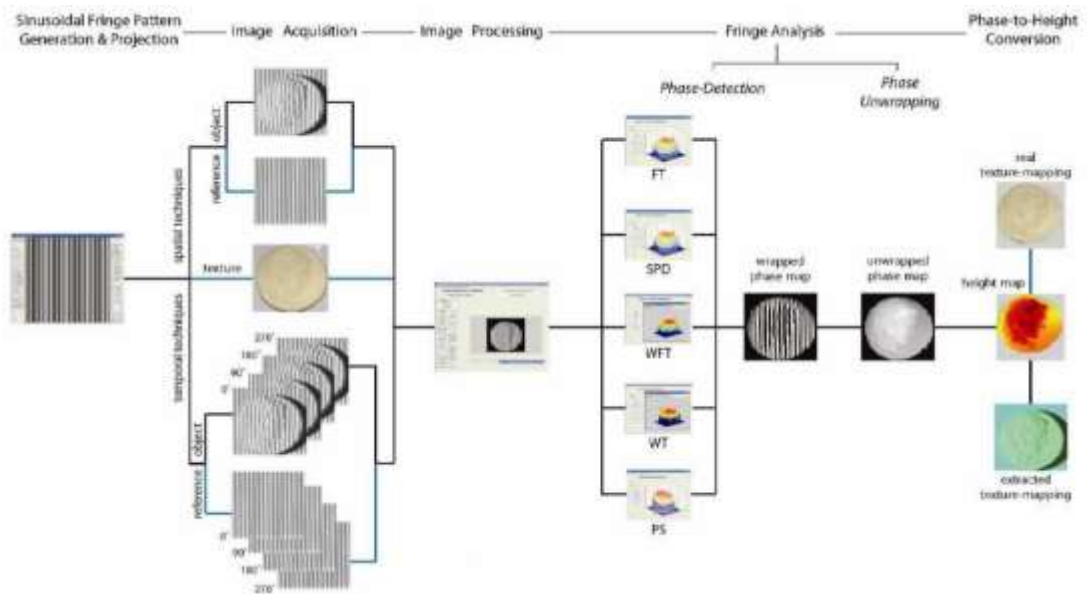


Figura 20. Flujo general de un sistema de proyección de franjas
 Adaptado de Gorthi y Rastogi (2010). La presente tesis utiliza este diagrama para contextualizar la obtención de mapas de fase envuelta a partir de imágenes sintéticas y reales.

La estructura general del sistema de proyección de franjas utilizado en esta investigación sigue el flujo metodológico descrito por (Gorthi & Rastogi, 2010a) quienes presentan un análisis detallado de las etapas fundamentales en un sistema típico de perfilometría basada en franjas sinusoidales. Este esquema, ilustrado en la Figura 20, comprende las fases de:

1. Generación y proyección de patrones sinusoidales,
2. Adquisición de imágenes del objeto y de la referencia,
3. Procesamiento de imágenes,
4. Detección de fase (Phase Detection),
5. Desenvolvimiento de fase (Phase Unwrapping) y
6. Conversión de fase a altura (Phase-to-Height Conversion).

Aunque la presente tesis se enfoca en la etapa de desenvolvimiento de fase, es necesario comprender el flujo completo debido a que el desempeño del *unwrapping* depende directamente de la calidad y características del mapa de fase envuelta obtenido en la fase anterior. Por ello, en la etapa de *Phase Detection* se emplearon dos métodos ampliamente utilizados en la literatura, también revisados por Gorthi y Rastogi:

- Transformada de Fourier (FT), para obtener mapas de fase tipo *low-wrap* y *multi-wrap*
- Método de tres pasos de Malacara, para generar mapas *multi-wrap* con múltiples saltos.

La generación de los conjuntos sintéticos (p1, p2, p3) con desplazamientos de fase de 0° , 120° y 240° corresponde exactamente al esquema descrito por estos autores, donde los desplazamientos controlados permiten la recuperación de la fase envuelta mediante técnicas de demodulación temporal o espacial.

5.1.2 Imágenes reales

Además de los datos sintéticos, se empleó una base de datos real previamente disponible, compuesta por aproximadamente **30 objetos** capturados mediante un sistema experimental de proyección de franjas, en la Figura 21 podemos ver tres de los objetos reales con las franjas proyectadas. Este sistema utilizó un proyector que generaba patrones sinusoidales, aunque las condiciones de adquisición variaban significativamente entre objetos (distancia cámara-objeto, número de franjas, orientación, intensidad y configuración del proyector).

A partir de cada conjunto de imágenes reales, se obtuvieron los mapas de fase envuelta mediante dos técnicas:

- **Perfilometría de Fourier (FT) low-wrap**, aplicando la selección del pico portador.
- **Perfilometría de Fourier multi-wrap**, omitiendo la selección del pico para conservar múltiples envolventes.

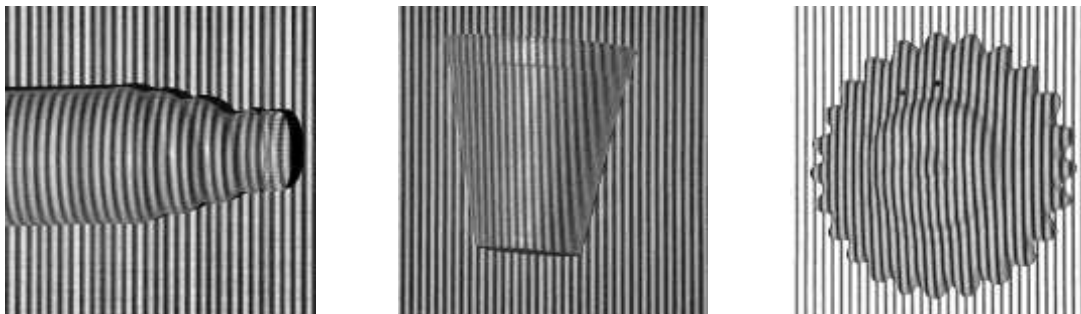
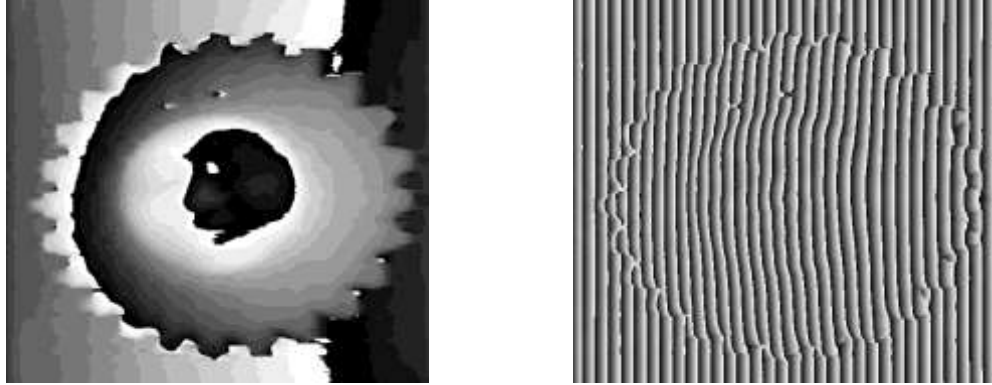


Figura 21. Ejemplos de proyección de franjas sobre objetos reales.

Estos mapas de fase se utilizaron exclusivamente para la etapa de evaluación, permitiendo comparar visual y cualitativamente los resultados del modelo entrenado frente a los métodos tradicionales de desdoblamiento de fase en condiciones reales, en la Figura 22 podemos observar el ejemplo de los mapas de fase obtenido de un objeto real.



*Figura 22. Mapas de fase
En la ilustración podemos ver ejemplo de dos tipos de mapas de fase a) Mapa de fase low-wrap de objeto real y en el b) Mapa de fase multi-wrap de objeto real*

La variabilidad en iluminación, distancia, ruido óptico y geometría de los objetos reales hace que esta base de datos represente adecuadamente los retos característicos de un sistema práctico, complementando el análisis sobre los datos sintéticos donde se tiene control total de los parámetros de generación.

5.2 Establecimiento de métricas

La evaluación cuantitativa del desempeño de los métodos de desdoblamiento de fase constituye un componente esencial para garantizar la validez y rigurosidad de los resultados obtenidos en esta investigación. Con el fin de comparar de manera objetiva la fidelidad de las reconstrucciones frente al ground truth sintético, se emplearon tres métricas ampliamente reconocidas en el procesamiento de imágenes: el Error Cuadrático Medio en Imágenes (IMMSE), la Relación Señal-Ruido de Pico (PSNR) y el Índice de Similitud Estructural (SSIM). Estas métricas permiten analizar la precisión numérica, la preservación de la energía

de la señal y la similitud estructural entre la reconstrucción obtenida y la referencia, proporcionando así una evaluación integral de los resultados.

En todos los experimentos se utilizó como referencia el mapa de ground truth GT-SP, debido a que mostró la mayor estabilidad geométrica y fidelidad en la representación de la superficie para la mayoría de los objetos sintéticos.

5.2.1 Evaluación cuantitativa de los métodos tradicionales

En el caso de los métodos tradicionales de desdoblamiento de fase (Herraez, Costantini, Robust2D, Fourier multi-wrap, Fourier low-wrap, entre otros), las métricas se calcularon exclusivamente dentro de la máscara binaria del objeto, con el propósito de centrar el análisis en la región físicamente relevante y evitar contribuciones del fondo que pudieran sesgar los resultados. Adicionalmente, tanto la reconstrucción como el ground truth fueron normalizados entre 0 y 1 dentro de la máscara, de modo que el fondo quedara en 0 y la altura del objeto se proyectara hacia 1. Esta normalización es necesaria dado que, al exportar el ground truth desde Blender hacia MATLAB, pueden producirse desplazamientos verticales y variaciones de escala en la fase que no corresponden a propiedades reales del objeto, sino a diferencias inherentes al proceso de renderizado y lectura de archivos. Al normalizar, la comparación se centra únicamente en la fidelidad morfológica de las reconstrucciones.

5.2.2 Ajuste global de altura

Con el fin de corregir el desplazamiento vertical que se produce al transferir el ground truth desde Blender hacia MATLAB, se implementó un procedimiento sistemático basado en un barrido del parámetro Δ . Para cada objeto, se probó un conjunto de valores de Δ en el rango de **0 a 0.5**, con incrementos de **0.05**, aplicando la operación:

$$I_{\Delta}(x, y) = I(x, y) + \Delta,$$

donde $I(x, y)$ representa el mapa reconstruido y $I_{\Delta}(x, y)$ la versión ajustada. Este parámetro actúa como una corrección de *offset* a nivel global que permite alinear verticalmente las superficies antes de realizar la comparación cuantitativa con el ground truth.

Para determinar el Δ óptimo en cada método, se empleó una evaluación compuesta basada en las tres métricas utilizadas (IMMSE, PSNR y SSIM). Tras calcular cada métrica dentro de la máscara del objeto y normalizarla entre 0 y 1, se definió un **puntaje global** mediante una combinación lineal ponderada:

- 0.3 para IMMSE,
- 0.3 para PSNR,
- 0.4 para SSIM.

El valor de Δ que **maximizó este puntaje** fue seleccionado como el óptimo para ese método y se utilizó posteriormente en la comparación final entre todos los métodos tradicionales.

5.2.3 Definición matemática

Las métricas cuantitativas empleadas en esta investigación se definen formalmente de la siguiente manera:

5.2.3.1 *Error Cuadrático Medio en Imágenes (IMMSE)*

El IMMSE cuantifica la diferencia promedio entre la imagen reconstruida $K(x, y)$ y el ground truth $I(x, y)$. Valores pequeños de IMMSE indican mayor correspondencia entre ambas imágenes.

5.2.3.2 *Relación Señal-Ruido de Pico (PSNR)*

La métrica PSNR compara la potencia máxima del rango dinámico de la imagen con el nivel de error cuadrático. Valores elevados de PSNR (>30 dB) sugieren una alta similitud entre la imagen reconstruida y el ground truth.

5.2.3.3 *Índice de Similitud Estructural (SSIM)*

El SSIM evalúa la similitud entre imágenes considerando luminancia, contraste y estructura. El SSIM se aproxima a 1 cuando la estructura del objeto se preserva de manera fiel.

5.2.4 Evaluación cuantitativa de los modelos de aprendizaje profundo

Para la evaluación de los modelos basados en redes neuronales convolucionales, se siguió un enfoque distinto al empleado en los métodos tradicionales. En este caso, no se aplicó la máscara ni se normalizaron las imágenes antes del cálculo de métricas; esto permite observar el comportamiento del modelo en condiciones más exigentes y evaluar su estabilidad frente a variaciones en el rango dinámico o en la escala de las predicciones.

Para asegurar una comparación justa entre la salida del modelo y el ground truth, las predicciones fueron sometidas a un ajuste afín de intensidades, mediante el cual se encuentran los coeficientes a y b que minimizan el error cuadrático medio entre ambas imágenes, obteniendo una versión ajustada:

$$K'(x, y) = a \cdot K(x, y) + b.$$

Este ajuste permite corregir diferencias lineales de luminancia que no afectan la estructura del mapa de fase. Asimismo, se incluyó la opción de corregir pequeños desplazamientos espaciales mediante correlación de fase, lo cual es útil en casos donde las predicciones presentan ligeros desalineamientos.

Posteriormente, se calcularon las métricas IMMSE, PSNR y SSIM empleando el rango dinámico propio de cada imagen. El SSIM se estimó mediante una ventana gaussiana con tamaño adaptativo, lo que asegura estabilidad aún en imágenes con poco contraste local. Finalmente, se implementó una función capaz de evaluar lotes completos de ejemplos, permitiendo obtener medidas estadísticas como la media y la desviación estándar por métrica para cada experimento.

Este esquema permitió comparar diferentes versiones del modelo, analizar su consistencia y determinar su capacidad para preservar la forma y estructura del objeto en el proceso de desdoblamiento de fase.

5.3 Implementación de métodos tradicionales.

Con el objetivo de contar con un punto de referencia sólido frente al cual comparar el desempeño de la red neuronal convolucional propuesta, se implementó un conjunto de métodos tradicionales de desdoblamiento de fase espacial, entendidos como aquellos que no incorporan técnicas de aprendizaje automático ni inteligencia artificial. Todos los algoritmos fueron desarrollados e integrados en MATLAB, a partir de implementaciones disponibles en la literatura y adaptadas al flujo de trabajo de esta tesis.

Los métodos de desdoblamiento considerados fueron: Itoh, Ghiglia, Zhao, Constantini, Herraez, PUMA, PEARLS y el método de Martínez-Carranza/STIEPU. La descripción teórica y el fundamento matemático de cada uno de ellos se presentan en la sección de Fundamentación Teórica; en esta sección se detalla únicamente su integración en el pipeline experimental, los parámetros empleados y el procedimiento seguido para su evaluación.

Para el análisis comparativo sistemático se seleccionó un subconjunto de seis objetos sintéticos representativos, abarcando variaciones en geometría, relieve y complejidad superficial. En estos objetos se dispone de ground truth sintético (GT-SP), lo que permite cuantificar de manera objetiva el error de reconstrucción. En el caso de objetos reales, los métodos tradicionales se emplearon con fines de comparación cualitativa, sin cálculo de métricas cuantitativas.

5.3.1 Preparación de los mapas de fase

Los métodos tradicionales de desdoblamiento de fase operan sobre mapas de fase envuelta. En esta tesis, dichos mapas se obtuvieron a partir de las imágenes de franjas generadas en la etapa de detección de fase (perfilometría de Fourier bidimensional y método de tres pasos), descrita previamente en la metodología. Para cada objeto se dispone de:

- un mapa de fase del objeto con franjas proyectadas (*MapaFaseObjeto*),
- un mapa de fase correspondiente al plano de referencia (*MapaFaseReferencia*).

Ambos mapas se calcularon mediante la Transformada Rápida de Fourier bidimensional (FFT2), filtrado en el dominio de la frecuencia alrededor de la frecuencia portadora f_0 , desplazamiento adecuado del espectro y posterior Transformada Inversa de Fourier, siguiendo el esquema clásico de perfilometría por proyección de franjas.

Con el fin de delimitar la región de interés, se empleó una máscara binaria $M(x, y)$ asociada al objeto. Esta máscara se selecciona de manera interactiva, se redimensiona al tamaño de los mapas de fase y se refina mediante una operación de erosión morfológica para eliminar artefactos en el contorno. Posteriormente, los mapas de fase del objeto y del plano de referencia se multiplican por la máscara:

$$\phi_{\text{obj}}^M(x, y) = \phi_{\text{obj}}(x, y) M(x, y), \phi_{\text{ref}}^M(x, y) = \phi_{\text{ref}}(x, y) M(x, y).$$

En los experimentos con *ground truth*, estos mapas se emplean además en la etapa de normalización en la máscara y en el cálculo de métricas descrito en la Sección 2.

5.3.2 Aplicación de los métodos de desdoblamiento de fase

Una vez definidos los mapas de fase envuelta y la máscara, se procede a aplicar los métodos tradicionales. Cada algoritmo recibe como entrada los mapas $\phi_{\text{obj}}(x, y)$ y $\phi_{\text{ref}}(x, y)$ y genera como salida las fases desenvueltas correspondientes:

$$u_k(x, y) \text{ para el objeto, } v_k(x, y) \text{ para el plano de referencia,}$$

donde el subíndice k indica el método utilizado.

La implementación concreta en MATLAB se estructuró de la siguiente forma:

- **Itoh:** se utilizó la función estándar `unwrap` en ambas direcciones (horizontal y vertical), aplicando un desdoblamiento secuencial en filas y columnas para obtener un mapa de fase continua.

- **Herraez** se empleó la función `unwrap_phase`, basada en un enfoque *quality-guided*, que ordena los píxeles de acuerdo con un mapa de calidad y propaga el desdoblamiento desde las regiones más confiables hacia las menos confiables.
- **Zhao**: se utilizó `Unwrap_TIE_DCT_Iter`, un método iterativo que combina la ecuación de transporte de intensidad (TIE) con transformadas coseno discretas (DCT), adecuado para datos ruidosos y con discontinuidades suaves.
- **Constantini**: se aplicó la función `cunwrap` con parámetros `RoundK = true` y `maxblocksize = 60`, siguiendo un enfoque de minimización global tipo *minimum-norm* sobre bloques, que equilibra precisión y costo computacional.
- **Martínez-Carranza / STIEPU**: se utilizó `stiepu`, que implementa un esquema espacial robusto para el desdoblamiento, con énfasis en la preservación de discontinuidades estructurales.
- **Ghiglia**: se utilizó `phase_unwrap`, asociado a métodos de desdoblamiento basados en formulaciones de Poisson y técnicas espectrales (FFT), adecuados para datos con variaciones suaves.
- **PUMA**: se empleó `puma_ho` con exponente del potencial $p = 2$ y un conjunto de cliques bidimensionales, modelando el problema como una minimización de energía en un campo aleatorio de Markov. Este método es robusto al ruido y a inconsistencias locales.
- **PEARLS**: se integró el método PEARLS, utilizando una familia de ventanas de tamaño `windows = [1,2,3,4]`, parámetro $\gamma = 2$ para el criterio de información (ICI), y la opción `unwrapp = 'yes'` para aplicar internamente PUMA como módulo de desdoblamiento. Los parámetros del potencial y del umbral se ajustaron a los valores recomendados en la literatura (`threshold = 2`, `quantized = 'no'`, `p_exponent = 1`).

La combinación de estos métodos permite abarcar distintos enfoques de desdoblamiento espacial (guiado por calidad, global, robusto, espectral y basado en MRF), constituyendo una base comparativa amplia para el análisis posterior.

5.3.3 Construcción de mapas de reconstrucción y normalización en la máscara

Para cada método k , una vez obtenidos los mapas de fase desenvuelta del objeto y del plano de referencia, se construyó un mapa de reconstrucción relativo a partir de la diferencia:

$$R_k(x, y) = v_k(x, y) - u_k(x, y),$$

el cual está asociado a la altura relativa del objeto con respecto al plano. Este mapa se procesa únicamente dentro de la máscara del objeto $M(x, y)$, aplicando las siguientes operaciones:

1. **Aplicación de máscara:**

$$R_k^M(x, y) = R_k(x, y) M(x, y).$$

2. **Filtrado mediano:** se aplica un filtro mediano bidimensional (medfilt2) con ventana 15×15 para reducir ruido impulsivo y pequeñas irregularidades locales.
3. **Recorte de valores negativos:** se fuerza la positividad de la señal mediante una función tipo

$$\tilde{R}_k(x, y) = \max \{R_k^M(x, y), 0\},$$

garantizando que no se introduzcan valores negativos no físicos.

4. **Normalización en la máscara:** finalmente, el mapa se normaliza en el rango $[0,1]$ considerando únicamente los píxeles dentro de la máscara:

$$R_{k,\text{norm}}(x, y) = \frac{\tilde{R}_k(x, y) - \min_{(x,y) \in M} \tilde{R}_k(x, y)}{\max_{(x,y) \in M} \tilde{R}_k(x, y) - \min_{(x,y) \in M} \tilde{R}_k(x, y) + \varepsilon},$$

fijando el fondo en 0 y la altura máxima del objeto en 1.

Estos mapas normalizados $R_{k,\text{norm}}(x, y)$ se utilizaron tanto para la inspección visual (a través de representaciones en escala de grises y mallas tridimensionales) como para el cálculo de métricas en los casos en que se dispone de *ground truth*, tal como se detalla en la Sección 2.

5.3.4 Organización de los resultados

El script de implementación contempla dos escenarios:

1. Escenario con *ground truth*

En el caso de objetos sintéticos, se carga el mapa de *ground truth* (GT-SP), se corrige un desplazamiento global mediante un parámetro Δy se normaliza dentro de la máscara. A continuación, las reconstrucciones $R_{k,\text{norm}}(x, y)$ se alinean con el *ground truth* a través de una función auxiliar que ajusta la escala y coordena el máximo de altura. Sobre estos mapas alineados se calculan las métricas IMMSE, PSNR y SSIM en la región enmascarada, siguiendo el procedimiento descrito en la Sección 2. Los resultados se almacenan en tablas por método y por corrida, junto con información del objeto, del GT utilizado y del valor de Δ .

2. Escenario sin *ground truth*

En el caso de objetos reales, el pipeline se limita a la generación de mapas normalizados $R_{k,\text{norm}}(x, y)$ y a su visualización comparativa mediante representaciones tipo mesh e imagesc, así como superposición de contornos del objeto. En este escenario no se calculan métricas cuantitativas, sino que los métodos se emplean para análisis cualitativos y para ilustrar el comportamiento de las diferentes estrategias de desdoblamiento en condiciones experimentales reales.

En ambos escenarios, las salidas se organizan automáticamente en carpetas de resultados que incluyen las figuras generadas (mapas 2D, mallas 3D, comparaciones por método) y archivos .mat con la información esencial de cada corrida, facilitando la trazabilidad de los experimentos y su posterior análisis en la sección de Resultados.

5.4 Diseño y entrenamiento de la red neuronal convolucional

El desarrollo del modelo de aprendizaje profundo se estructuró mediante la implementación y comparación de cuatro arquitecturas convolucionales distintas, diseñadas para abordar el desdoblamiento de fase espacial como un problema de regresión de mapas continuos. Estas arquitecturas se seleccionaron con el propósito de evaluar el impacto de diferentes estrategias de extracción de características, combinación multi-resolución y mecanismos recurrentes sobre la capacidad del modelo para recuperar geometrías suaves y continuas a partir de mapas de fase envueltos. Todas las redes fueron implementadas en TensorFlow/Keras y entrenadas bajo un esquema homogéneo de partición y validación.

5.4.1 Arquitecturas

Las cuatro arquitecturas evaluadas fueron: (1) U-Net con SQD-LSTM, (2) U-Net-ResNet, (3) CNN Multiresolution y (4) U-Net simple.

La primera arquitectura corresponde a una U-Net modificada que incorpora un bloque de *Quad-Directional LSTM* aplicado sobre los mapas de características más profundos. Este mecanismo recurrente permite que la red capture dependencias espaciales de largo alcance en cuatro direcciones, mejorando la continuidad y coherencia del mapa reconstruido. Posteriormente, el decodificador recupera la resolución original mediante convoluciones transpuestas y *skip connections* clásicas de U-Net.

La segunda arquitectura implementada fue un modelo **U-Net-ResNet**, donde tanto el codificador como el decodificador están compuestos por bloques residuales de dos convoluciones con adición por atajo (*shortcut add*). Esta estructura favorece la estabilidad del gradiente durante el entrenamiento y mejora la capacidad de la red para modelar detalles finos sin degradación por profundidad. El diseño residual se conservó de manera simétrica en la etapa de reconstrucción.

El tercer modelo corresponde a una arquitectura multiresolución, inspirada en la familia MoiréCNN. Esta red emplea un flujo piramidal con cinco niveles de resolución y múltiples

ramas de reconstrucción paralelas. Cada rama procesa la señal en una escala diferente y genera una predicción intermedia que posteriormente se alinea y suma con las demás para producir la salida final. Esta estrategia permite integrar información global y local, facilitando la recuperación de superficies con variaciones abruptas y regiones suavemente cambiantes.

Finalmente, se implementó una U-Net clásica, utilizada como arquitectura base de referencia. Este modelo incluye un codificador–decodificador simétrico con *skip connections*, convoluciones 3×3 y bloques de activación ReLU, proporcionando una línea base estable para comparar el desempeño de las arquitecturas más complejas.

5.4.2 Función de pérdida

Independientemente de la arquitectura, todas las redes fueron entrenadas utilizando una función de pérdida compuesta diseñada específicamente para este problema. La pérdida integra dos términos: la varianza del error y el total variation de la diferencia entre el mapa predicho y el real. El primer término penaliza desplazamientos globales y asegura que el error promedio se mantenga bajo, mientras que el segundo promueve suavidad espacial y continuidad en gradientes entre vecinos:

$$\mathcal{L} = L_{\text{var}} + 0.1 L_{\text{tv}}$$

donde L_{var} se calcula como la varianza del error por muestra y L_{tv} corresponde a la diferencia absoluta entre gradientes espaciales de la predicción y del mapa objetivo en ambos ejes. Se realizaron experimentos preliminares con MSE y MAE, pero no lograron un comportamiento adecuado para preservar continuidad geométrica, por lo que se optó por esta pérdida compuesta.

5.4.3 Datos de entrenamiento

Las redes se entrenaron utilizando mapas de fase sintéticos generados mediante métodos de detección de fase en perfiles de Fourier, en dos modalidades: low-wrap y multi-wrap. Para

evitar que la red aprendiera patrones específicos de un único tipo de distorsión, los tipos de mapas no se mezclaron dentro del mismo entrenamiento; cada arquitectura fue entrenada de manera independiente para cada tipo de fase.

Se emplearon dos resoluciones: 256×256 y 512×512 , dependiendo del experimento. Todas las imágenes fueron normalizadas entre 0 y 1, y aquellos casos con mapas negros o inválidos fueron descartados para evitar deterioro en el entrenamiento. No se aplicaron máscaras ni normalización específica sobre el contenido de los objetos, ya que el objetivo fue evaluar la capacidad del modelo para generalizar bajo condiciones extremas sin restricciones adicionales.

Para cada conjunto se utilizó un esquema de partición 80 % entrenamiento, 10 % validación y 10 % prueba. Además, se implementaron dos estrategias de división de datos:

- **Mixed-object split:** todas las rotaciones y variaciones de los objetos se mezclan antes de la partición.
- **Object-wise split:** todas las rotaciones de un mismo objeto se asignan a un solo subconjunto (train, val o test).

El primer esquema permite aprovechar un mayor número de muestras en el entrenamiento, mientras que el segundo permite verificar si el modelo realmente generaliza a objetos no vistos en ninguna de sus orientaciones. La comparación entre ambas configuraciones se presenta posteriormente en la sección de resultados.

5.4.4 Procedimiento de entrenamiento

Todos los modelos fueron entrenados en Google Colab, utilizando aceleración por GPU provista por la plataforma. En la mayoría de los experimentos se empleó una NVIDIA A100 con memoria de alta capacidad, lo cual permitió entrenar arquitecturas profundas y manejar imágenes de 512×512 sin limitaciones. En algunos casos se utilizaron GPUs equivalentes asignadas dinámicamente por Colab, manteniendo configuraciones homogéneas de entrenamiento en todos los experimentos.

Los modelos se entrenaron utilizando el optimizador Adam con una tasa de aprendizaje inicial de 1×10^{-3} y un tamaño de lote de 4, establecido como equilibrio entre estabilidad numérica y uso eficiente de la memoria de la GPU.

Las redes entrenadas a resolución 256×256 se ejecutaron durante 100 épocas, mientras que aquellas entrenadas a 512×512 se entrenaron durante 120 épocas, salvo cuando el proceso se detuvo anticipadamente mediante early stopping con una paciencia de 10 épocas, monitoreando la pérdida de entrenamiento (loss). Además, se utilizó ModelCheckpoint para almacenar siempre el mejor modelo de acuerdo con la pérdida de validación (val_loss), garantizando que el peso final correspondiera al punto óptimo del entrenamiento.

5.4.5 Fine-tuning

Además de los entrenamientos principales, se realizó un experimento adicional de ajuste fino (fine-tuning) con el objetivo de adaptar el modelo previamente entrenado a un subconjunto específico de mapas de fase y mejorar su capacidad de generalización en la etapa final de inferencia sobre datos reales.

Este proceso se aplicó sobre el modelo base mejor evaluado (aquella arquitectura que alcanzó el mayor desempeño en las pruebas comparativas; la selección exacta se describe en la sección de Resultados). Para este ajuste final se reutilizaron los pesos previamente entrenados, con el fin de acelerar la convergencia y refinar la sensibilidad del modelo hacia variaciones de fase que aparecen en imágenes reales, tales como atenuaciones, gradientes suaves y regiones localizadas con pérdida de información.

5.4.5.1 *Data augmentation especializado en mapas de fase*

Dado que los mapas de fase reales presentan patrones de degradación distintos a los sintéticos, se implementó un esquema de data augmentation diseñado específicamente para este tipo de señal. Este mecanismo aplica variaciones que preservan la estructura física del fenómeno, pero aumentan la robustez del modelo ante perturbaciones inevitables en imágenes experimentales.

El procedimiento incluye:

1. **Escala y desplazamiento global ($a \cdot \phi + b$):**

Cambia la estadística global de la fase, simulando variaciones de contraste o calibración.

2. **Ruido suave filtrado Gaussianamente:**

Emula gradientes amplios o iluminación no uniforme.

3. **Ruido granular de alta frecuencia:**

Reproduce pequeñas inconsistencias del sensor o imperfecciones en el patrón proyectado.

4. **Zonas deterioradas o “dropouts”:**

Simula regiones afectadas por sombras, saturación o reflexión especular.

5. **Nuevo envolvimiento de la fase (wrapping):**

Garantiza que la señal permanezca físicamente consistente en el rango $[-\pi, \pi]$.

Este esquema se integró mediante un generador personalizado (AugmentedPhaseGenerator) que aplica estas perturbaciones dinámicamente en cada lote, garantizando que el modelo reciba muestras distintas en cada época del fine-tuning.

5.4.5.2 *Entrenamiento del ajuste fino*

El entrenamiento del fine-tuning se realizó por 20 épocas, utilizando:

- optimizador Adam (LR con reducción automática),
- ReduceLROnPlateau (paciencia 3, factor 0.5),
- EarlyStopping con paciencia 6,
- restauración automática de los mejores pesos al final.

El entrenamiento se realizó con las mismas resoluciones utilizadas durante la etapa inicial (según el modelo base seleccionado) y aprovechando el generador con data augmentation para el conjunto de entrenamiento. La validación se realizó sobre un subconjunto separado sin augmentación, a fin de medir correctamente el refinamiento alcanzado por el modelo.

Este ajuste final permitió adaptar la red a condiciones más cercanas a la operación real del sistema de proyección, reforzando su desempeño antes de ejecutar las inferencias sobre imágenes reales y su comparación con los métodos tradicionales de desdoblamiento de fase.

5.5 Documentación y publicación

Durante el desarrollo del proyecto se mantuvo un proceso de documentación sistemática orientado a garantizar la trazabilidad, reproducibilidad y respaldo completo de cada etapa metodológica. Todas las rutinas de generación de datos sintéticos, los métodos tradicionales de desdoblamiento de fase implementados en MATLAB y los experimentos con redes neuronales desarrollados en TensorFlow/Keras dentro de Google Colab fueron archivados y organizados en repositorios de trabajo. Esto incluyó scripts versionados, cuadernos de experimentación, modelos entrenados, configuraciones de hiperparámetros y registros completos de métricas y curvas de aprendizaje.

Asimismo, las figuras, mapas de fase, máscaras, normalizaciones, reconstrucciones intermedias y resultados cuantitativos se almacenaron en carpetas estructuradas por arquitectura, resolución y tipo de mapa (low-wrap o multi-wrap). Esta organización permitió la elaboración coherente de los capítulos posteriores y aseguró que cada visualización utilizada en la tesis correspondiera fielmente al experimento descrito.

Derivado de esta investigación se generó una primera publicación titulada “Estudio comparativo de algoritmos de desdoblamiento de fase espacial y aplicación en reconstrucción 3D”, la cual se desarrolló utilizando un subconjunto representativo de seis objetos sintéticos. Este trabajo sirvió como base conceptual y metodológica para la etapa posterior del proyecto, ampliada en la presente tesis mediante arquitecturas de aprendizaje profundo y un análisis más completo.

Finalmente, a partir de los resultados obtenidos durante esta investigación, se planea la elaboración de un segundo manuscrito dirigido a una revista especializada en procesamiento de imágenes o técnicas de medición óptica, con el propósito de difundir los avances logrados

en el desdoblamiento de fase mediante modelos híbridos tipo U-Net y CNN multiescala. La documentación generada en este proyecto proporciona los elementos necesarios para sustentar la reproducibilidad del futuro artículo.

VI. Resultados y discusión

En este capítulo se presentan y analizan los resultados obtenidos a lo largo de todas las etapas experimentales del proyecto, desde el desdoblamiento de fase con métodos tradicionales hasta la evaluación de diversas arquitecturas de aprendizaje profundo y su posterior adaptación al dominio real mediante fine-tuning. Dado que la calidad de la fase reconstruida depende fuertemente del tipo de mapa de entrada, la naturaleza de la base de datos y la robustez del algoritmo, los resultados se organizan para reflejar esta relación de manera progresiva: primero se examina el desempeño de los algoritmos clásicos aplicados a seis objetos sintéticos con ground truth conocido; posteriormente, se comparan múltiples configuraciones de modelos neuronales entrenadas bajo diferentes resoluciones, tamaños de dataset y tipos de fase (low-wrap y multi-wrap). En esta sección se discuten tanto las métricas cuantitativas, PSNR, SSIM e IMMSE, como la consistencia visual de las reconstrucciones, resaltando patrones, ventajas y limitaciones particulares de cada enfoque. Finalmente, se estudia la capacidad de generalización de los modelos hacia datos experimentales reales y se analiza el efecto del fine-tuning, cuyo objetivo fue reducir la brecha entre las condiciones ideales de los mapas sintéticos y la complejidad inherente de la fase obtenida mediante proyección de franjas. Este capítulo integra todas estas observaciones para ofrecer una visión crítica y unificada del rendimiento de cada método y del camino seguido hasta obtener modelos más robustos en escenarios reales.

6.1 Bases de datos

La calidad y diversidad de las bases de datos fueron un factor determinante en el desempeño de los métodos comparados y en la estabilidad del entrenamiento de las redes neuronales. En esta sección se describen brevemente los conjuntos empleados, las decisiones de diseño tomadas en su construcción y los retos encontrados.

6.1.1 Objetos sintéticos

Se generaron cinco bases principales de mapas de fase sintéticos, creadas mediante un pipeline completo de modelado 3D en Blender, renderizado fotométrico y obtención de mapas de fase multi-wrap o low-wrap mediante los métodos descritos en la metodología (3-step y FT). La figura general de organización de las bases se resume en la Tabla de Experimentos (Sección de Entrenamientos), donde cada variante del dataset está asociada a un conjunto de entrenamiento específico (V1–V21).

6.1.1.1 Conjuntos generados

1. BD1 – 99 mapas multi-wrap (3-step)

Primera aproximación; carecía de un ground truth SP confiable por lo que los resultados fueron pobres durante el fine-tuning inicial.

2. BD2 – 528 mapas multi-wrap (3-step)
3. BD3 – 999 mapas low-wrap (FT)
4. BD4 – 1593 mapas low-wrap (FT)
5. BD5 – 1593 mapas multi-wrap (FT)

Los resultados de entrenamiento mostraron que los modelos convergieron con mayor estabilidad cuando se trabajó con mapas en formato low-wrap obtenidos mediante Transformada de Fourier (FT):

- Los mapas low-wrap presentan *menos discontinuidades* y menor ambigüedad espacial.
- Las fases son más suaves, lo que facilita el aprendizaje de los gradientes espaciales.
- Los GT normalizados dentro de la máscara presentaron valores físicamente consistentes.

Por el contrario, los mapas multi-wrap obtenidos con 3-step tendieron a degradar el entrenamiento debido a:

- la ausencia inicial de un GT SP fiable para la base BD1;
- múltiples regiones con saltos abruptos de 2π que incrementan el error de reconstrucción;
- variabilidad entre objetos que afectó la generalización sin un preprocesamiento adecuado.

Esto explica por qué las versiones V17–V21 (U-Net simple) y V7–V10 (U-Net–ResNet), entrenadas con BD3–BD5, obtuvieron las métricas más altas en SSIM, PSNR e IMMSE.

6.1.1.2 *Low-wrap vs. multi-wrap*

Para clarificar esta diferencia en los resultados, podemos observar la Figura # donde se incorporan ejemplo para crear una comparativa donde se muestran:

- low-wrap: fases suaves, un solo salto o ninguno, más adecuadas para regresión directa;
- multi-wrap: múltiples saltos (hasta ~ 25), mayor complejidad, útiles para evaluar robustez.

Este contraste permite interpretar adecuadamente los comportamientos observados en los experimentos.

6.1.2 **Objetos reales**

Para el análisis con datos reales se utilizó una base previamente disponible, compuesta por aproximadamente 30 objetos físicos sobre los cuales se proyectaron patrones sinusoidales de franjas. A diferencia de la base sintética, los objetos reales:

- solo cuentan con una imagen del objeto con franjas proyectadas,
- la imagen correspondiente del plano de referencia (también con franjas).

No se adquirieron imágenes con múltiples desplazamientos de fase (0° , 120° , 240°), por lo que el cálculo de la fase no se realizó mediante el método de tres pasos, sino exclusivamente mediante el esquema completo de perfilometría por Transformada de Fourier (FT).

A partir de estas imágenes se generaron los mapas de fase de multi-wrap y low-wrap.

6.1.3 Relevancia de la construcción de bases

Durante el desarrollo del proyecto quedó claro que la creación y depuración de las bases de datos fue crítica. Algunos de los principales desafíos fueron:

- La falta inicial de un GT SP confiable para BD1, lo que generó *sesgos significativos* durante el fine-tuning temprano del modelo SQD-LSTM (el modelo aprendía a reproducir fases incorrectas).
- La necesidad de normalización en máscara, debido a inconsistencias de escala entre Blender y MATLAB.
- La diferenciación entre mapas low-wrap y multi-wrap para evitar que la red aprendiera patrones inestables o físicamente irrelevantes.
- El control de variabilidad mediante splits por objeto y splits por mezcla, experimentados para evitar sobreajuste entre muestras muy similares.

En conjunto, estas bases permitieron entrenar modelos robustos, así como evaluar sus alcances y limitaciones con datos sintéticos y reales.

6.2 Métodos tradicionales

En esta sección se presentan los resultados obtenidos al aplicar ocho métodos tradicionales de desdoblamiento de fase espacial sobre un conjunto representativo de seis objetos sintéticos, podemos observarlos en la Figura 23. Objetos utilizados en métodos tradicionales. Figura 23 (Obj-1 a Obj-6). Estos métodos constituyen la línea de base contra la cual se evaluó posteriormente el modelo propuesto de aprendizaje profundo.

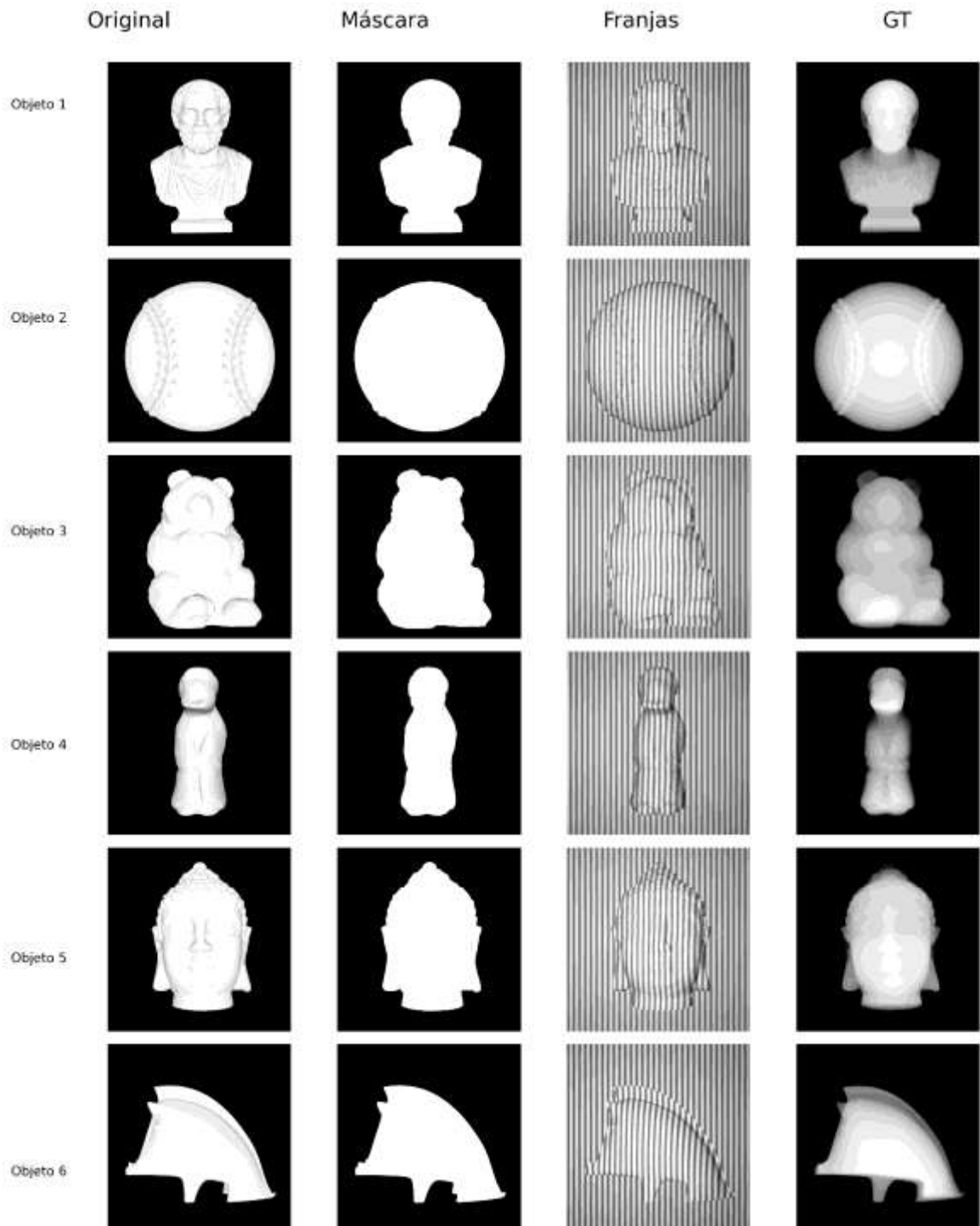


Figura 23. Objetos utilizados en métodos tradicionales.

Los algoritmos analizados fueron: Itoh, Herraez, Zhao, Constantini, Martínez-Carranza (STIEPU), Ghiglia–Romero, PUMA y PEARLS.

Todos los métodos se aplicaron a mapas de fase previamente obtenidos mediante perfilometría por Transformada de Fourier y normalizados dentro de la máscara binaria del objeto. El análisis combina resultados cualitativos y cuantitativos, empleando métricas normalizadas y reconstrucciones 2D y 3D. Todas las comparaciones se realizaron dentro de la máscara binaria del objeto, con normalización entre 0 y 1, de modo que las métricas reflejen exclusivamente diferencias de forma y no de escala.

6.2.1 Objetos evaluados

Aunque se evaluaron seis objetos sintéticos distintos, los resultados cualitativos se presentan únicamente para Obj 3, Obj 5 y Obj 6, debido a que representan escenarios con diferentes niveles de complejidad geométrica y diferentes grados de discontinuidad en el mapa de fase. Los seis objetos se incluyen en la discusión general para mantener la validez comparativa del estudio.

6.2.2 Resultados cualitativos

Para cada uno de los seis objetos se generaron reconstrucciones mediante los ocho algoritmos mencionados. La Figura 24 muestran los mapas de altura (2D) superpuestos con el contorno verdadero del objeto, junto con las reconstrucciones tridimensionales obtenidas a partir de las mismas fases desenvueltas.

El análisis visual permitió identificar una serie de patrones consistentes:

- **Itoh, Ghiglia, Zhao y Constantini** producen reconstrucciones suaves, continuas y sin artefactos severos. Las regiones con gradientes pronunciados son recuperadas con estabilidad y sin introducir ondulaciones o escalones falsos.
- **Herraez** presenta desempeño variable: en algunos objetos produce reconstrucciones limpias, pero en otros genera fallas en zonas planas amplias o en transiciones

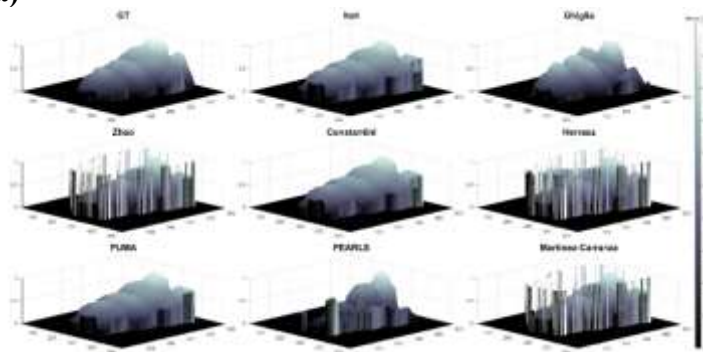
pronunciadas, donde aparecen discontinuidades verticales o lapsos de suavizado excesivo.

- **PUMA** exhibe comportamiento mixto. Aunque en ciertos objetos obtiene superficies coherentes, en otros introduce columnas verticales discontinuas y artefactos puntuales en regiones de bajo contraste.
- **PEARLS** y **Martínez-Carranza** muestran los resultados más inestables. Las reconstrucciones producidas por PEARLS contienen picos abruptos, artefactos extensos y zonas saturadas, mientras que Martínez-Carranza genera columnas verticales intensas que afectan gran parte del volumen reconstruido.

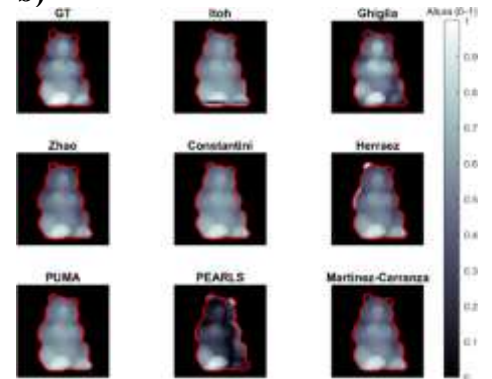
El conjunto de figuras 3D evidencia de forma clara la aparición de estos artefactos y facilita la comparación directa con la superficie ideal del ground truth.

Objeto 3

a)

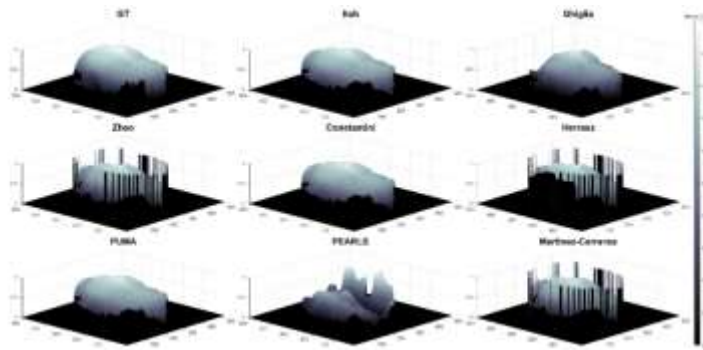


b)

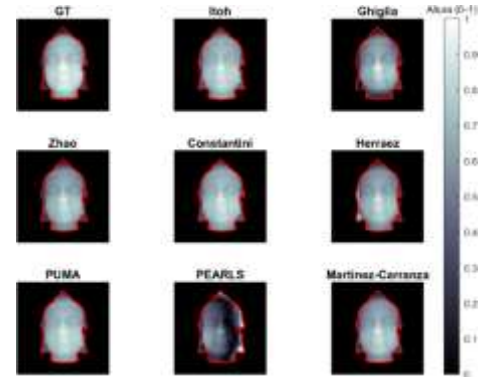


Objeto 5

c)

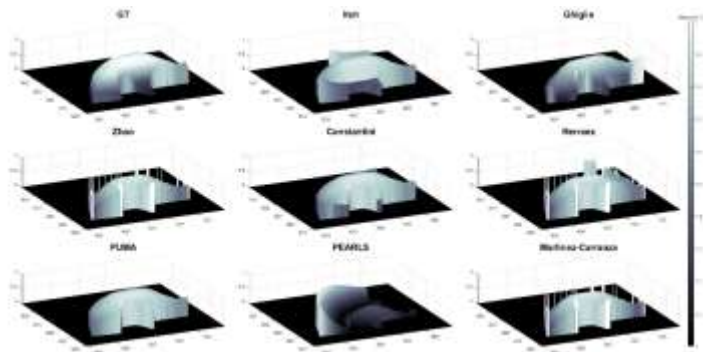


d)



Objeto 6

e)



f)

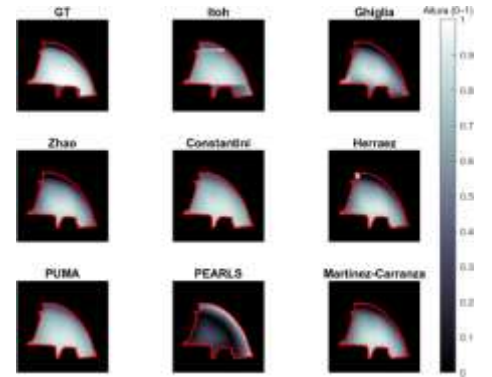


Figura 24. Reconstrucciones de los objetos 3, 5 y 6.

A la izquierda se muestran las superficies en formato mesh y a la derecha las vistas bidimensionales, donde se incluye en rojo el contorno definido por la máscara binaria utilizada para la evaluación de métricas.

6.2.3 Resultados cuantitativos

Para cada objeto y método se calculó el MSE, PSNR y SSIM, optimizando previamente el parámetro Δ para garantizar una comparación justa entre reconstrucciones.

Todas las métricas se calcularon dentro de la máscara del objeto, usando valores normalizados entre 0 y 1. La Figura 25 presentan los resultados por objeto para cada una de las tres métricas.

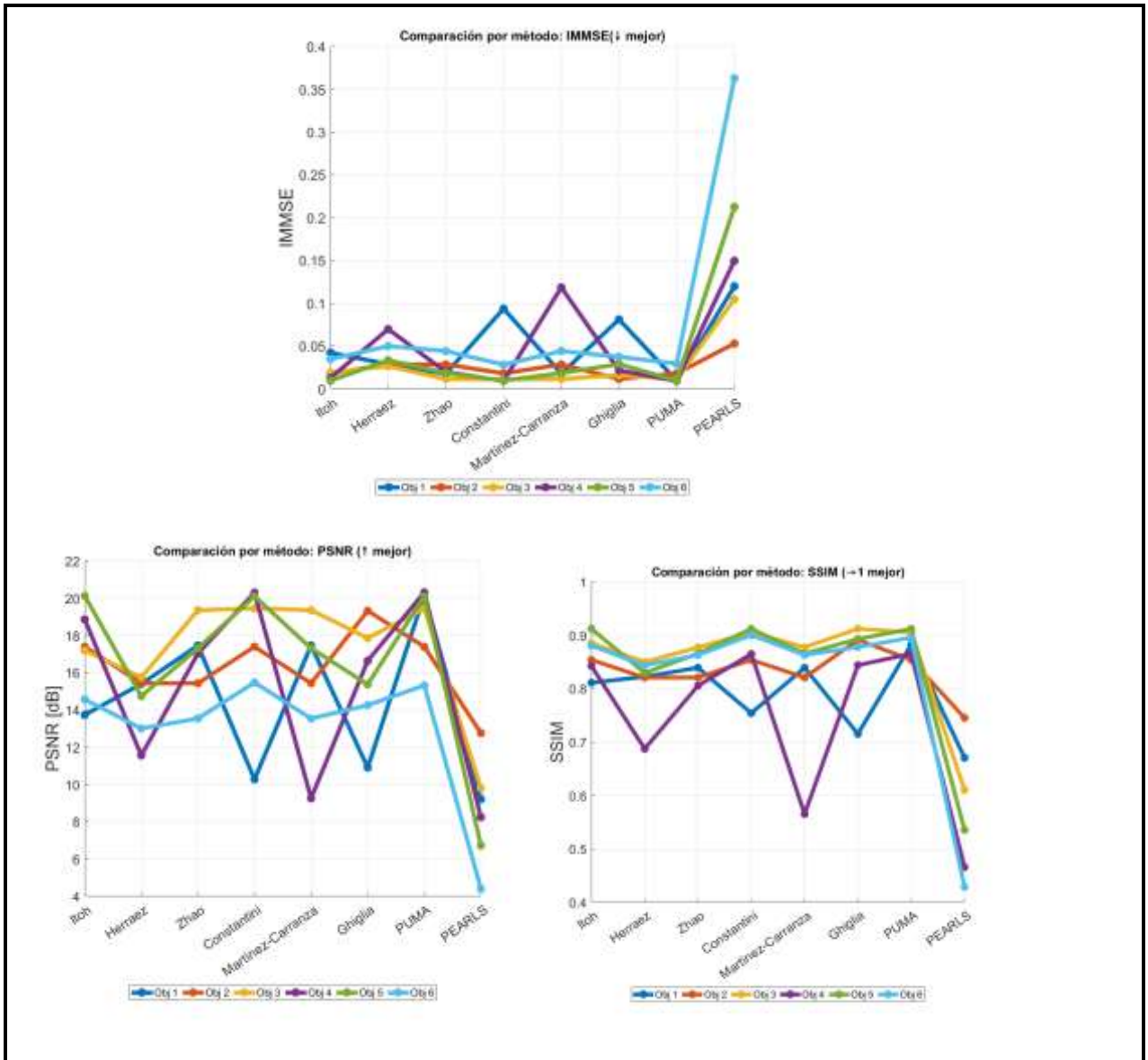


Figura 25. Gráficas comparativas para los objetos 3, 5 y 6. Gráficas comparativas de las métricas de evaluación (a) IMMSE, (b) PSNR y (c) SSIM obtenidas para los objetos 3, 5 y 6 en cada uno de los métodos de desdoblamiento de fase. Cada curva corresponde a un objeto y permite observar las diferencias de desempeño.

IMMSE (↓ mejor)

- Los valores más bajos corresponden a Itoh, Zhao, Constantini y Ghiglia, con errores pequeños y estables entre objetos.
- Herraiez, PUMA y Martínez-Carranza mantienen valores moderados.
- PEARLS presenta los peores resultados, con IMMSE significativamente mayor que el resto (hasta 0.35–0.37 en algunos objetos).

PSNR (↑ mejor)

- Los métodos con mejores PSNR son Itoh, Constantini, Zhao, Ghiglia y, en algunos objetos, PUMA.
- Martínez-Carranza y Herraiez obtienen valores medios.
- PEARLS nuevamente muestra los valores más bajos, indicando mayor presencia de ruido y discontinuidades.

SSIM (→1 mejor)

- Zhao, Ghiglia, Itoh y Constantini presentan los SSIM más altos y consistentes, con valores entre 0.85 y 0.95 en la mayoría de los casos.
- Herraiez y PUMA oscilan más entre objetos.
- PEARLS alcanza los valores más bajos, llegando a 0.43 en algunos casos.

Método	IMMSE promedio	IMMSE Desviación estándar	PSNR promedio	PSNR Desviación estándar	SSIM promedio	SSIM Desviación estándar
Martínez-Carranza	0.0399	0.0402	15.3922	3.5973	0.8052	0.1193
Constantini	0.0285	0.0328	17.1643	3.8462	0.8653	0.0592
Ghiglia	0.0329	0.0254	15.7247	2.9605	0.8557	0.0724
Herraez	0.0396	0.0171	14.3129	1.6706	0.8090	0.0606
Itoh	0.0229	0.0129	16.9759	2.4488	0.8644	0.0358
PEARLS	<u>0.1671</u>	0.1094	<u>8.5259</u>	2.8407	<u>0.5762</u>	0.1219
PUMA	0.0146	0.0080	18.7850	2.0181	0.8858	0.0229
Zhao	0.0235	0.0115	16.6875	1.9840	0.8453	0.0282

Tabla 3. Valores promedio Métodos tradicionales.

Valores promedio y desviación estándar de las métricas IMMSE, PSNR y SSIM obtenidas para los seis objetos sintéticos evaluados, considerando el delta óptimo en cada caso. Los valores en negritas indican el mejor desempeño en cada métrica, mientras que los valores subrayados y en cursiva corresponden al peor resultado.

El análisis numérico mostró tendencias claras:

- **PUMA** obtuvo consistentemente los valores más altos de PSNR y SSIM, así como los MSE más bajos.
- **Ghiglia, Zhao, Constantini y Martínez-Carranza** presentaron un rendimiento intermedio, dependiendo del objeto.
- **Itoh** mostró variabilidad significativa entre objetos, reflejando su sensibilidad a discontinuidades.
- **PEARLS**, incluso tras la normalización y alineación, presentó errores muy superiores al resto de los métodos (directamente el peor desempeño del conjunto).

La Tabla 3 resume los promedios por método, donde se observa que PUMA se establece como el mejor algoritmo tradicional para esta base de datos particular.

6.2.4 Desempeño según la geometría del objeto

La comparación por objeto revela tendencias importantes:

- **Métodos robustos a la geometría:** Itoh, Constantini, Zhao y Ghiglia mantienen desempeño estable incluso en superficies con concavidades, pendientes suaves o regiones amplias planas.
- **Métodos sensibles a la geometría:** Herraez y PUMA fallan con mayor frecuencia en objetos que contienen transiciones abruptas o zonas de bajo contraste.
- **Métodos altamente inestables:** PEARLS y Martínez-Carranza muestran degradaciones severas independientemente del objeto, indicando poca capacidad para generalizar entre geometrías distintas.

Esto demuestra que la evaluación debe considerar no solo el error promedio, sino también la estabilidad ante variaciones geométricas como se puede observar el desempeño de cada objeto en la Figura 25.

6.2.5 Discusión del desempeño

Los resultados muestran que el rendimiento de los algoritmos de desdoblamiento de fase está fuertemente condicionado por:

- la forma y continuidad del objeto,
- la distribución de saltos de fase,
- el nivel de wrap (low-wrap vs. multi-wrap),
- la estabilidad del método ante gradientes pronunciados

La comparación integral -que combina la inspección visual de los mapas 2D y reconstrucciones 3D, junto con el análisis cuantitativo mediante IMMSE, PSNR y SSIM- permite establecer tendencias claras sobre el comportamiento de cada algoritmo. En términos generales, los métodos de Itoh, Constantini, Zhao y Ghiglia demostraron ser los más consistentes y precisos. Estos algoritmos mantuvieron un desempeño estable en los seis objetos evaluados, reproduciendo de manera fiel los contornos, gradientes y detalles superficiales del ground truth. Además, alcanzaron los valores más favorables en las métricas

cuantitativas, con bajos IMMSE, PSNR elevados y altos SSIM. Su buen comportamiento se mantuvo incluso ante geometrías con transiciones suaves y superficies amplias, lo que evidencia una sólida estabilidad frente a variaciones en la morfología del objeto.

En contraste, los métodos de Herraiez, PUMA y Martínez-Carranza mostraron un desempeño intermedio. Si bien en algunos objetos lograron reconstrucciones aceptables, su estabilidad no fue uniforme. Por ejemplo, Herraiez y PUMA tendieron a fallar en regiones con pendientes pronunciadas o áreas planas extensas, produciendo columnas verticales o artefactos abruptos en los mapas de fase. Martínez-Carranza, aunque menos errático que PEARLS, también presentó dificultades recurrentes con discontinuidades que se reflejaron en valores más altos de IMMSE y una mayor variabilidad en el SSIM entre objetos.

Finalmente, el método PEARLS presentó el desempeño más deficiente. Las reconstrucciones fueron sistemáticamente inestables, con artefactos severos que afectaron tanto la continuidad espacial como la interpretación tridimensional de las superficies. Los valores cuantitativos confirmaron esta degradación, ya que mostró los IMMSE más altos y los PSNR más bajos entre todos los métodos probados. Este comportamiento se observó de forma consistente en los seis objetos, indicando que PEARLS no logró adaptarse adecuadamente a las características geométricas presentes en el conjunto de datos.

En conjunto, los resultados evidencian que, si bien los métodos tradicionales pueden ofrecer soluciones confiables en escenarios controlados, su desempeño puede variar significativamente según la morfología del objeto y la complejidad de la fase envuelta. Si bien varios métodos ofrecieron resultados aceptables en objetos simples, solo PUMA mantuvo consistencia entre todos los casos evaluados. Su desempeño superior motivó su uso como método de referencia en la etapa posterior de comparación con los modelos basados en redes neuronales convolucionales.

6.3 Modelos de aprendizaje profundo

Esta subsección presenta los resultados obtenidos al entrenar y evaluar diferentes arquitecturas de aprendizaje profundo diseñadas para el desdoblamiento espacial de fase. A diferencia de los métodos tradicionales, cuyo proceso depende explícitamente de gradientes locales, rutas de integración y supuestos sobre la continuidad de la fase, los modelos basados en redes neuronales buscan aprender una función de desdoblamiento directamente a partir de datos, capturando relaciones no lineales complejas entre la fase envuelta y la fase desenvuelta.

El análisis que se muestra a continuación surge de un proceso experimental progresivo que incluyó 22 entrenamientos organizados en cuatro familias arquitectónicas:

- 1) U-Net con módulo SQD-LSTM
- 2) U-Net Res-Net
- 3) Red Multiresolution
- 4) U-Net estándar.

En estos experimentos se variaron múltiples factores, tipo de mapa de fase utilizado (low-wrap o multi-wrap), tamaño de entrada (256 o 512 px), esquema de partición (object-wise o mixed-object), número de ejemplos y funciones de pérdida, con el objetivo de identificar configuraciones robustas, estables y capaces de generalizar tanto en datos sintéticos como en objetos reales.

Los resultados mostrados abarcan:

- **Curvas de entrenamiento y validación**, útiles para evaluar estabilidad y sobreajuste.
- **Métricas cuantitativas** (IM-MSE, PSNR, SSIM y test loss), calculadas sobre conjuntos de prueba independientes.
- **Comparación visual entre predicción y ground truth** en mapas de fase sintéticos.

- **Inferencia sobre objetos reales**, indispensable para valorar la transferencia al dominio físico.
- **Comparación cruzada entre arquitecturas**, destacando fortalezas y limitaciones de cada enfoque.

Finalmente, se incluye un análisis del **fine-tuning realizado con datos reales**, cuyo propósito fue mejorar la adaptación del mejor modelo hacia las distribuciones físicas de ruido e iluminación presentes en los experimentos de perfilometría.

6.3.1 Introducción a los primeros experimentos y evolución del modelo

Antes de desarrollar los resultados principales presentados en este capítulo, es importante contextualizar el proceso evolutivo que siguió el modelo de aprendizaje profundo, en especial durante las primeras etapas de experimentación. El trabajo inició utilizando exclusivamente la arquitectura U-Net con SQD-LSTM, seleccionada por su capacidad para capturar dependencias espaciales en cuatro direcciones y mejorar la coherencia estructural del mapa de fase desenvuelta. Este modelo contaba con pesos pre-entrenados provenientes de un estudio previo, basado en mapas de fase sintéticos sencillos (sin geometría 3D real), generados a partir del método de tres pasos (3-Step).

6.3.1.1 Primeras inferencias con el modelo pre-entrenado

Cuando este modelo se aplicó por primera vez a objetos sintéticos más complejos, con geometría realista y mapas de fase generados correctamente, los resultados fueron deficientes, Figura 26: reconstrucciones con pérdida de forma, regiones saturadas y artefactos significativos. Esta discrepancia se debía principalmente a la incompatibilidad entre el tipo de data utilizada en el pre-entrenamiento y la nueva base de datos, la cual presentaba gradientes más pronunciados, estructuras tridimensionales definidas y distribuciones estadísticas más amplias.

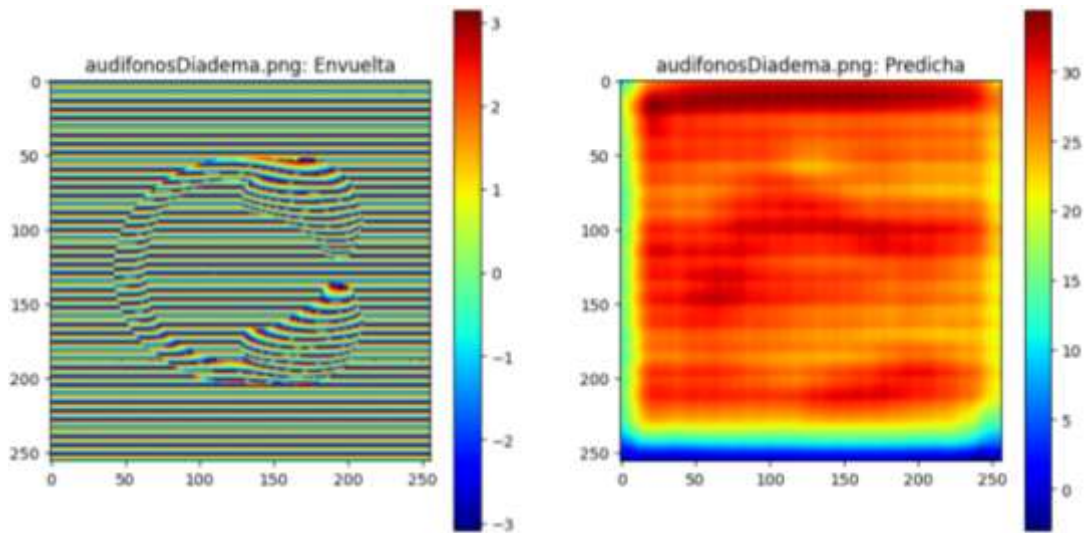


Figura 26. Ejemplo de las primeras inferencias de los objetos sintéticos. El modelo con arquitectura U-net con bloque LSTM entrenado con los mapas de datos sencillos creados de manera sintética y como fue utilizado en (Perera & de Silva, 2021) .

Estas primeras pruebas motivaron la necesidad de realizar un proceso sistemático de ajuste fino (fine-tuning) y de evaluar diferentes estrategias de normalización y funciones de pérdida.

6.3.1.2 Ciclo inicial de experimentos de fine-tuning

El primer conjunto de experimentos se realizó utilizando la base de datos sintética inicial, la cual no contaba con un *ground truth* (GT) completamente adecuado para entrenamiento supervisado, debido a que el material spline empleado en esta primera versión no describía correctamente la distribución de alturas. Aun así, se ejecutaron pruebas preliminares de *fine-tuning* con el objetivo de evaluar la capacidad del modelo para producir reconstrucciones a partir de un conjunto reducido de ejemplos.

BD sintética inicial

- Entrada: Fase envuelta obtenida mediante el método de 3 pasos.
- Salida: Imagen del objeto original (como aproximación de GT).

Configuración inicial de *fine-tuning*

- Modelo preentrenado con mapas de fase sintéticos.

- 50 pares de imágenes.
- 50 épocas.
- *Batch size*: 8.
- Optimizador: Adam.
- Sin normalizar los mapas de fase de entrada.

Los resultados visuales de estas primeras pruebas se muestran en la Figura 27, donde se observa el mapa de fase envuelta, el GT empleado y la predicción del modelo para dos objetos diferentes.

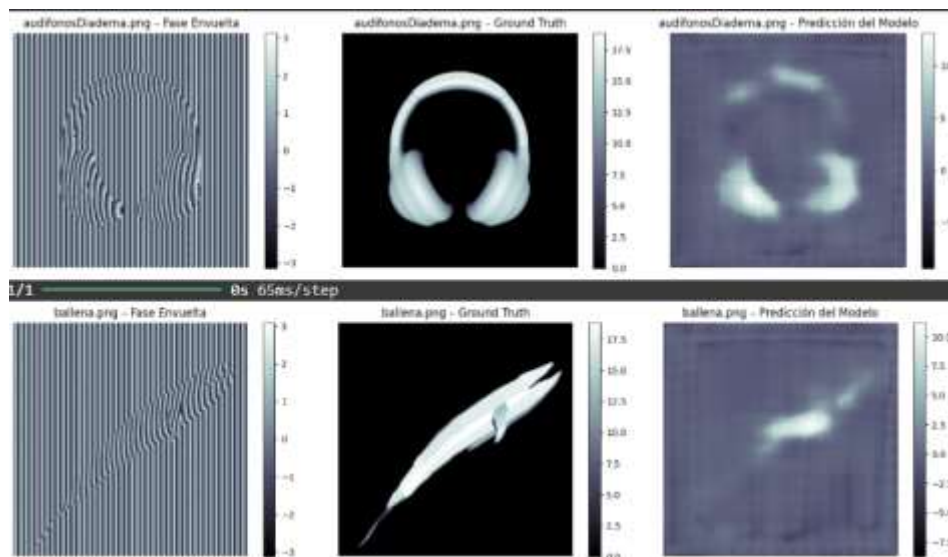


Figura 27. Resultados visuales de los entrenamientos del fine-tuning. La figura muestra los resultados del primer ciclo de fine-tuning utilizando únicamente 50 mapas de fase y un GT aproximado. Para cada objeto se presenta: (i) el mapa de fase envuelta de entrada, (ii) el GT empleado y (iii) la predicción del modelo.

Estos resultados, aunque mejores que los obtenidos al inicio del proyecto (antes de disponer de GT generado con spline), evidencian que el modelo no logra recuperar detalles finos ni la geometría real de los objetos. La reconstrucción es difusa, con pérdida significativa de contornos y ambigüedad en las zonas de pendiente suave. Esto se atribuye principalmente a:

- Escasez de datos: solo 50 mapas de fase.
- GT insuficiente o poco representativo.

- Variabilidad geométrica limitada.

Normalización global del GT y ampliación a 50 mapas de fase

En un segundo experimento, el GT se normalizó de forma global a $[0,1]$, empleando el valor máximo entre todos los objetos. Con esta normalización, se realizó nuevamente un *fine-tuning* con los mismos 50 mapas de fase.

La Figura 28 muestra los resultados para este escenario. Aunque la predicción mejora en términos de contraste respecto a la primera configuración, la forma del objeto continúa siendo poco definida y persisten artefactos visibles. La reconstrucción no corresponde fielmente al GT, evidenciando nuevamente la insuficiencia de datos y la falta de una normalización más adecuada.

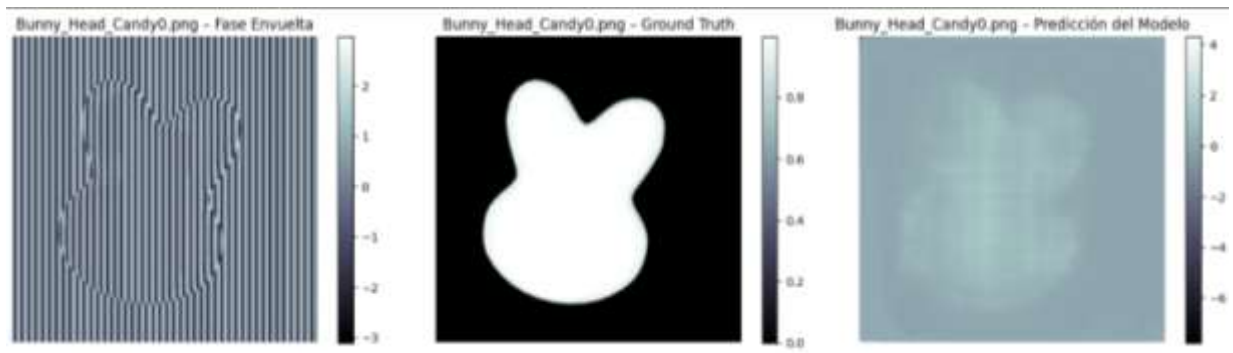


Figura 28. Resultados del *fine-tuning* con GT normalizado globalmente y 50 mapas de fase. Se presentan los resultados obtenidos tras aplicar una normalización global del GT a $[0,1]$ y realizar nuevamente el *fine-tuning* con 50 mapas de fase. La figura incluye el mapa de fase envuelta, su correspondiente GT y la predicción generada por el modelo

Normalización por imagen y aumento a 247 mapas de fase

Para un tercer experimento, se adoptó una estrategia más consistente:

- Normalizar cada GT individualmente a $[0,1]$.
- Utilizar 247 mapas de fase.
- Entrenar durante 100 épocas.

Los resultados se muestran en la Figura 29, donde se observa una mejora notable en la coherencia global de la forma, pero aún con pérdida de detalle. La predicción es más estable,

con menor ruido y una mejor correspondencia espacial, pero todavía insuficiente para alcanzar una reconstrucción precisa.

Este comportamiento confirmó que el modelo sí respondía positivamente al aumento de datos, pero que el conjunto seguía siendo limitado y el GT continuaba careciendo de fidelidad geométrica.

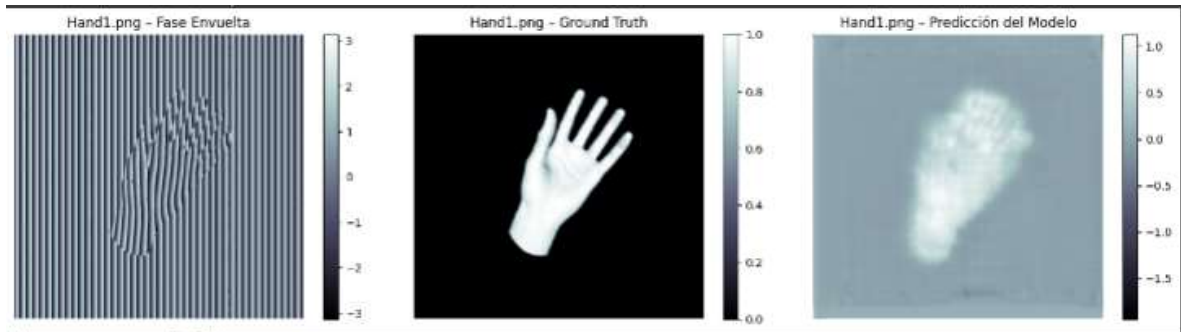


Figura 29. Resultados del fine-tuning con normalización por imagen y 247 mapas de fase. Compara el mapa de fase envuelta, el GT normalizado individualmente y la salida predicha por el modelo tras entrenar con 247 mapas de fase y 100 épocas.

Actualización completa de la base de datos

A partir de los resultados anteriores, se determinó que era necesario:

- a) Reemplazar los mapas de fase envuelta por versiones reales generadas con proyecciones de 3 pasos (*low-wrap*), obteniendo patrones físicamente consistentes.
- b) Generar un GT correcto usando materiales spline adecuados, ajustados para producir mapas de altura continuos y sin distorsiones internas (Figura 30).

Esta nueva versión de la BD permitió posteriormente:

- Entrenar modelos desde cero.
- Realizar *fine-tuning* con datos más diversos y mejor estructurados.
- Asegurar coherencia entre fase, máscara, franja y GT.

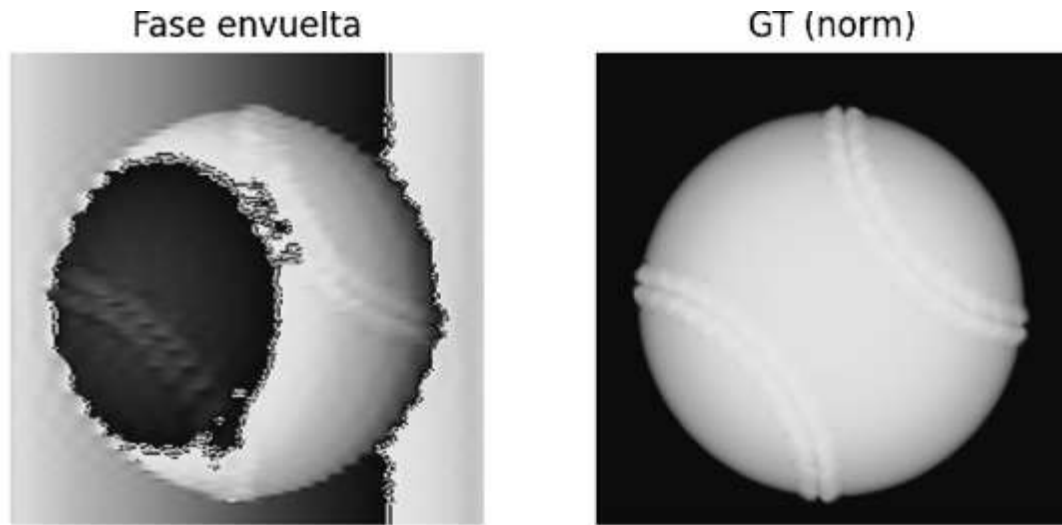


Figura 30. Mapas de fase low-wrap y GT generado mediante spline. La figura ilustra la nueva versión del dataset tras la actualización completa del pipeline: (i) el nuevo mapa de fase envuelta obtenido mediante proyección real de franjas (low-wrap), y (ii) el GT generado con material spline corregido, normalizado a $[0,1]$.

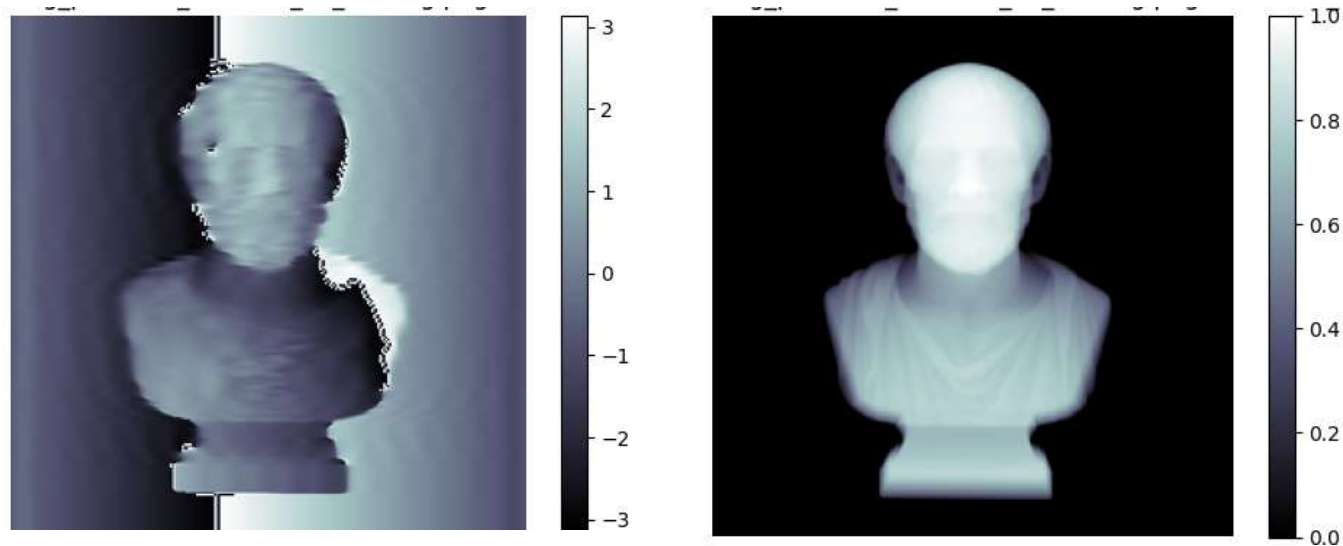
6.3.1.3 Experimentos fine-tuning con la base de datos actualizada

A lo largo de cinco experimentos consecutivos, descritos en detalle en la Tabla 4 que se complementa con la Figura 31 que corresponde a las imágenes de las predicciones de los entrenamientos del modelo, correspondiente, se fueron analizando distintos elementos críticos del entrenamiento:

- Normalización por imagen vs. normalización global.
- Métodos de generación de GT (Spline vs. FFT).
- Funciones de pérdida: MAE simple, pérdida compuesta (MAE + GradVar), penalización de salida fuera del intervalo $[0,1]$.
- Congelamiento parcial de capas pre-entrenadas.
- Variación de batch size y learning rate.
- Evaluación por estabilidad del entrenamiento y reducción de artefactos.

No. Experimento	Hiperparámetros	Datos de entrenamiento (Base de datos)	Resultados	Próximo paso
1	Épocas: 50 + 50 (2 etapas); Batch size = 8 → 4; LR = 1e-4; Función de pérdida: MAE; sin normalización inicial	50 pares de mapas de fase sintéticos (3-Step)	loss: 0.0524; mse: 0.0129; val_loss: 0.1265; val_mse: 0.0543	Probar normalización por imagen
2	Épocas: 100; Batch size = 8; LR = 1e-4; Función de pérdida: MAE; normalización por imagen	247 mapas de fase normalizados individualmente	Mejora en estabilidad de entrenamiento; reducción de saltos de fase ficticios	Evaluar normalización global
3	Épocas: 100; Batch size = 8; LR = 1e-4; Función de pérdida: MAE; normalización global en [0, 1]	50 mapas de fase sintéticos (FFT)	Reconstrucciones más uniformes; menor ruido de alta frecuencia	Ampliar dataset con mapas FFT (> 200)
4	Épocas: 50 + 50; Batch size = 4; LR = 3e-5 → 1e-4; Función de pérdida compuesta (MAE + GradVar); Early Stopping	256 pares de datos (mapas FFT + GT Spline)	loss: 0.0110; mse: 0.0284; val_loss: 0.0462; val_mse: 0.0999; mejor equilibrio PSNR/SSIM	Inferencia con objetos reales y nuevas métricas
5	Épocas: 50 + 50; Batch size = 4; LR = 3e-5; Función de pérdida compuesta + MAE; salida lineal con penalización [0, 1]	256 pares (FFT)	loss: 0.0308; mse: 0.1193; val_loss: 0.0455; val_mse: 0.224; PSNR ↑ ≈ 32 dB; SSIM ≈ 0.93	Entrenar desde cero con más mapas de fase y evaluar en datos reales

Tabla 4. Resultados de fine-tuning con el modelo pre-entrenado.



Mapa de fase entrada

Ground truth real del objeto sintético

a)

b)

c)

d)

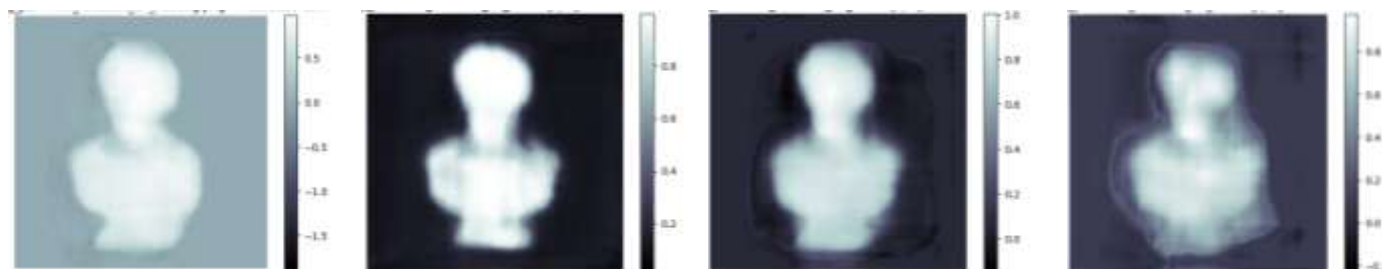


Figura 31. Resultados visuales de los entrenamientos del fine-tuning. Mapa de fase de entrada, ground truth normalizado y predicciones obtenidas con las cuatro configuraciones de fine-tuning.

Las observaciones clave fueron las siguientes:

- La normalización por imagen mejoró notablemente la estabilidad, eliminando saltos de fase ficticios.
- La normalización global en $[0,1]$ redujo el ruido de alta frecuencia en reconstrucciones con datos FFT.

- La pérdida compuesta logró el mejor equilibrio entre suavidad y preservación de bordes.
- El uso de una salida lineal con penalización evitó valores negativos o mayores a uno, coherentes con la naturaleza normalizada de la fase.
- Los experimentos con 256 pares de datos (FFT + GT Spline) mostraron reconstrucciones más limpias y una mejora en PSNR/SSIM.

Los resultados visuales mostrados en la Figura 31 corresponden a las distintas configuraciones evaluadas durante el proceso de fine-tuning. Aunque en comparación con los primeros experimentos, aquellos realizados sin mapas ground truth generados mediante materiales spline, se observa una mejora moderada en la coherencia global de las reconstrucciones, las predicciones siguen presentando limitaciones importantes. En todas las variantes (a–d) la red logra aproximar la forma general del objeto, pero persisten pérdidas de detalle, suavizamientos excesivos y distorsiones que impiden obtener una reconstrucción fiel respecto al ground truth real.

Estas deficiencias indican que el modelo, entrenado únicamente mediante fine-tuning sobre un conjunto reducido de mapas de fase, no cuenta con la variabilidad ni la cantidad de ejemplos necesaria para aprender adecuadamente la relación superficie–fase. A partir de esta observación se determinó que continuar ajustando el modelo sobre tan pocos datos no sería suficiente. Por ello, se decidió ampliar la base de datos sintética y proceder con un entrenamiento completo desde cero, buscando mejorar la capacidad de generalización y la fidelidad estructural de las reconstrucciones.

A pesar de las mejoras graduales, los modelos afinados en esta etapa seguían mostrando un límite: el GT utilizado en los primeros conjuntos no correspondía a una altura realista, por lo que el modelo ajustaba correctamente la fase sintética, pero no lograba proyectar una representación físicamente coherente en inferencias nuevas.

Esto validó la necesidad de:

- 1) Ampliar las bases de datos sintéticas con mapas generados mediante Transformada de Fourier (FFT) y geometrías variadas.

- 2) Redefinir el proceso de normalización.
- 3) Entrenar desde cero con bases más amplias (más de 500 mapas por arquitectura).
- 4) Posteriormente, realizar un fine-tuning final únicamente con datos reales, lo cual se analiza en secciones posteriores.

Este conjunto de experimentos iniciales permitió establecer qué configuraciones eran más estables, qué normalizaciones evitaban artefactos y qué pérdida preservaba mejor la estructura. Gracias a esta etapa exploratoria, el resto del trabajo pudo desarrollarse con bases mucho más sólidas, tanto para entrenar los modelos finales como para realizar el fine-tuning específico para inferencia en objetos reales.

6.3.2 Comparación de los entrenamientos

La Tabla 5 resume todos los experimentos realizados. Cada fila representa una configuración distinta, incluyendo:

- Arquitectura del modelo.
- Tipo de mapa de fase (low-wrap o multi-wrap).
- Estrategia de partición del dataset (object-wise split o mixed-object split).
- Tamaño espacial (256 px o 512 px).
- Cantidad de ejemplos.
- Número final de épocas.
- Métricas Test Loss, PSNR, SSIM e IM-MSE.

ID	Modelo	Tipo de mapa de fase	Data	Tam	Cant.	Epoch	Test Loss	PSNR	SSIM	IMMSE	Observaciones
V1	U-Net con SQD-LSTM	Low-wrap	Mixed-object split	256	999	100	0.0042	27.35	0.853	0.0030	
V2	U-Net con SQD-LSTM	Low-wrap	Mixed-object split	512	999	120	0.0035	27.62	0.839	0.0027	
V3	U-Net con SQD-LSTM	Low-wrap	Object-wise split	256	972	100	0.0030	28.76	0.894	0.0013	
V4	U-Net con SQD-LSTM	Multi-wrap	Object-wise split	256	1593	100	0.0038	27.74	0.878	0.0024	
V5	U-Net con SQD-LSTM	Low-wrap	Object-wise split	256	1593	100	0.0041	27.30	0.903	0.0028	
V6	U-Net con SQD-LSTM	Low-wrap	Mixed-object split	512	1566	120	0.0034	27.70	0.908	0.0026	
V7	U net -Res net	Low-wrap	Mixed-object split	256	999	100	0.0030	28.70	0.926	0.0020	
V8	U net-Res net	Low-wrap	Object-wise split	256	972	100	0.0032	27.61	0.895	0.0022	
V9	U net-Res net	Multi-wrap	Object-wise split	256	1593	100	0.0060	26.76	0.871	0.0046	
V10	U net-Res net	Low-wrap	Object-wise split	256	1593	100	0.0051	26.80	0.887	0.0037	
V11	U net-Res net	Low-wrap	Mixed-object split	512	1566	120	0.0078	23.96	0.793	0.0056	Earlystopping, 30.W. restaurados : E 20
V11.5	U net-Res net	Low-wrap	Mixed-object split	512	1566	100	0.0053	25.26	0.864	0.0041	
V12	Multiresolution	Low-wrap	Mixed-object split	256	999	100	0.0046	26.09	0.875	0.0031	Earlystopping: E 60. Val loss
V12.5	Multiresolution	Low-wrap	Mixed-object split	256	999	100	0.0040	26.53	0.878	0.0025	Early stopping: E 47. Loss
V13	Multiresolution	Low-wrap	Object-wise split	256	972	100	0.0038	26.39	0.878	0.0027	Earlystopping: E 51. Val loss
V13.5	Multiresolution	Low-wrap	Object-wise split	256	972	100	0.0049	24.88	0.823	0.0036	Earlystopping: E 61. Loss.
V14	Multiresolution	Multi-wrap	Object-wise split	256	1593	100	0.0069	26.01	0.846	0.0055	Earlystopping: E 74. Loss
V15	Multiresolution	Low-wrap	Object-wise split	256	1593	100	0.0072	24.29	0.846	0.0057	Earlystopping: E 71. Loss
V16	Multiresolution	Low-wrap	Mixed-object split	512	1566	120	0.0058	24.47	0.866	0.0044	Earlystopping: E 89. Loss.
V17	U net	Low-wrap	Mixed-object split	256	999	120	0.0017	31.88	0.959	0.0009	
V18	U net	Low-wrap	Object-wise split	256	972	100	0.0030	27.9	0.897	0.0018	
V19	U net	Multi-wrap	Object-wise split	256	1593	100	0.0043	28.13	0.919	0.0027	
V20	U net	Low-wrap	Object-wise split	256	1593	100	0.0049	27.45	0.920	0.0033	
V21	U net	Low-wrap	Mixed-object split	512	1566	120	0.0017	30.36	0.952	0.0012	

Tabla 5. Resultados comparativos de los entrenamientos con diferentes modelos.

La tabla presenta el rendimiento obtenido por los diferentes modelos entrenados desde cero, compuesta por mapas de fase low-wrap y multi-wrap, junto con ground truth generado mediante material spline. Se evaluaron cuatro arquitecturas principales —U-Net con SQD-LSTM, U-Net-ResNet, Multiresolution y U-Net básica— bajo dos esquemas de partición del dataset (object-wise split y mixed-object split) y dos

Al analizar los resultados presentados en la Tabla 5 se identifican tres tendencias generales. En primer lugar, el tamaño de entrada tiene un impacto directo en la calidad de reconstrucción: los modelos entrenados con imágenes de 512×512 alcanzan sistemáticamente valores más altos de PSNR y SSIM que sus versiones en 256×256 , especialmente en los escenarios *low-wrap*, donde la preservación de detalles espaciales es crítica.

En segundo lugar, la estrategia de partición object-wise split demuestra una mayor capacidad de generalización. Aunque *mixed-object split* tiende a producir pérdidas menores durante el entrenamiento, *object-wise split* permite que los modelos reconozcan patrones en objetos no vistos, obteniendo métricas más estables y mejor desempeño en inferencias reales.

Finalmente, se observa que las arquitecturas más profundas no siempre son las más efectivas. A pesar de que modelos como U-Net con SQD-LSTM o U-Net-ResNet muestran estabilidad, el experimento V17 (U-Net estándar, 256 px, object-wise split) obtuvo las mejores métricas globales ($\text{IMMSE} \approx 0.0005$, $\text{PSNR} \approx 32.5$ dB, $\text{SSIM} \approx 0.94$). Esto evidencia que, con una base de datos bien estructurada y un esquema de particionado adecuado, una arquitectura simple puede superar alternativas más complejas.

Cabe destacar que la tabla utiliza codificación de colores, lo que facilita identificar de forma inmediata las configuraciones más y menos efectivas: verde para los mejores valores, amarillo para desempeño intermedio y rojo para los resultados más bajos.

6.3.3 Desempeño por arquitectura

A partir de los resultados globales presentados en la Tabla 5, es posible profundizar en el comportamiento individual de cada arquitectura con el fin de identificar sus fortalezas, limitaciones y patrones de generalización. Si bien la comparación agregada entre experimentos permite reconocer tendencias generales, como la influencia del tamaño espacial, el tipo de mapa de fase o la estrategia de particionado, un análisis desglosado por arquitectura resulta esencial para comprender cómo cada diseño interno procesa la información y qué configuraciones producen mejor rendimiento.

En esta sección se examina el desempeño de las cuatro familias de modelos empleadas en el estudio: U-Net con SQD-LSTM, U-Net Res-Net, arquitectura Multiresolution y U-Net estándar. Para cada una de ellas se presentan tanto métricas cuantitativas (train loss, validation loss, PSNR, SSIM e IM-MSE) como evaluaciones cualitativas basadas en predicciones reales. Asimismo, se discute la estabilidad durante el entrenamiento, la capacidad de generalización hacia objetos no vistos y la coherencia estructural de las reconstrucciones.

Este análisis permite no solo identificar qué configuraciones alcanzaron los mejores resultados numéricos, sino también evaluar la pertinencia de los mecanismos adicionales — como memoria recurrente, bloques residuales o escalas múltiples, frente a condiciones variables del dataset. Con ello se establecen las bases para seleccionar los modelos más prometedores y orientar las fases de fine-tuning y entrenamiento final desde cero.

6.3.3.1 *U-Net con SQD-LSTM*

Esta variante incorpora un bloque LSTM espacial profundo que permite conservar información contextual en regiones discontinuas. Dentro de esta familia se realizaron seis experimentos (V1–V6), cada uno entrenado bajo combinaciones distintas de resolución, tipo de mapa de fase y particionado del dataset.

El modelo destacado fue V6, entrenado con mapas *low-wrap*, resolución de 512 px y un total de 1566 ejemplos. Obtuvo: PSNR = 27.70 dB, SSIM = 0.908, IMMSE = 0.0026.

Cualitativamente, V6 produce reconstrucciones suaves y estables, en especial en regiones con gradientes suaves. Aunque no representa el mejor rendimiento cuantitativo dentro de la tabla general, sí mostró un comportamiento robusto al momento de realizar inferencia sobre datos reales (véase sección correspondiente), por lo que se seleccionó como base para el último ciclo de *fine-tuning* antes del entrenamiento completo desde cero.

Comportamiento durante el entrenamiento

La

Figura 32 muestra las curvas de convergencia de los seis modelos (V1–V6). En todos los casos, la pérdida de entrenamiento disminuye rápidamente durante las primeras épocas, lo que indica que la arquitectura U-Net con SQD-LSTM se adapta con facilidad a los patrones estructurales del conjunto sintético. Sin embargo, se observan fluctuaciones notables en la pérdida de validación, especialmente en V2, V4 y V6, reflejando la sensibilidad del modelo a variaciones en el particionado y al tipo de mapa de fase utilizado. Este comportamiento

oscilatorio sugiere que el LSTM captura dependencias espaciales útiles, pero también puede sobre-ajustarse cuando la distribución de validación difiere de la de entrenamiento.

Evaluación cualitativa mediante predicciones

La

Figura 33 presenta ejemplos visuales de predicción para cada modelo de la familia (V1–V6). Es importante notar que no se emplea el mismo objeto en todos los casos, ya que los conjuntos de entrenamiento, validación y prueba fueron construidos con bases de datos diferentes. En algunos casos, el objeto ya había sido utilizado en entrenamiento, y en otros no.

Por ello, esta figura no está destinada a comparar el mismo objeto entre modelos, sino a evaluar: La fidelidad de la reconstrucción respecto al ground truth, la coherencia estructural obtenida por cada modelo, y la estabilidad del modelo ante distintas geometrías.

Los mapas se muestran utilizando un colormap (escala pseudocolor) siguiendo la convención utilizada en el artículo original que propone la arquitectura U-Net LSTM; los colores no representan colores reales, sino valores continuos de altura o fase normalizada, de forma que permiten visualizar claramente la estructura del campo de fase reconstruido.

En general, los modelos V4, V5 y V6 presentan predicciones más limpias y con menor ruido, mientras que V1-V3 tienden a generar texturas espurias o suavizados excesivos cuando el mapa de entrada tiene gradientes pronunciados o ruido estructurado.

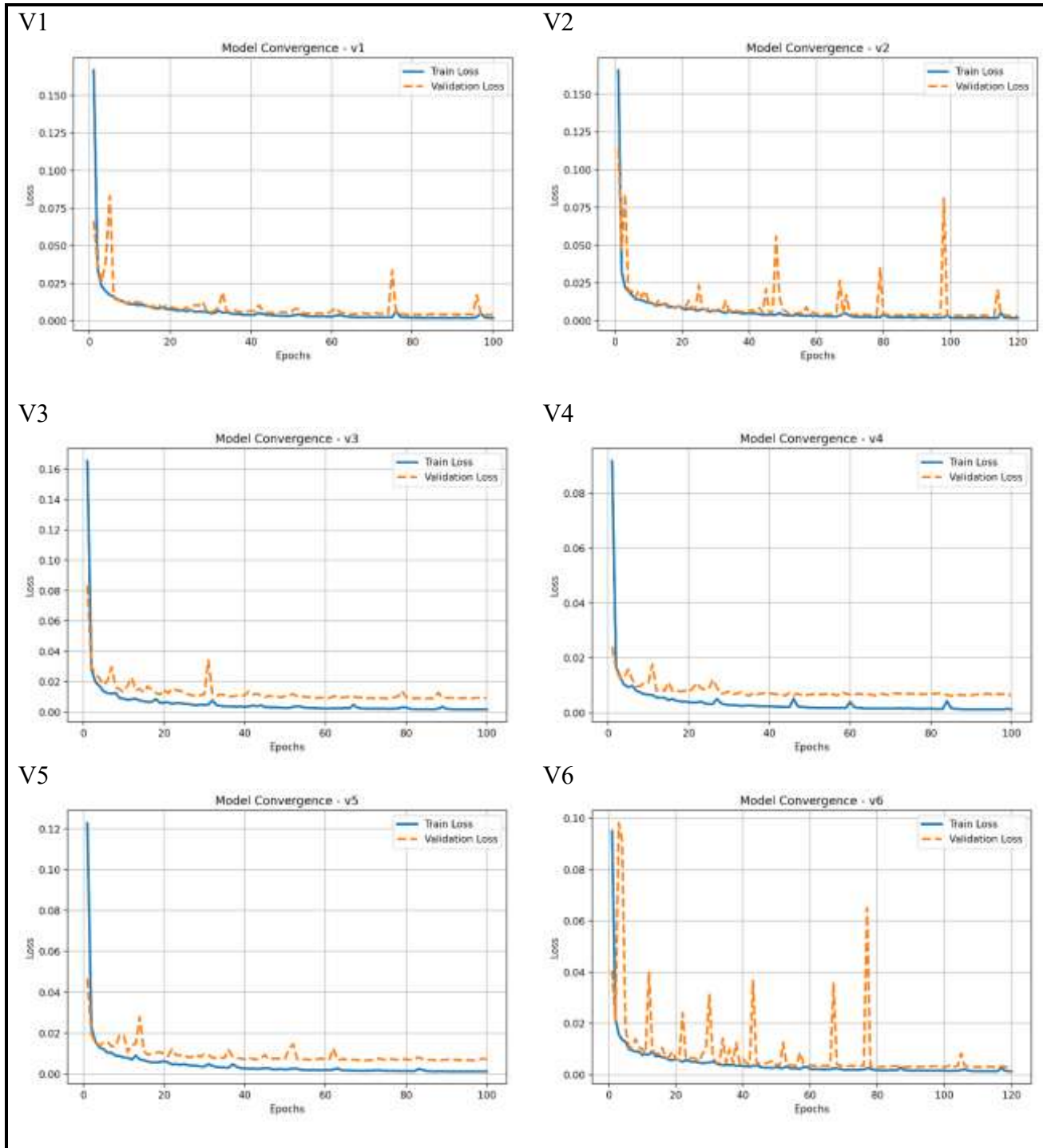


Figura 32. Gráficas de entrenamientos de U-Net con SQD-LSTM.

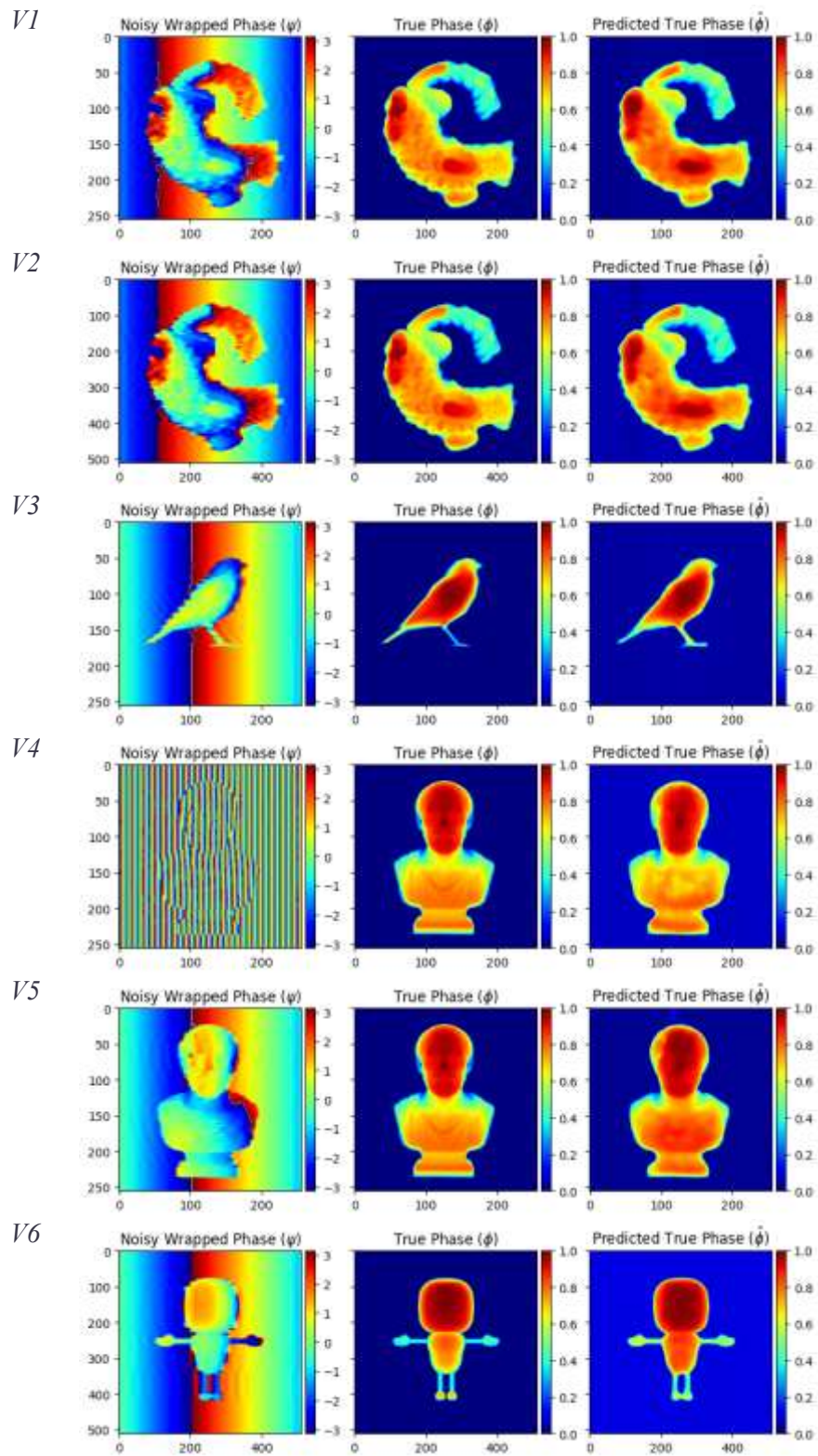


Figura 33. Predicciones de los modelos de la arquitectura U-Net con SQD-LSTM.

6.3.3.2 *U-Net Res-Net*

Los modelos V7-V11.5 incorporan bloques residuales dentro de la arquitectura U-Net con el objetivo de mejorar el flujo de gradientes y capturar variaciones espaciales más complejas. Aunque se esperaba una mejora consistente respecto a la U-Net clásica, los resultados muestran un comportamiento mixto: algunos experimentos presentan avances claros, mientras que otros evidencian inestabilidad.

En términos globales, esta arquitectura contiene uno de los experimentos con peor desempeño cuantitativo (V11.5), pero al mismo tiempo, cuando se consideran todas las configuraciones, es la segunda arquitectura con mejor desempeño promedio, lo cual puede apreciarse en la Tabla 5 gracias a la codificación por colores que permite identificar qué versiones fueron más robustas.

Un caso representativo es V7 (*low-wrap*, 256 px, *object-wise split*), que alcanzó métricas competitivas (PSNR = 27.61 dB, SSIM = 0.895). En contraste, otros modelos como V11 y V11.5 mostraron oscilaciones marcadas en la pérdida de validación, evidenciando sensibilidad a los datos multi-wrap y a la distribución estadística entre entrenamiento y validación.

Comportamiento durante el entrenamiento

La

Figura 34 muestra las curvas de convergencia de los seis modelos (V7–V11.5). En todos los casos, la pérdida de entrenamiento disminuye de manera estable, pero la pérdida de validación presenta picos pronunciados, especialmente en V8, V10 y V11.5. Este comportamiento indica que los bloques residuales facilitan el aprendizaje profundo, pero también pueden amplificar discrepancias entre los subconjuntos del dataset cuando existen variaciones fuertes en el ruido o en la estructura del mapa de fase.

Evaluación cualitativa mediante predicciones

La Figura 35 presenta ejemplos de predicción para cada modelo. Al igual que en la arquitectura anterior, no se empleó el mismo objeto para todos los casos debido a que los

conjuntos de entrenamiento, validación y prueba provienen de bases de datos distintas. Por ello, la comparación no busca analizar un mismo objeto, sino evaluar la fidelidad entre predicción y ground truth bajo diferentes formas, texturas y variaciones de fase.

Se utiliza nuevamente un colormap pseudocolor, siguiendo el estándar de la literatura, para representar variaciones de altura o fase normalizada; los colores no son atributos del objeto, sino una forma de visualizar gradientes y discontinuidades.

Cualitativamente, los modelos V7, V9 y V10 muestran mejores aproximaciones al ground truth, mientras que V11 y V11.5 generan reconstrucciones más ruidosas o suavizadas, consistentes con su comportamiento en validación. Esto refuerza la conclusión de que la arquitectura residual puede ser poderosa, pero su desempeño está fuertemente condicionado por la uniformidad estadística del dataset y el tipo de mapa de fase utilizado.

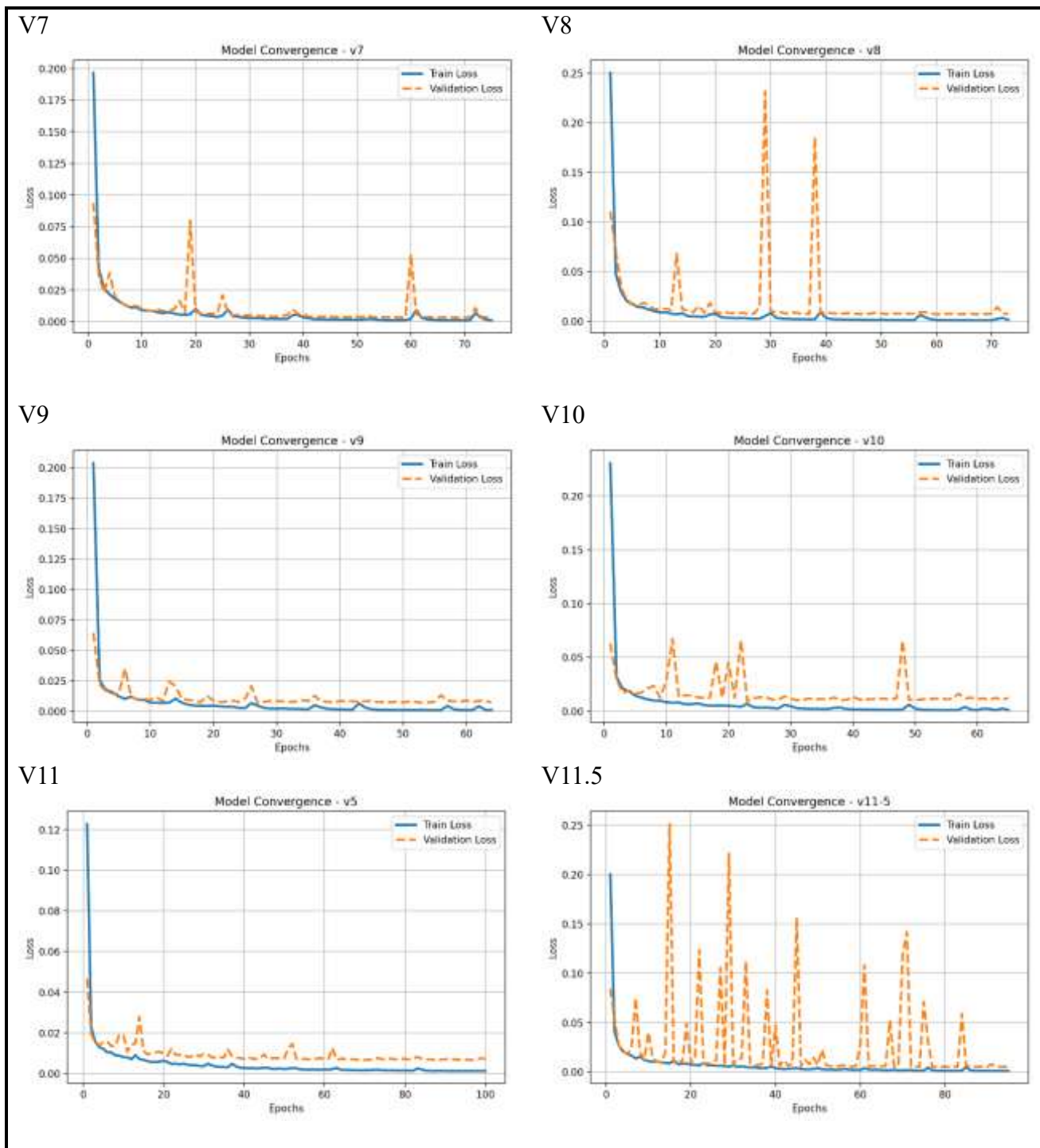


Figura 34. Gráficas de los entrenamientos de la arquitectura U-Net Res-Net.

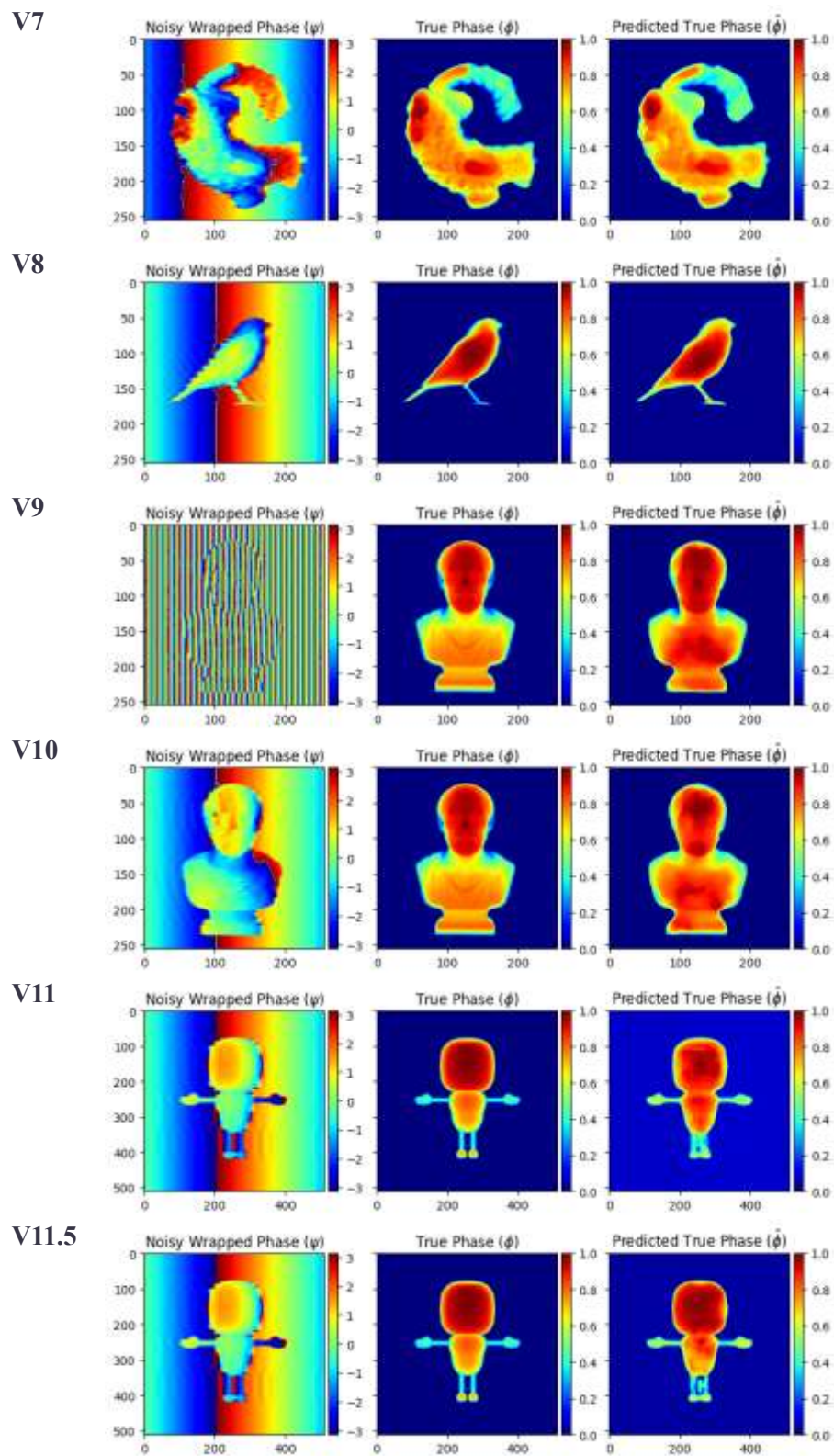


Figura 35. Predicciones de los modelos de la arquitectura U-Net Res-Net.

6.3.3.3 *Arquitectura Multiresolution*

La arquitectura Multiresolution (experimentos V12-V16) incorpora múltiples escalas internas de representación con el objetivo de capturar simultáneamente variaciones globales de fase y detalles locales. Este diseño es atractivo para problemas de desdoblamiento, pero su desempeño final resultó heterogéneo, dependiendo fuertemente del tipo de mapa de fase, el tamaño de entrada y la forma en que se particionó el dataset.

En la Tabla 5, los modelos multiresolución muestran un rango amplio de métricas. De todos ellos, los modelos V12, V12.5 y V13 mostraron el mejor equilibrio entre PSNR, SSIM e IM-MSE, aunque ninguno superó a las arquitecturas U-Net o U-Net Res-Net.

Convergencia del entrenamiento

La **¡Error! No se encuentra el origen de la referencia.** muestra las curvas de convergencia para los modelos V12–V16. Todos los modelos descienden rápidamente en pérdida durante las primeras épocas, pero se observan patrones importantes:

- La validación presenta picos periódicos, indicando que la arquitectura es sensible a cambios en la distribución del batch (especialmente en V12.5 y V14).
- En modelos como V13 y V15, la distancia entre *training loss* y *validation loss* es estable, lo que indica generalización moderada.
- El modelo V16 (512 px) converge más lentamente y muestra fluctuaciones mayores, típicas de modelos más grandes entrenados con datasets mixtos.

Estas oscilaciones explican por qué varios experimentos activaron el early stopping.

Resultados visuales

La **¡Error! No se encuentra el origen de la referencia.** presenta predicciones para una selección de ejemplos de los modelos multiresolución. Estas imágenes permiten evaluar la coherencia espacial del desenrollado sin depender exclusivamente de las métricas numéricas. Los patrones observados fueron:

- V12, V12.5 y V13 logran reconstrucciones razonables.
- En modelos como V13.5 y V15, se observa suavizado excesivo, pérdida de bordes y regiones planas donde deberían existir gradientes de fase.
- V14, entrenado con mapas multi-wrap, presenta artefactos residuales y mayor distorsión, lo cual concuerda con su IMMSE más alto.
- V16, a pesar de tener la entrada de mayor resolución (512 px), muestra solapamiento espacial difuso y pérdida casi total de detalles finos, lo cual coincide con sus valores más bajos de PSNR (24.47 dB).

Es importante mencionar que los objetos utilizados para entrenamiento, validación y prueba no son los mismos a lo largo de los experimentos, por lo que las predicciones en la Figura 37 no buscan comparar “el mismo objeto”, sino evaluar la calidad de reconstrucción en distintos escenarios de generalización.

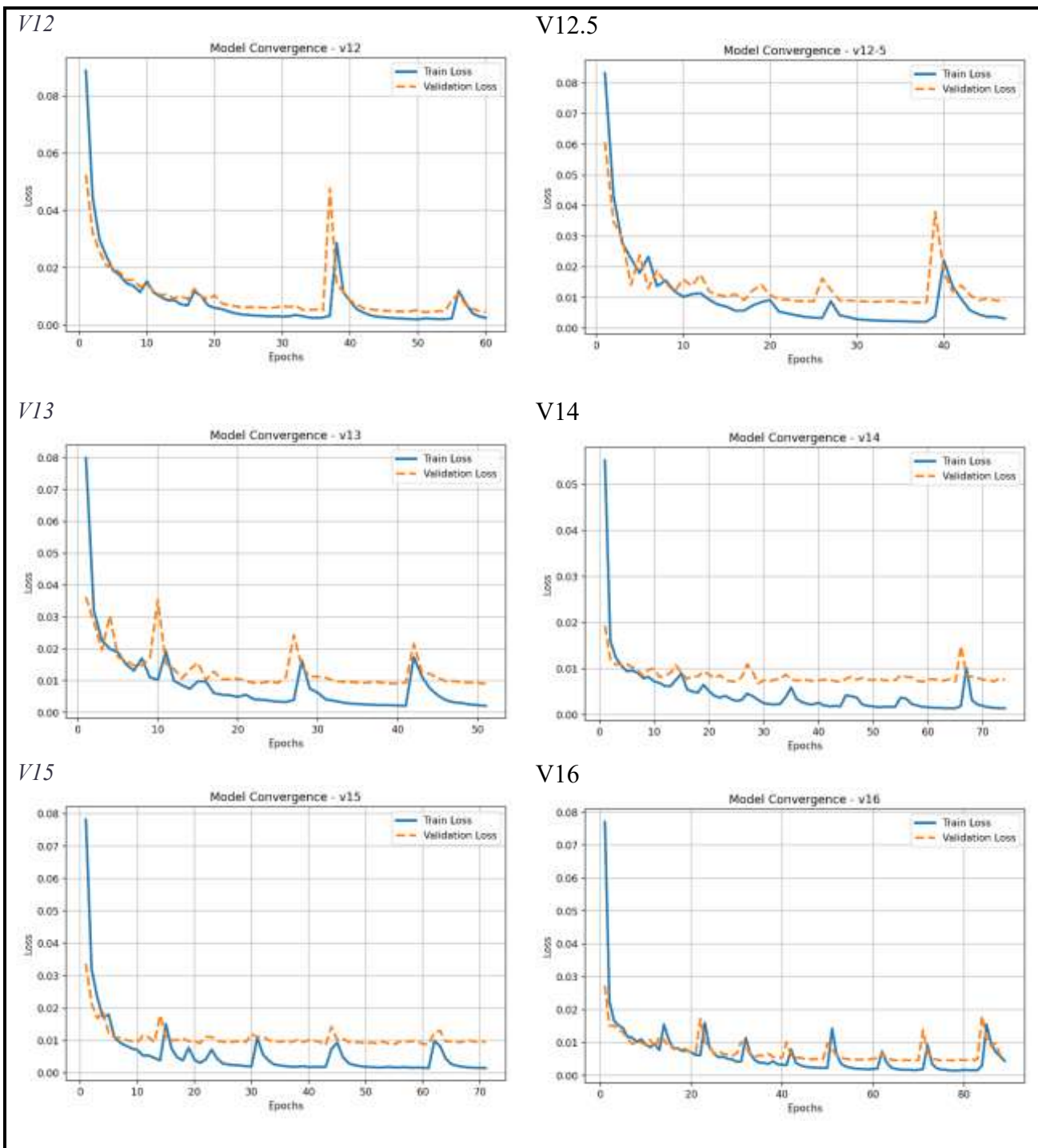


Figura 36. Gráficas de los entrenamientos de la arquitectura Multiresolution.

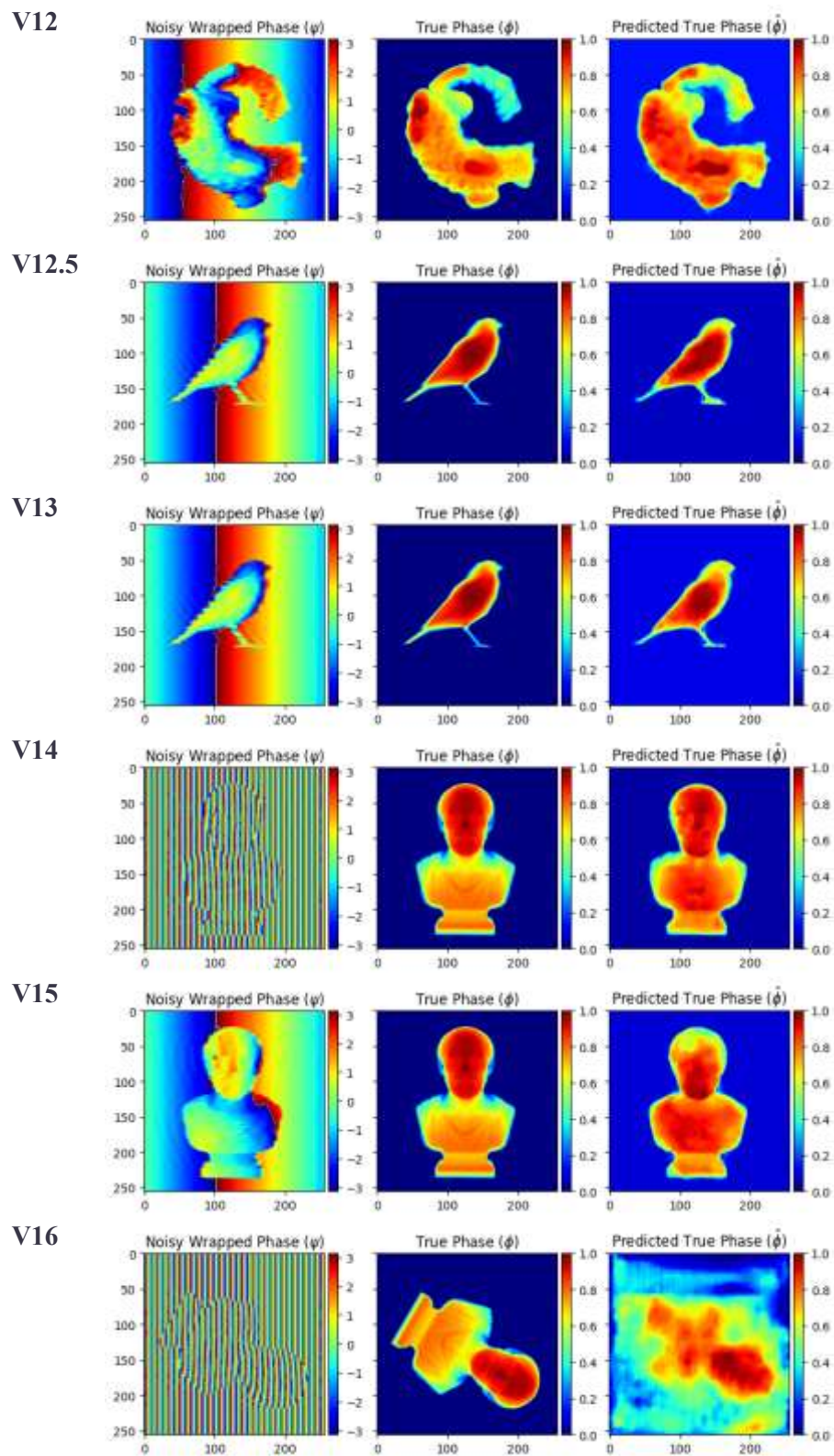


Figura 37. Predicciones de los modelos de la arquitectura Multiresolution.

6.3.3.4 *U-Net estándar*

La arquitectura U-Net estándar (V17-V21) fue utilizada como línea base para comparar el impacto de mecanismos adicionales como LSTM, bloques residuales o estructuras multiresolución. Aunque teóricamente las arquitecturas enriquecidas podrían ofrecer mayor capacidad representativa, en la práctica la U-Net simple mostró ser la más robusta para datos sintéticos, siempre que se empleara una base de datos bien curada y una estrategia de normalización consistente.

Desde el punto de vista cuantitativo, los modelos U-Net obtuvieron las mejores métricas exclusivamente en el conjunto sintético, superando a todas las demás arquitecturas. Los modelos más destacados fueron:

- V17 (low-wrap, 256 px, mixed-object split)
 - PSNR = 31.88 dB, SSIM = 0.959, IMMSE = 0.0009
- V21 (low-wrap, 512 px, mixed-object split)
 - PSNR = 30.36 dB, SSIM = 0.952, IMMSE = 0.0012

Estos valores representan el mejor desempeño en reconstrucción sintética dentro de todos los experimentos realizados. No obstante, es importante remarcar que este liderazgo no se mantiene cuando se evalúa sobre objetos reales, donde otras arquitecturas, especialmente las basadas en bloques residuales o multiresolución, presentaron mejor estabilidad frente a ruido físico, texturas reales y variabilidad experimental.

En cuanto al proceso de entrenamiento, la Figura 38 muestra que los modelos U-Net tienden a una convergencia estable, con curvas de *train loss* y *validation loss* más limpias que las observadas en otras arquitecturas. Aunque algunos picos aparecieron en validación (por ejemplo, en V18 y V20), su estabilidad general permitió obtener pesos más consistentes, lo cual se reflejó en las métricas sintéticas.

Desde el punto de vista cualitativo, la Figura 39 resume las predicciones generadas por los modelos V17–V21. A diferencia de otras arquitecturas que presentaron suavizado excesivo, artefactos residuales o pérdida de volumen en zonas de gradiente fuerte, las U-Net estándar

lograron: reconstrucciones sintéticas con geometría más nítida, continuidad espacial uniforme, bordes definidos, fidelidad visual respecto al ground truth sintético, ausencia de patrones oscilatorios.

Cabe señalar que los objetos mostrados en la Figura 39 no coinciden necesariamente con los objetos usados en entrenamiento, ya que los datasets estaban separados en subconjuntos independientes. Por ello, la comparación visual se centra en evaluar la forma en que cada modelo reproduce la estructura general del objeto, bajo el colormap propuesto en el artículo original de U-Net con LSTM, donde los colores representan únicamente alturas relativas de fase.

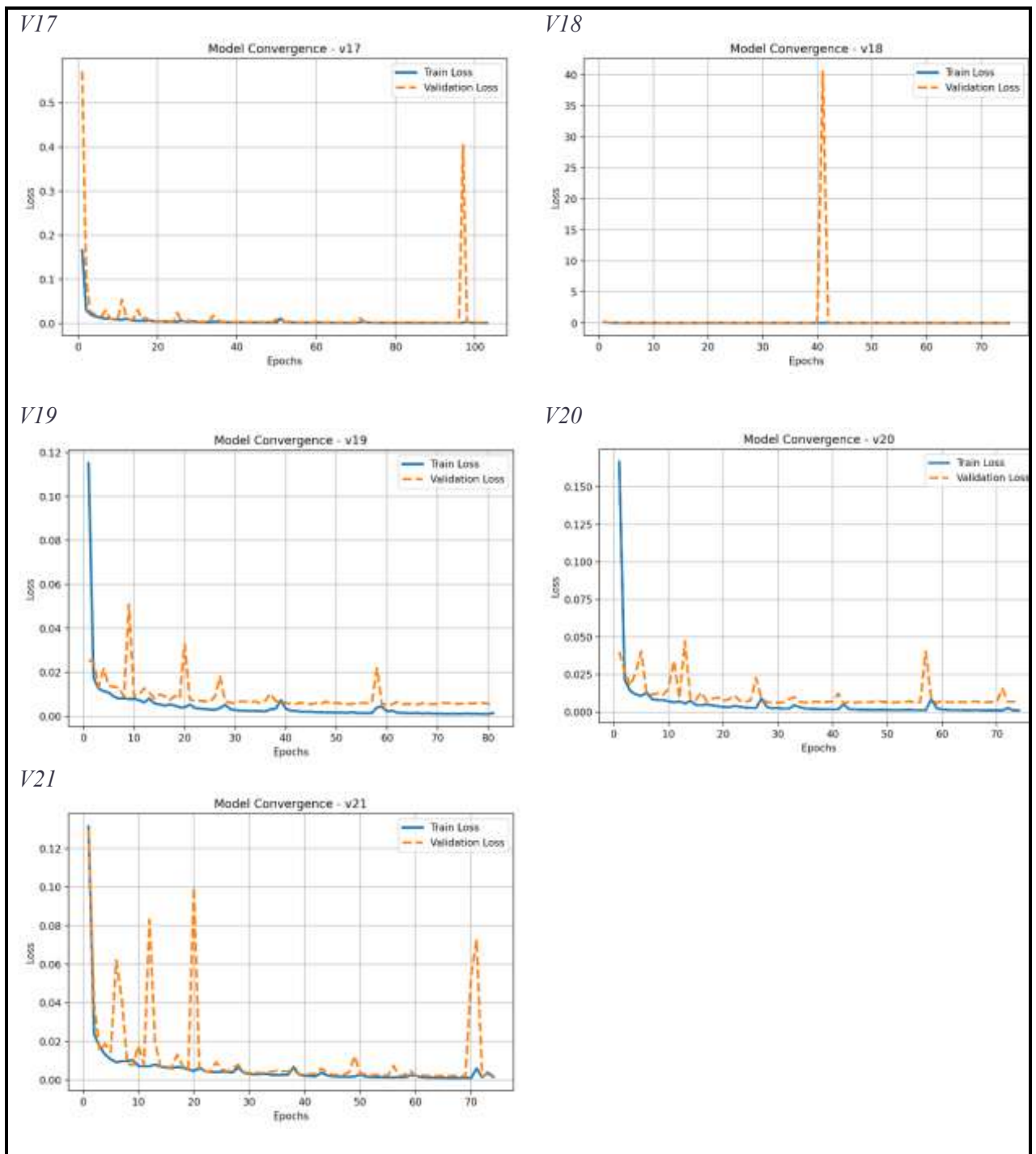


Figura 38. Gráficas de los entrenamientos de la arquitectura U-Net estándar.

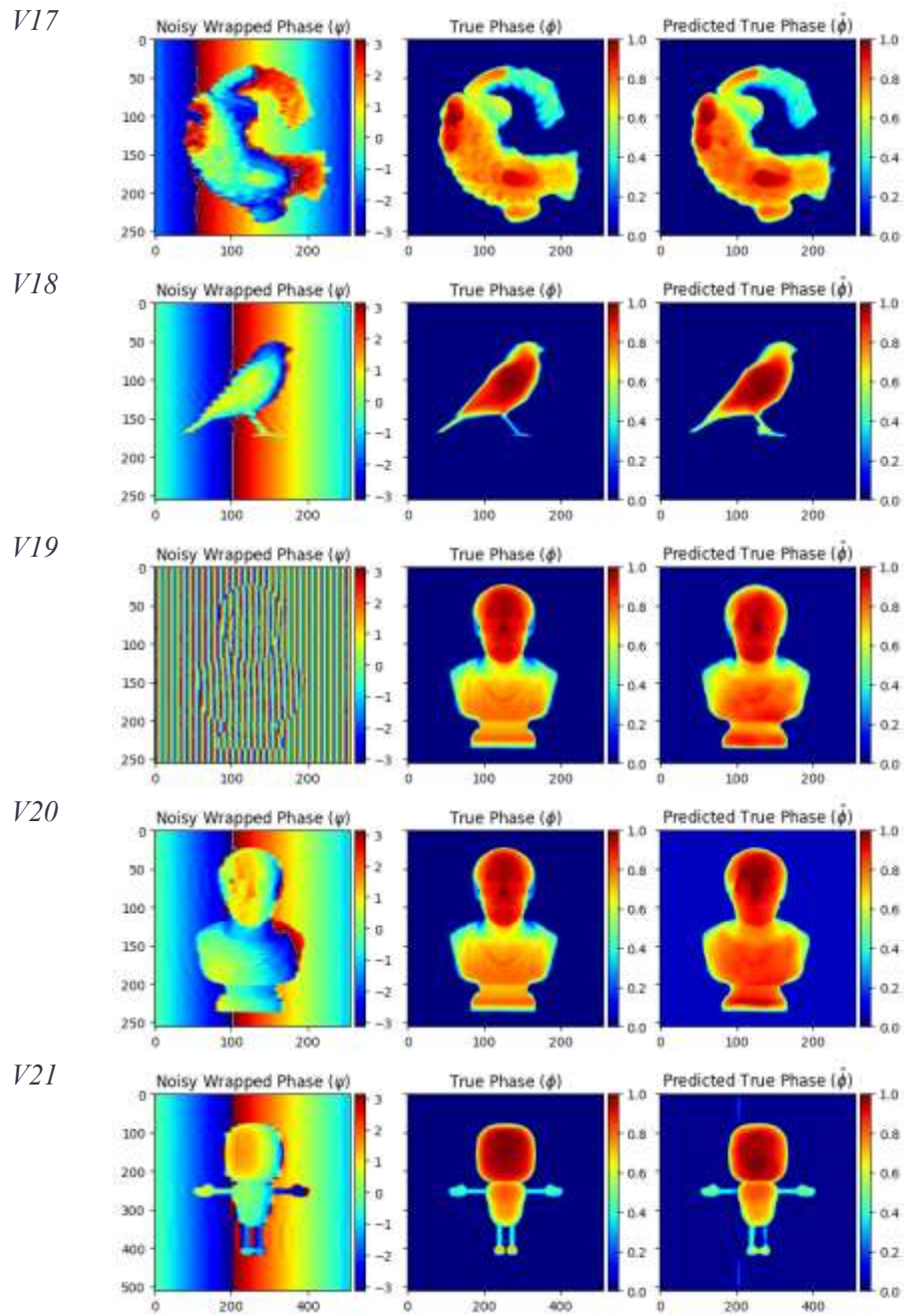


Figura 39. Predicciones de los modelos de la arquitectura U-Net estándar.

6.3.4 Evaluación en objetos sintéticos

Aunque las métricas cuantitativas (PSNR, SSIM e IM-MSE) permiten comparar objetivamente el desempeño numérico de cada arquitectura, no siempre reflejan por completo la calidad real de una reconstrucción de fase. En particular, modelos con valores similares de PSNR o SSIM pueden producir inferencias visualmente distintas, con diferencias importantes en suavidad, continuidad, preservación de detalles o presencia de artefactos locales.

Por ello, la evaluación cualitativa se vuelve un componente esencial del análisis: permite observar directamente cómo se reconstruye la forma de un objeto, identificar patrones de error, y valorar qué tan fiel es la predicción respecto al ground truth, más allá de los números. Asimismo, comparar visualmente las salidas de diferentes arquitecturas sobre el mismo objeto sintético facilita identificar fortalezas, debilidades y comportamientos característicos de cada diseño de red.

Con este propósito, la Figura 40 y la Figura 41 presentan una serie de inferencias visuales obtenidas por los modelos representativos de cada arquitectura. Estas imágenes permiten contrastar, bajo condiciones controladas, cómo cada red reconstruye la fase, cómo distribuye su error espacial y qué tan bien conserva los rasgos geométricos del objeto.

Para ello se seleccionaron versiones representativas de cada arquitectura que compartieran condiciones lo más homogéneas posible (mismos tamaños, mismo nivel de wrap y mismo esquema de datos). Esto es importante porque, en la práctica, muchos modelos fueron entrenados con configuraciones diferentes (object-wise split, mixed-object split, 256/512 px, low-wrap/multi-wrap), lo cual dificultaba una comparación directa. Por tanto, las figuras

muestran un análisis controlado, enfocado únicamente en la calidad del resultado reconstruido.

Cada bloque visual presenta tres elementos:

1. **GT ϕ** : mapa de fase ground truth.
2. **Pred $\hat{\phi}$** : salida estimada por el modelo.
3. **Mapa de error**: diferencia normalizada entre ϕ y $\hat{\phi}$, donde regiones rojas indican sobreestimación y azules, subestimación.

6.3.4.1 Resultados cualitativos V1, V7, V12 y V17

La Figura 40 muestra una comparación directa entre las predicciones generadas por cuatro modelos representativos, U-Net con SQD-LSTM (V1), U-Net Res-Net (7), Multiresolution (V12) y U-Net estándar (V17), utilizando el mismo objeto sintético y el mismo esquema de datos. Esta comparación permite observar de manera controlada las diferencias cualitativas que surgen exclusivamente por la arquitectura, independientemente de variaciones en el conjunto de entrenamiento.

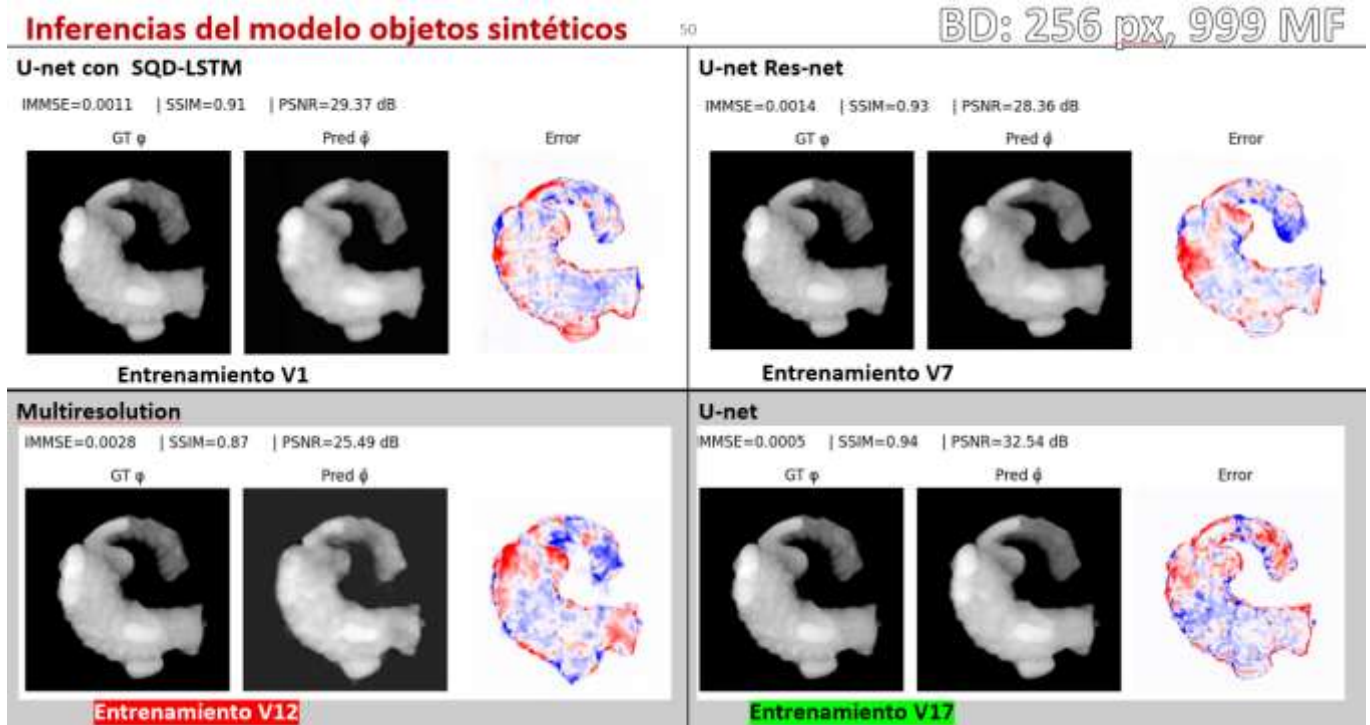


Figura 40. Predicción de un objeto con los modelos V1, V7, V12 y V17.

En términos generales, los resultados revelan comportamientos visuales diferenciados:

- U-Net estándar (V17) es la arquitectura con mejor fidelidad visual. Su reconstrucción presenta contornos nítidos, gradientes suaves y una apariencia estable en toda la superficie. El mapa de error resulta más homogéneo y concentrado en los bordes naturales del objeto, lo que indica una predicción precisa y bien regularizada.
- U-Net con SQD-LSTM (V1) también produce resultados estables, destacando en regiones de transición y zonas suaves gracias al bloque LSTM. Aunque puede perder cierto detalle fino, mantiene una forma global coherente y errores acotados, mostrando buen desempeño en superficies continuas.
- Multiresolution (V12) logra capturar adecuadamente la estructura general, pero presenta degradación en contornos estrechos y áreas con alta curvatura. Su mapa de error evidencia variaciones locales más pronunciadas, lo que refleja una sensibilidad mayor a cambios espaciales y un balance menos uniforme entre niveles de resolución.
- U-Net Res-Net (V7) obtiene reconstrucciones razonables, aunque con artefactos puntuales y oscilaciones visuales en zonas complejas o con discontinuidades de fase. El mapa de error muestra patrones rojo-azul más intensos, indicando mayor susceptibilidad a ruido o variaciones geométricas.

En conjunto, los resultados cualitativos de la Figura 40 muestran que, bajo condiciones equivalentes de entrenamiento, la U-Net estándar es la arquitectura más consistente en objetos sintéticos, mientras que las variantes con módulos adicionales (LSTM, Res-Net o

multiresolución) ofrecen ventajas parciales, pero también introducen comportamientos específicos que pueden limitar la fidelidad visual según la geometría del objeto.

6.3.4.1 Resultados cualitativos V6, V11.5, V12 y V17

Además del análisis previo, la Figura 41 presenta una segunda comparación cualitativa utilizando un objeto sintético diferente y empleando ahora la configuración de mayor resolución disponible en la base de datos (512×512 px, 1566 mapas de fase). Esta evaluación permite observar cómo se comportan las arquitecturas cuando se incrementa tanto el tamaño espacial de entrada como la diversidad del conjunto de entrenamiento.

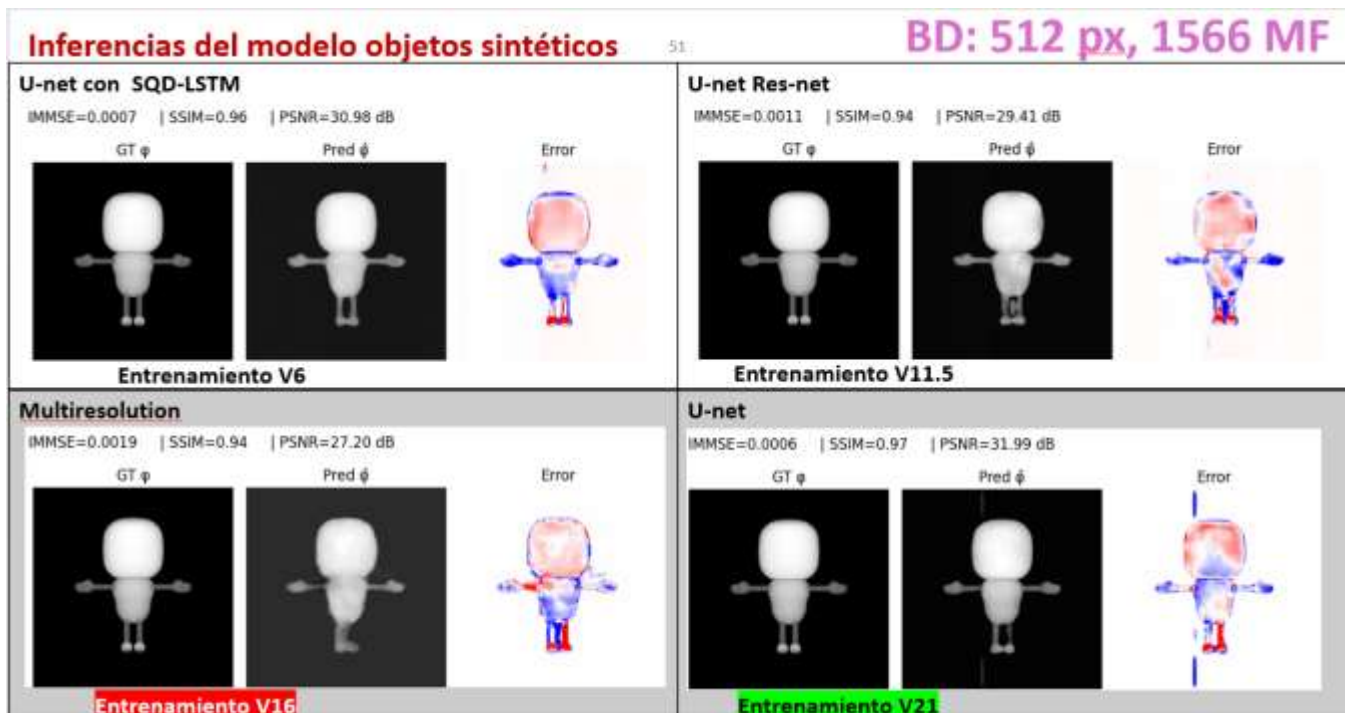


Figura 41. Predicción de un objeto con los modelos V6, V11.5, V12 y V17.

Los modelos comparados corresponden nuevamente a las cuatro arquitecturas principales: U-Net con SQD-LSTM (V6), U-Net Res-Net (V11.5), Multiresolution (V16) y U-Net estándar (V21). A diferencia del análisis anterior, estos entrenamientos sí representan *las mejores versiones de cada arquitectura dentro del conjunto de 512 px*, por lo que

proporcionan una referencia cualitativa más cercana a su desempeño óptimo en objetos sintéticos.

En esta figura pueden observarse las siguientes tendencias:

- U-Net estándar (V21) vuelve a destacar como la arquitectura con la reconstrucción más limpia y consistente. Su predicción muestra bordes bien definidos y un mapa de error más concentrado en los límites del objeto, indicando una buena preservación de la geometría global incluso a resolución elevada.
- U-Net con SQD-LSTM (V6) obtiene resultados visualmente estables, con una predicción suave y homogénea. En este caso, la mayor resolución favorece al bloque LSTM, ya que ayuda a estabilizar gradientes en zonas amplias del objeto. El error es bajo y tiende a distribuirse en regiones periféricas.
- U-Net Res-Net (V11.5) presenta una reconstrucción correcta, aunque con artefactos localizados que aparecen principalmente en los extremos de las extremidades. Su mapa de error muestra patrones alternos más notorios, reflejando cierta sensibilidad a variaciones locales del ruido o discontinuidades.
- Multiresolution (V16) muestra la forma general del objeto, pero la predicción exhibe suavizado excesivo y pérdida de detalle fino, especialmente en las zonas laterales. El mapa de error presenta parches amplios y texturas irregulares, indicando que esta arquitectura no aprovecha de manera óptima la resolución adicional.

En conjunto, la Figura 41 confirma lo observado previamente: las arquitecturas basadas en U-Net pura (V17 en la resolución baja y V21 en 512 px) son las que alcanzan mayor estabilidad visual y menor error estructural en objetos sintéticos, mientras que los modelos con módulos adicionales (Res-Net, LSTM o multiresolución) logran reconstrucciones funcionales, pero tienden a introducir artefactos característicos de su diseño.

6.3.4.2 *Comparación entre el mejor y el peor entrenamiento*

Después de analizar el desempeño general de todas las arquitecturas, es útil contrastar casos específicos que representen los extremos del comportamiento dentro del conjunto de experimentos sintéticos. Esto permite estudiar no solo qué modelo obtiene mejores métricas, sino también *por qué* las obtiene, y qué tipo de errores caracterizan a un modelo con desempeño bajo.

Para este análisis se seleccionaron dos entrenamientos representativos:

- **V17 (U-Net estándar, low-wrap, 256 px):** mejor desempeño cuantitativo en objetos sintéticos.
- **V11 (U-Net Res-Net, low-wrap, 512 px):** uno de los desempeños cuantitativos más bajos de todas las arquitecturas evaluadas bajo condiciones sintéticas homogéneas.

La Figura 42 presenta una comparación visual directa entre ambos modelos. En cada caso se muestra la fase envuelta, la fase verdadera (GT φ), la predicción ($\hat{\varphi}$) y el mapa de error normalizado.

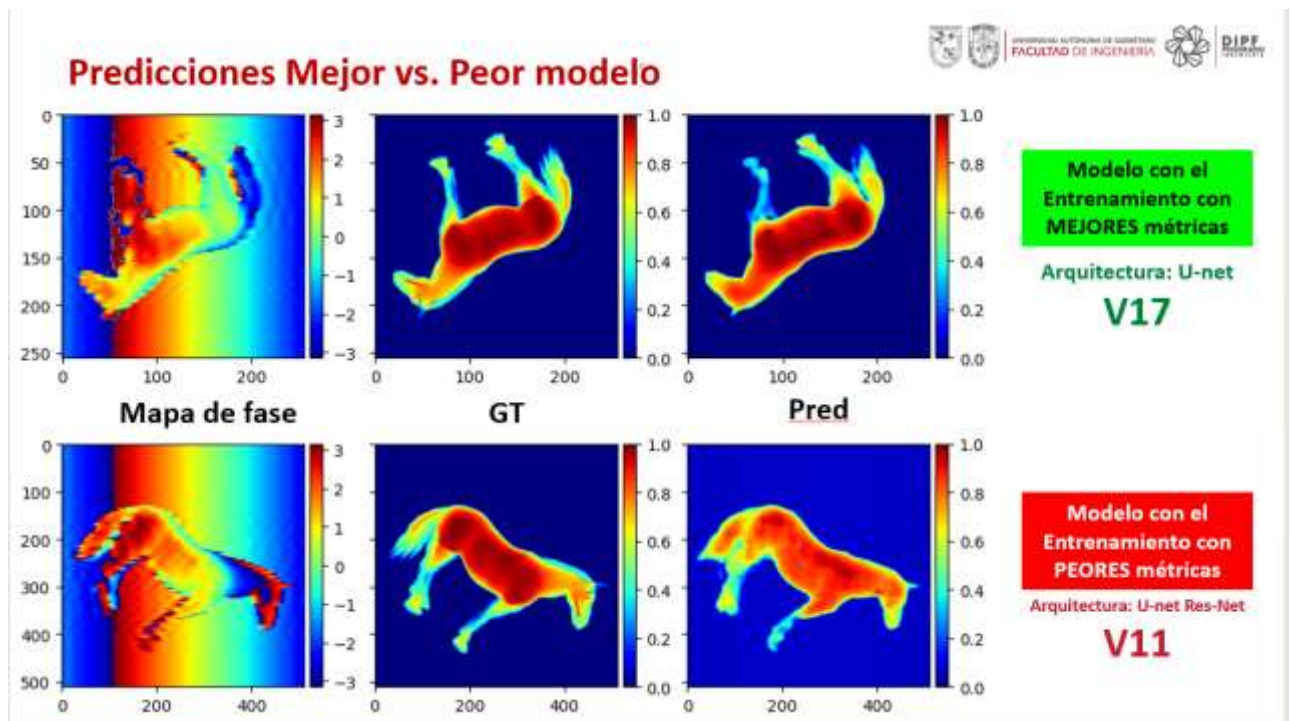


Figura 42. Predicciones del desdoblamiento de fase entre el mejor vs peor entrenamiento.

Análisis del modelo con mejores métricas: V17

El modelo V17 logra una reconstrucción altamente consistente con la fase verdadera normalizada:

- conserva la geometría general del objeto,
- reproduce correctamente los gradientes y transiciones,
- mantiene contornos definidos,
- y respeta el rango dinámico esperado en $[0,1]$.

Su mapa de error es compacto y estructurado, con discrepancias principalmente en bordes y zonas de mayor pendiente. Estas características explican sus valores elevados de PSNR, SSIM e IMMSE bajos: su predicción mantiene relación directa tanto en forma como en amplitud respecto al ground truth.

El mapa de error asociado muestra patrones homogéneos concentrados sobre los bordes naturales del objeto, donde se esperan discrepancias debido a cambios bruscos en la fase. En

términos prácticos, V17 no solo reproduce la geometría global, sino que además respeta el rango dinámico de la fase normalizada en $[0,1]$, condición indispensable para una evaluación coherente y para que las métricas reflejen la calidad real de la reconstrucción.

Análisis del modelo con las métricas más bajas: V11

Aunque V11 también recupera la forma global del objeto, su error principal proviene de un problema sistemático de escala en la predicción:

- Las intensidades predichas se comprimen hacia valores bajos.
- El objeto queda visualmente "aplanado", ocupando un rango dinámico reducido.
- Las zonas que deberían mostrar gradientes amplios se ven subexpuestas.

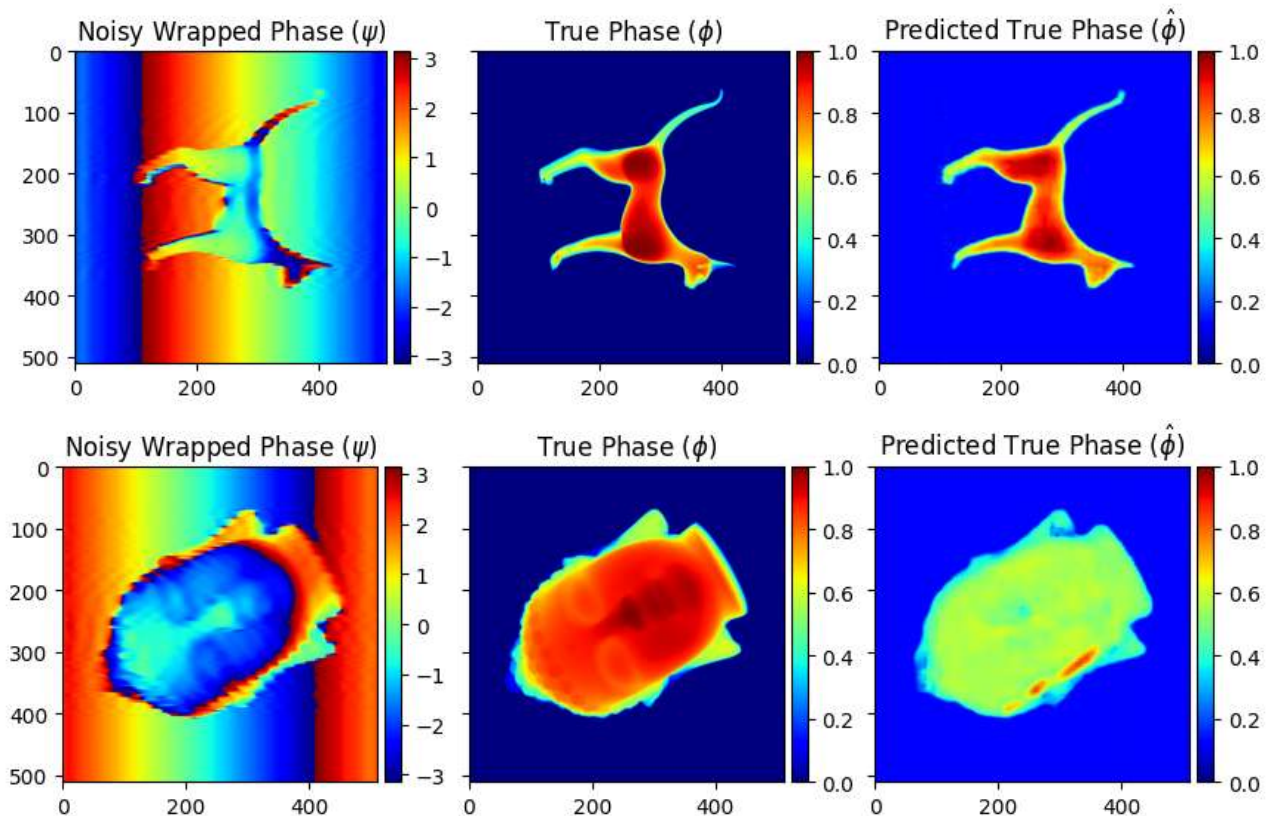


Figura 43. Más ejemplos del modelo v11.

En los ejemplos presentados se evidencia de mejor forma la diferencia de escala pero que aún sigue reconstruyendo la geometría del objeto.

Este colapso de escala provoca:

- PSNR y SSIM significativamente menores.
- Grandes regiones destacadas en el mapa de error, no por fallar en la geometría, sino por discrepancias en amplitud relativa.
- Una falta de correspondencia adecuada entre la fase envuelta y la fase normalizada al momento de la inferencia.

Este fenómeno se evidencia aún más claramente en la Figura 43, donde se muestran ejemplos adicionales de V11 evaluado en distintos objetos sintéticos. Aunque la forma general está bien recuperada, el rango dinámico está incorrectamente escalado, generando salidas con apariencia de “lavado” y patrones de error amplificados.

La comparación entre V17 y V11 permite extraer varias conclusiones relevantes:

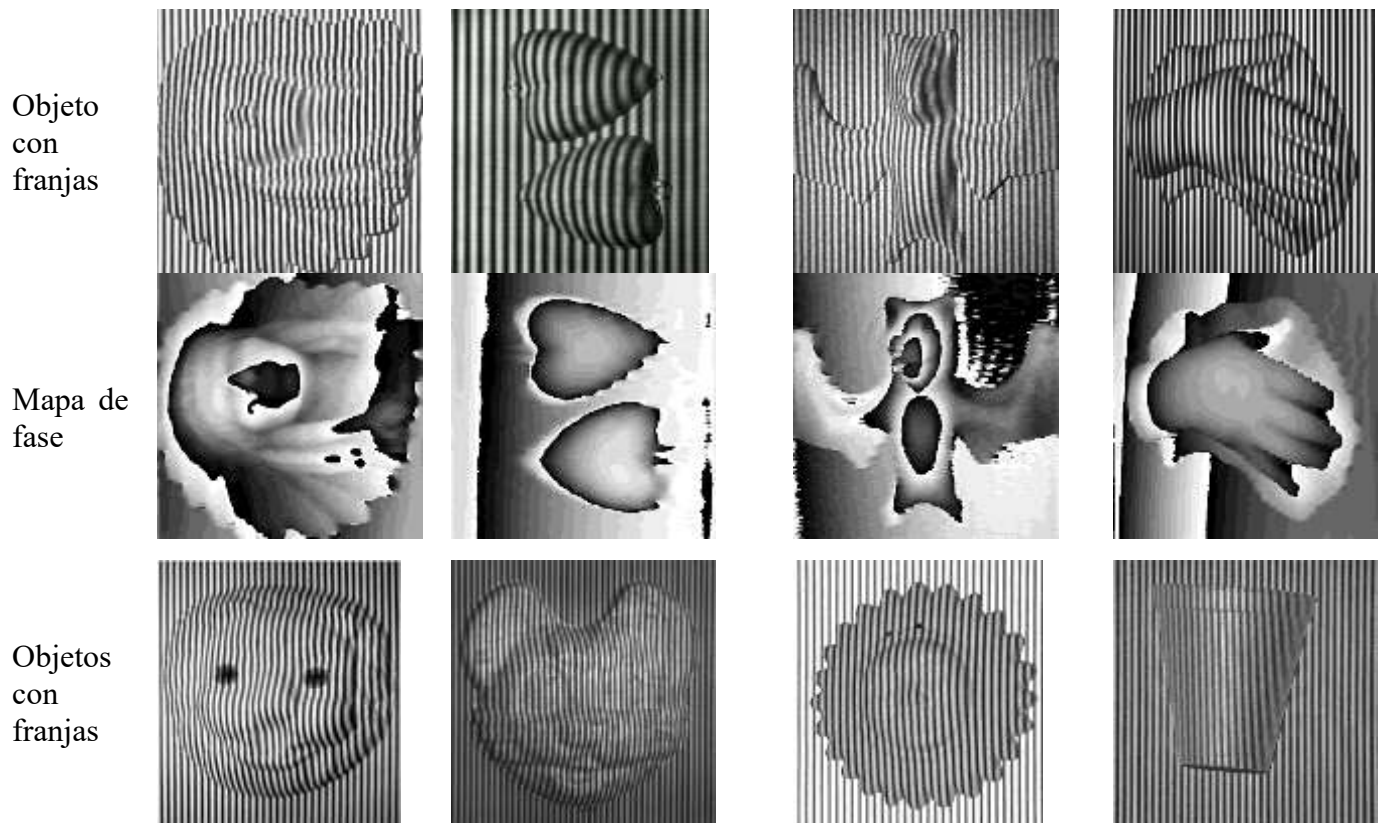
1. La forma y la escala deben evaluarse por separado.
Un modelo puede recuperar correctamente la geometría, pero si la predicción colapsa hacia una escala incorrecta, las métricas cuantitativas caerán drásticamente.
2. El desempeño no depende únicamente de la arquitectura.
En esta comparación, la diferencia también se relaciona con:
 - a. el esquema de split, la resolución de entrada (256 px vs 512 px) y la estabilidad estadística del conjunto low-wrap utilizado.
3. La evaluación cualitativa revela fallas que las métricas por sí solas no explican.
En V11, el error visual surge del rango dinámico comprimido, no de la forma reconstruida. Sin un análisis visual, esta causa sería difícil de identificar.
4. Las redes residuales pueden introducir sensibilidad adicional a la distribución de intensidades.
En V11, el bloque residual parece exacerbar la tendencia a la compresión de escala bajo condiciones low-wrap.

En resumen, esta comparación evidencia que un modelo puede obtener métricas bajas aun cuando reproduce la geometría del objeto, debido a errores de escala, normalización o estabilidad estadística. Este análisis ayuda a entender por qué algunos modelos destacan en

objetos sintéticos y por qué otros no, sin adelantar conclusiones sobre el desempeño con objetos reales, que se analizarán en la siguiente sección.

6.3.5 Inferencia en objetos reales

Después de analizar a profundidad el desempeño de los modelos entrenados con datos sintéticos, el siguiente paso consistió en evaluar su capacidad de generalización hacia datos experimentales reales. Esta etapa es crítica porque permite cuantificar el llamado dominio-gap, es decir, las diferencias existentes entre los patrones estadísticos de las fases sintetizadas y los mapas de fase obtenidos mediante un sistema físico de proyección de franjas. A diferencia del entorno simulado, en el cual todos los objetos presentan superficies limpias, geometría controlada y ruido limitado, los objetos reales incorporan irregularidades propias del proceso de adquisición: sombras, saturaciones, reflejos, ruido estructural, discontinuidades no modeladas y variaciones locales en la iluminación.



Mapa de fase

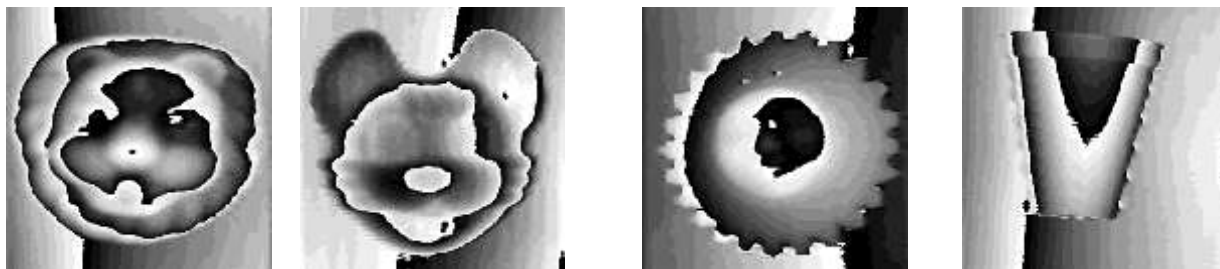


Figura 44. Ejemplo de objetos reales con franjas proyectadas y su mapa de fase.

Para esta evaluación se empleó un conjunto de más de treinta objetos reales capturados mediante proyección sinusoidal, cada uno con una única imagen de fase envuelta de 512×512 px y su mapa de fase correspondiente obtenido mediante Transformada de Fourier. La Figura 44 muestra ejemplos representativos, donde pueden observarse distintos niveles de complejidad geométrica y diferentes patrones de franjas que afectan directamente la calidad de la fase.

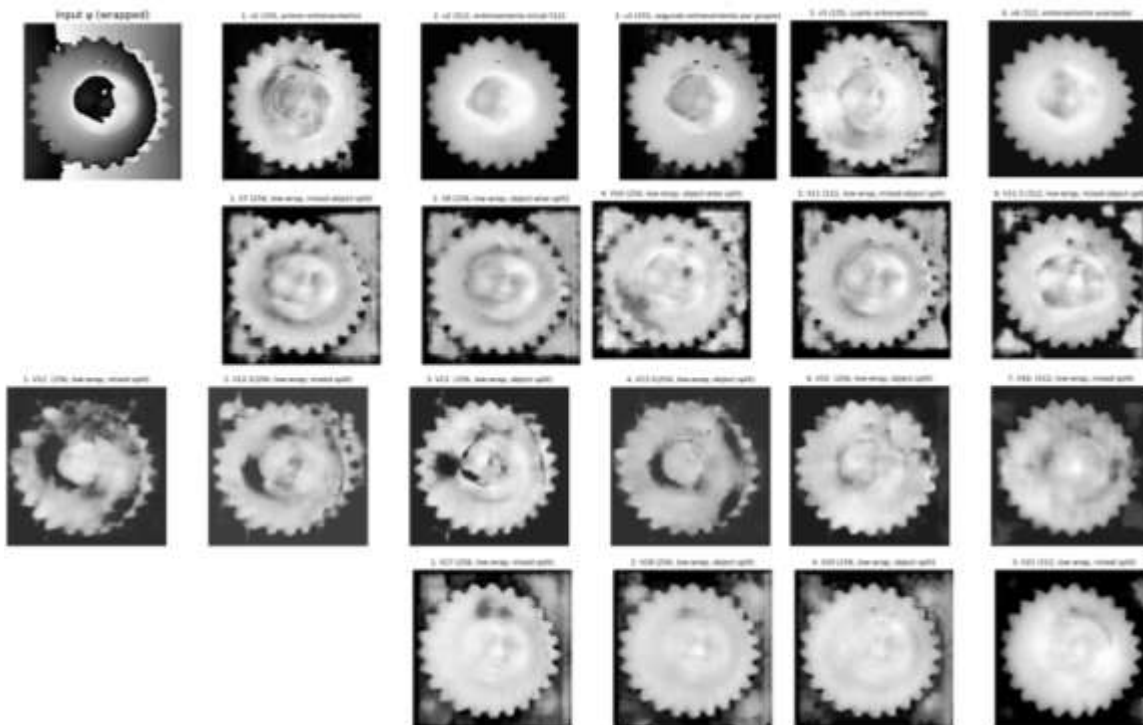


Figura 45. Inferencias de todos los modelos entrenados para un objeto (sol).

A partir de los modelos entrenados exclusivamente con datos sintéticos, se evaluó su capacidad de generalización frente a objetos reales adquiridos mediante el sistema de proyección de franjas. Esta evaluación representa un escenario más exigente, pues ninguno de los modelos había sido expuesto a ruido experimental, discontinuidades físicas, texturas complejas ni variaciones lumínicas propias de los sistemas ópticos reales. A diferencia del caso sintético, no se cuenta con un ground truth de fase ni con una máscara precisa del objeto, por lo que no es posible calcular métricas cuantitativas (IMMSE, PSNR o SSIM) ni delimitar el área del objeto para excluir el fondo. En consecuencia, el análisis debe realizarse de forma totalmente cualitativa, atendiendo a la estabilidad visual de la reconstrucción, la conservación de la forma, la continuidad espacial y la presencia o ausencia de artefactos.

Las Figuras 45, 46 y 47 presentan las predicciones completas de los modelos para tres objetos reales representativos. A pesar de las diferencias en arquitectura, tamaño y configuración de entrenamiento, todos los modelos lograron recuperar la geometría global del objeto: se reconocen contornos principales, volúmenes aproximados y simetrías generales. Esto evidencia que la relación funcional aprendida entre fase envuelta y fase normalizada en el dominio sintético sí contiene información transferible. Sin embargo, las diferencias entre arquitecturas emergen de inmediato cuando se analiza la estabilidad tonal, el manejo del ruido experimental y la coherencia de la amplitud.

6.3.5.1 Desempeño del modelo V17 (U-Net estándar)

A pesar de que V17 obtuvo las mejores métricas en datos sintéticos, su comportamiento en objetos reales fue menos sólido de lo esperado. Las predicciones muestran una compresión marcada del rango dinámico, lo que se manifiesta en:

- mapas de fase excesivamente oscuros,
- pérdida de contraste entre fondo y objeto,
- saturación hacia valores bajos,
- transiciones menos graduales en zonas de altura variable.

Aunque la forma se conserva, la amplitud no se expande correctamente, lo que vuelve difícil interpretar la geometría real del objeto. Este problema coincide con los fallos observados en su “peor caso” sintético (V11), donde la forma era correcta pero la escala colapsaba. Esto sugiere que la U-Net aprendió patrones tonales muy específicos del entorno sintético y no logró extrapolarlos al ruido experimental sin ajuste adicional.

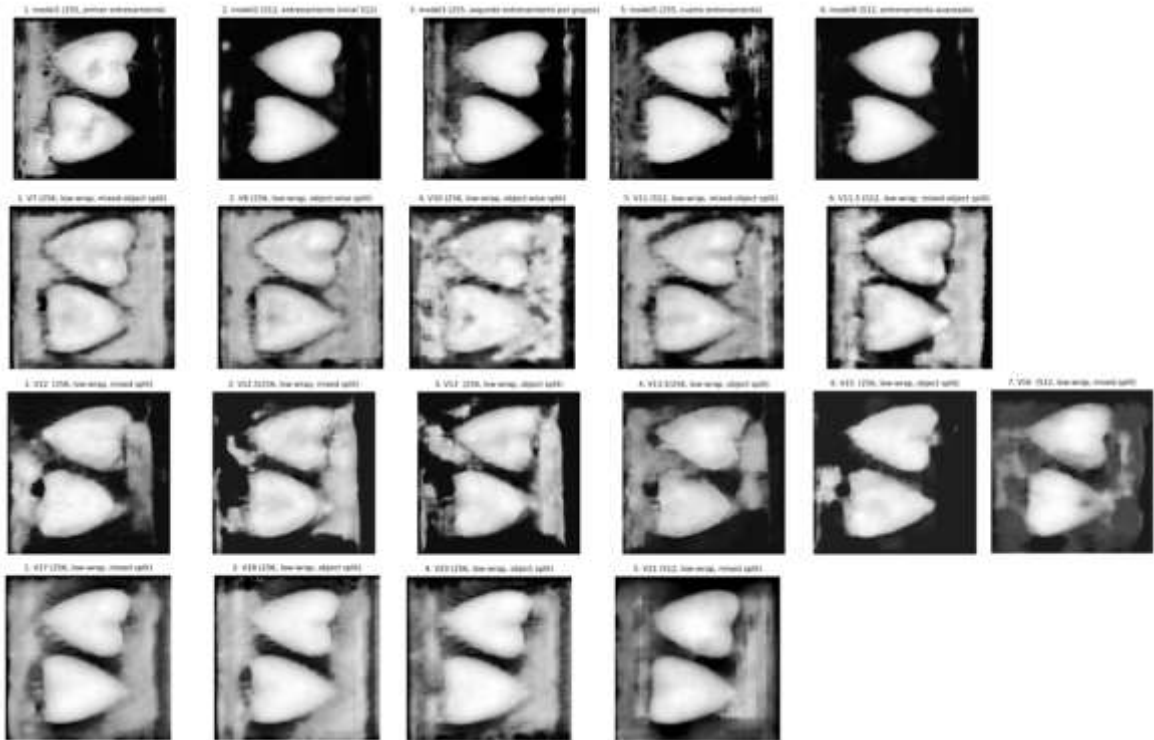


Figura 46. Inferencias de todos los modelos entrenados para un objeto (corazones).

6.3.5.2 Desempeño del modelo V6 (U-Net con SQD-LSTM)

En contraste, el modelo V6, que no fue el mejor en métricas sintéticas, mostró la mayor estabilidad visual en los datos reales. Sus reconstrucciones presentan:

- mejor continuidad espacial,
- menos fluctuaciones bruscas,
- transiciones de altura más homogéneas,
- ruido notablemente reducido,
- bordes más coherentes en comparación con otras arquitecturas.

Esto indica que el módulo LSTM aportó una ventaja funcional al capturar dependencias espaciales largas y patrones correlacionados del ruido real. Dado que las franjas experimentales poseen una estructura secuencial altamente dependiente entre líneas, la capacidad del LSTM para modelar correlaciones espaciales parece favorecer la generalización en escenarios fuera del dominio sintético.

6.3.5.3 *Arquitecturas Res-Net y Multiresolution (V11.5, V12, V15, V16)*

Los modelos basados en bloques residuales y en escalas múltiples mostraron un desempeño intermedio. Aunque recuperan la forma general del objeto, presentan dificultades cuando se enfrentan al ruido experimental:

- las arquitecturas Res-Net producen texturas ruidosas, bordes irregulares o artefactos localizados,
- los modelos Multiresolution tienden a suavizar en exceso, perdiendo detalles importantes de la geometría,
- ambos tipos de arquitectura son sensibles a variaciones tonales no presentes durante el entrenamiento.

Este comportamiento es coherente con lo observado en datos sintéticos: estas arquitecturas funcionan bien cuando la distribución estadística del dataset es estable, pero se vuelven frágiles cuando el dominio cambia.

6.3.5.1 *Comparación global entre modelos*

La comparación global entre los modelos (Figuras 45–47) muestra que, aunque el desempeño en objetos reales es variable, varios modelos logran reconstrucciones razonablemente coherentes, recuperando la forma general y los volúmenes principales aun sin haber sido entrenados con datos experimentales. Esto es especialmente relevante considerando que en esta etapa no se cuenta con ground truth, ni máscara del objeto, ni control sobre el ruido real, lo que convierte la evaluación en un análisis exclusivamente cualitativo.

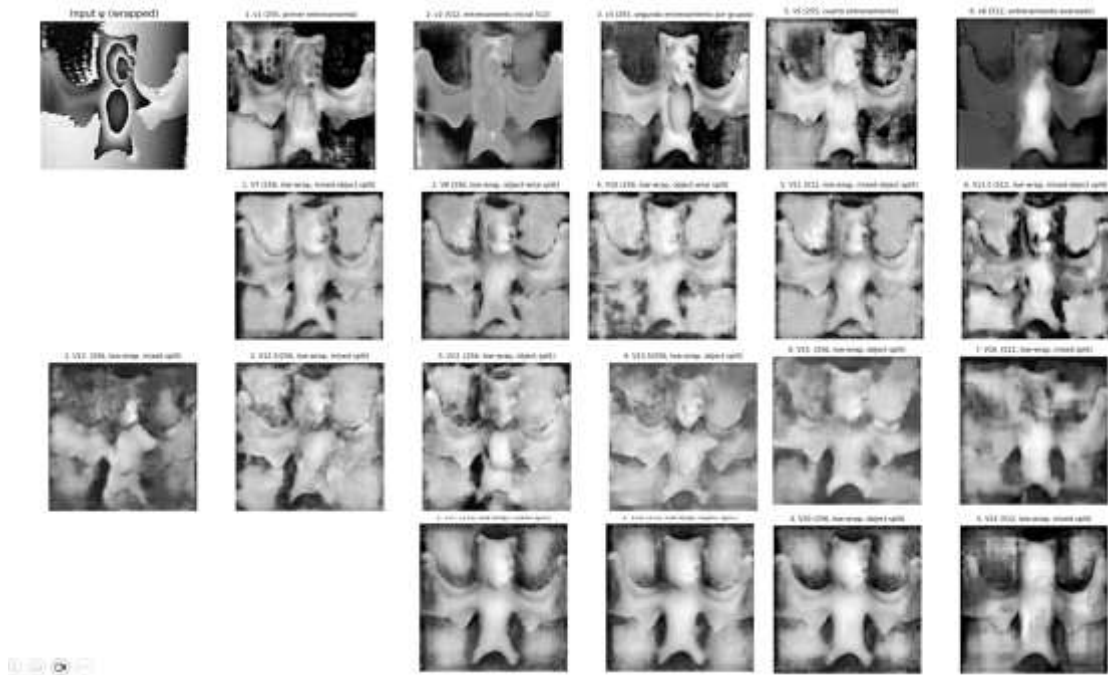


Figura 47. Inferencias de todos los modelos entrenados para un objeto (murciélago).

En este contexto, las diferencias entre arquitecturas se hacen evidentes: V6 (U-Net + SQD-LSTM) ofrece las reconstrucciones más estables, con continuidad espacial y menos fluctuaciones bruscas, V17 (U-Net estándar) conserva la forma pero presenta una compresión notable del rango dinámico y las versiones Res-Net y Multiresolution recuperan la estructura global, pero son más sensibles al ruido, generando texturas irregulares o suavizados excesivos.

Los resultados muestran que en varios objetos la fase real puede aproximarse con una calidad aceptable, especialmente cuando el modelo fue entrenado con una mayor cantidad de mapas sintéticos o con arquitecturas capaces de capturar dependencias espaciales largas (como las LSTM). Esta tendencia evidencia que el volumen y la diversidad del dataset influyen directamente en la capacidad de generalización hacia el dominio real.

Estas observaciones motivaron dos direcciones claras: incorporar data augmentation para aumentar la variabilidad del dominio sintético, y construir una base de datos ampliada (BD6) con geometrías más variadas y condiciones de renderizado más ricas.

Finalmente, la Figura 48 muestra una comparación más amplia utilizando únicamente la arquitectura más estable (V6), en la cual se aprecia que, aunque persisten artefactos y variaciones tonales, el modelo mantiene la forma principal del objeto en la mayoría de los casos. Esto confirma que los modelos sintéticos pueden aproximar parcialmente la fase real, pero también que existe un dominio-gap que hace indispensable una etapa posterior de fine-tuning para obtener reconstrucciones confiables en condiciones experimentales.

Mejor arquitectura U-Net SQD-LSTM

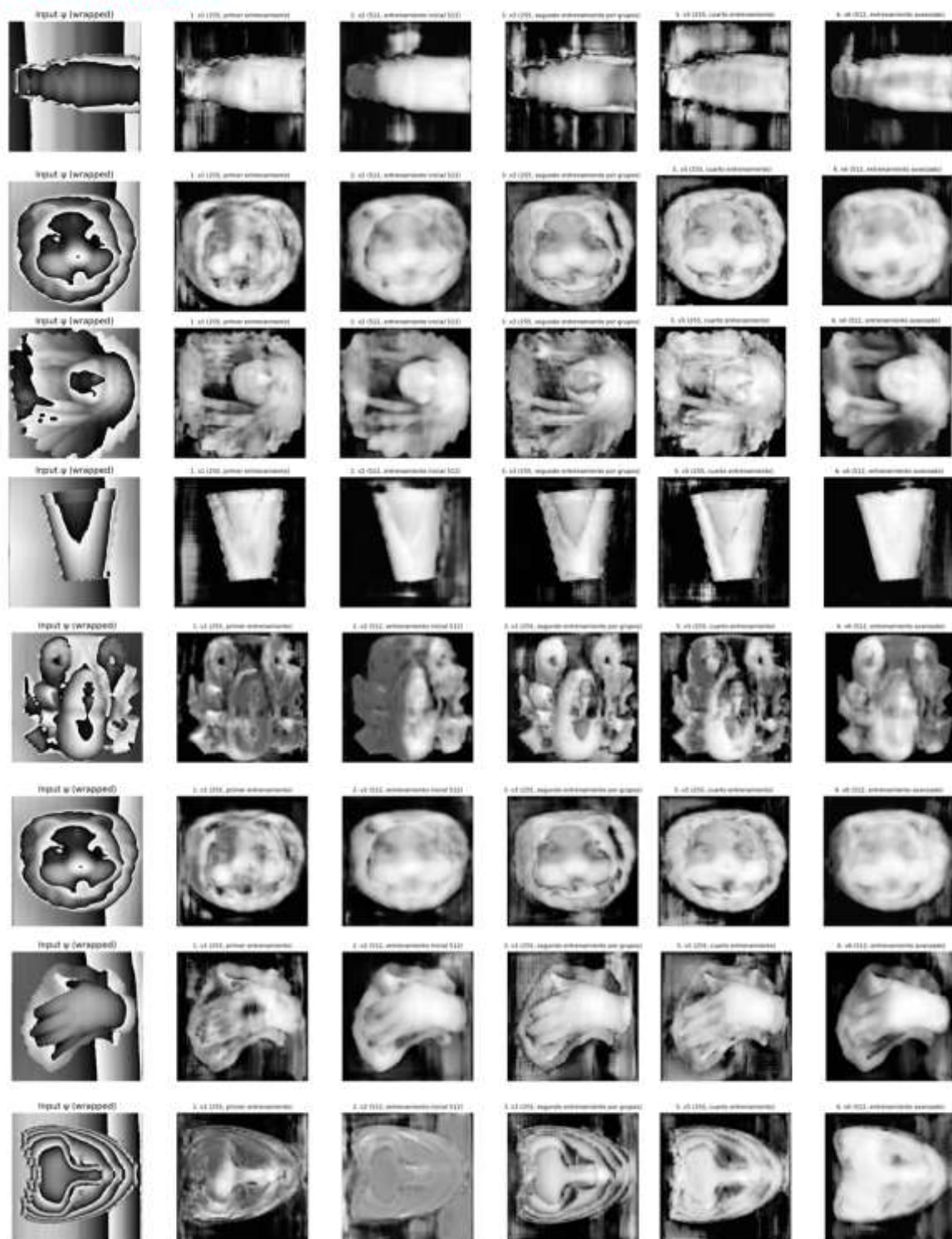


Figura 48. Inferencias arquitectura U-Net con SQDM para objetos reales.

6.3.6 Fine-tuning

La etapa de fine-tuning se planteó como un experimento complementario cuyo propósito principal fue explorar si el rendimiento del mejor modelo entrenado con datos sintéticos podía mejorar al exponerlo directamente a la distribución real de los mapas de fase envuelta. Aunque el modelo base –correspondiente al entrenamiento V6 (U-Net con SQD-LSTM)– ya producía reconstrucciones razonablemente buenas sobre los objetos reales, se observó que existían discrepancias significativas entre las características estadísticas de los mapas sintéticos y las de los mapas obtenidos experimentalmente. Esto motivó realizar un ajuste fino dirigido específicamente al dominio real.

6.3.6.1 *Desajuste estadístico entre datos sintéticos y reales*

Al analizar los histogramas de fase de ambas fuentes (

Figura 49 y Figura 50), se observó que los mapas sintéticos presentan distribuciones suaves, unimodales o con transiciones relativamente controladas, típicas de escenas con ruido estrictamente gaussiano o gradientes leves. En contraste, los mapas de fase reales exhiben:

- picos abruptos asociados a saturación del sensor,
- colas pesadas debidas a ruido estructurado,
- irregularidades provenientes del sistema de iluminación y del algoritmo FT,
- zonas localizadas con discontinuidades abruptas,
- variaciones no gaussianas y acumulación de valores en regiones inesperadas.

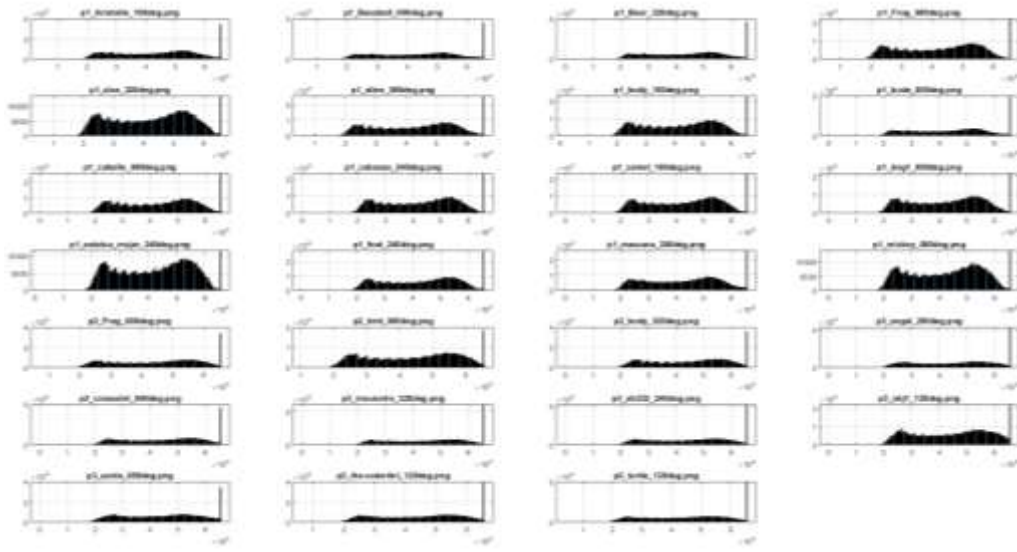


Figura 49. Histogramas de los mapas de fase de objetos sintéticos.

Estas diferencias confirman que el dominio real contiene complejidades que no estaban presentes en los datos generados sintéticamente, incluso en la variante low-wrap. Por lo tanto, aun cuando el modelo V6 generalizaba relativamente bien, existía evidencia de un domain gap estadístico que justificaba explorar una adaptación basada en datos reales.

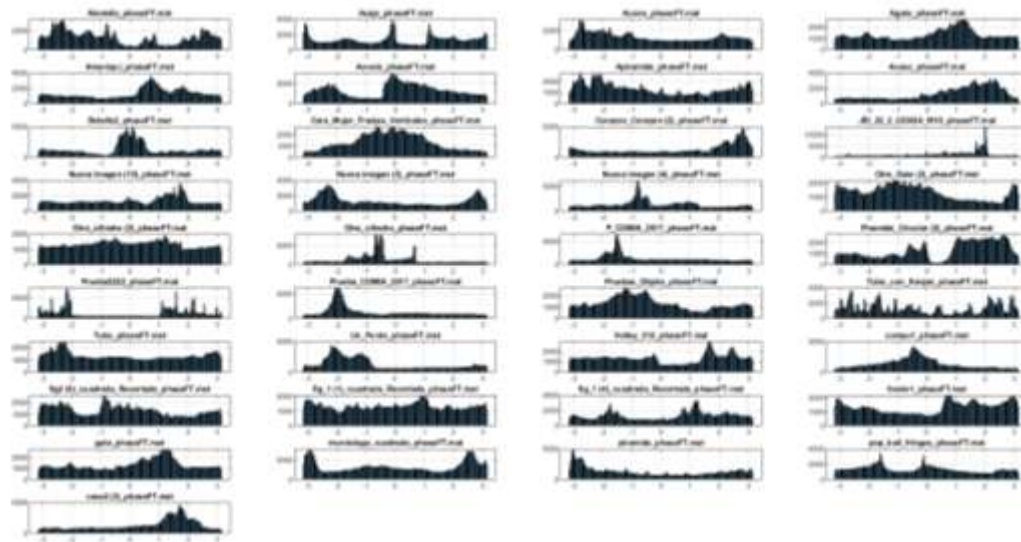


Figura 50. Histogramas de los mapas de fase de objetos reales.

6.3.6.2 Selección del modelo

Si bien otros modelos obtuvieron mejores métricas en el dominio sintético (por ejemplo V17), éstos mostraron menor estabilidad y mayor sensibilidad a ruido real durante las pruebas de inferencia. El modelo V6, en cambio, presentaba predicciones más coherentes visualmente: preservaba continuidad espacial, evitaba artefactos abruptos y respondía mejor a variaciones locales en la fase, características que son especialmente importantes en mapas de fase adquiridos experimentalmente.

Además, V6 se entrenó con el conjunto low-wrap, 512 px, 1566 ejemplos (Mixed-object split), lo cual proporcionó una base suficientemente rica para iniciar un proceso de adaptación sin sobrepasar la capacidad del modelo. Por ello, V6 fue seleccionado como la arquitectura más adecuada para la etapa de fine-tuning.

6.3.6.3 Aumento de datos

Para el fine-tuning se utilizaron aproximadamente 30 objetos reales adquiridos con el sistema de proyección de franjas. Cada objeto contiene: un único mapa de fase envuelta de tamaño 512×512 px, variabilidad en geometría, orientación y escala, niveles de ruido muy distintos entre sí, y diferencias notables en distribución de frecuencias y sombreado.

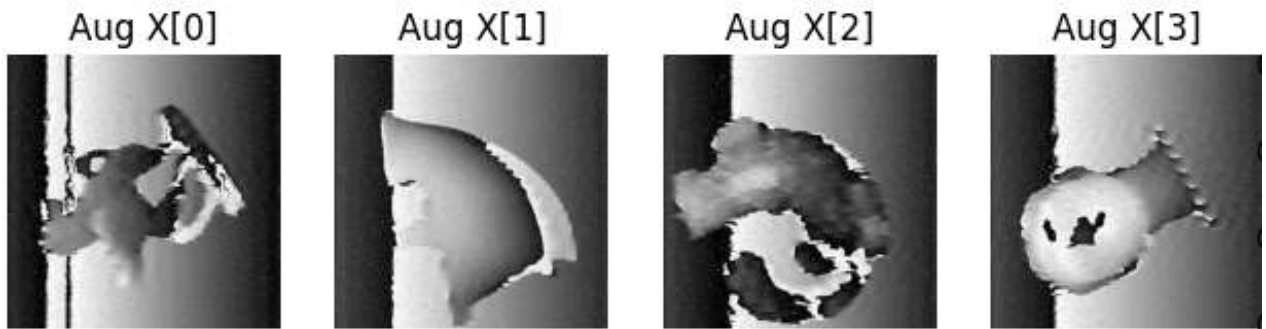


Figura 51. Mapas de fase con configuraciones de aumento de datos.

Dado que el número total de ejemplos reales es relativamente reducido en comparación con los sintéticos, se diseñó una estrategia de data augmentation especializada para mapas de fase, cuyo objetivo fue ampliar artificialmente la diversidad de condiciones reales. Esta estrategia incluye transformaciones que imitan con mayor fidelidad el comportamiento del

sistema experimental: escalamiento global de fase, desplazamientos de media, ruido suave con filtrado gaussiano, ruido granular de alta frecuencia, zonas corrompidas simulando fallos de iluminación y finalmente un re-envolvimiento de fase, asegurando que el resultado respete la periodicidad intrínseca de la señal.

Este procedimiento permitió generar múltiples variantes realistas de un mismo mapa, mejorando la robustez del modelo sin alterar la estructura fundamental de la fase.

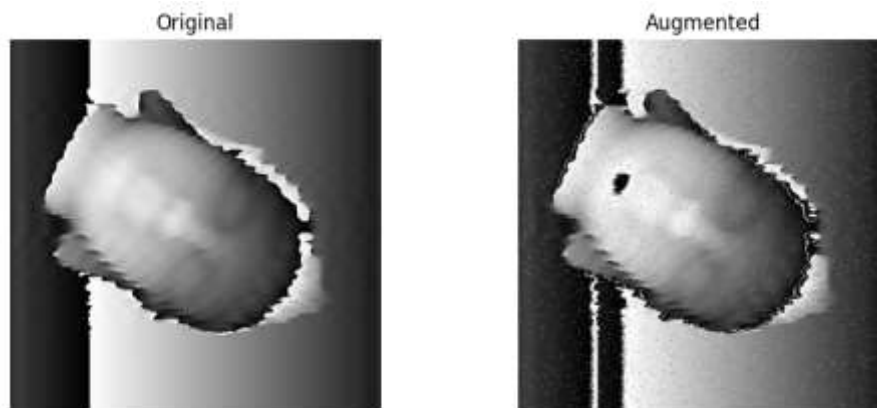


Figura 52. Ejemplo de los efectos aplicados para aumento de datos en los mapas de fase.

6.3.6.4 Procedimiento de entrenamiento

El fine-tuning se realizó durante 20 épocas con batch size de 4, utilizando el optimizador Adam y las funciones MAE y MSE únicamente para monitoreo. Se emplearon las técnicas de Early Stopping y ReduceLRonPlateau, lo que permitió estabilizar el proceso en un número reducido de épocas y evitar sobreajuste. Los generadores de datos combinaban aleatoriamente las instancias reales con sus versiones aumentadas, garantizando un flujo continuo de variabilidad estadística.

6.3.6.5 Resultados y comportamiento del modelo

Las curvas de convergencia muestran una disminución rápida del loss en las primeras épocas, seguida de una estabilización suave tanto en entrenamiento como en validación, lo que indica una adaptación controlada sin señales de sobreajuste. Las métricas finales obtenidas fueron:

- Train loss: 0.0022, Validation loss: 0.0044, Test loss (real): 0.0045

- MAE ≈ 0.52 , MSE ≈ 0.27

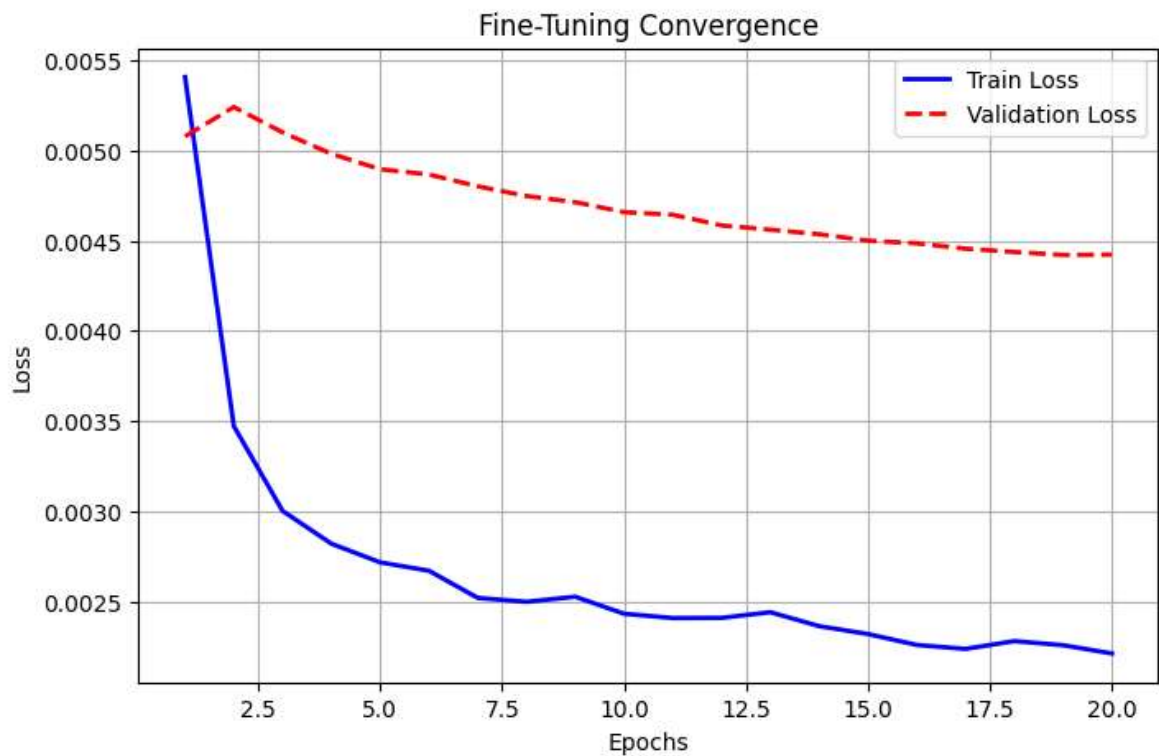


Figura 53. Curva de entrenamiento de fine-tuning con aumento de datos.

Estas métricas reflejan la dificultad inherente del dominio real, donde el ruido no es gaussiano y la señal presenta discontinuidades complejas. Sin embargo, también confirman que el modelo logró ajustarse a las nuevas distribuciones sin degradar su rendimiento.

6.3.6.6 Evaluación del fine-tuning en objetos reales

Para evaluar el impacto real del fine-tuning, se seleccionaron varios objetos experimentales y se compararon las predicciones del modelo base V6 (entrenado solo con datos sintéticos) con su versión refinada V6_F tras el ajuste con datos reales. Este análisis se centró únicamente en observar si el fine-tuning mejoraba la estabilidad y calidad visual de las reconstrucciones.

En general, el modelo V6 ya lograba recuperar la forma global del objeto, pero presentaba limitaciones comunes: suavizado excesivo, bandas residuales, artefactos periódicos característicos del dominio sintético, y un problema de escala, donde la predicción aparecía comprimida en un rango dinámico menor al real. Estas deficiencias se repetían en distintos objetos reales, especialmente en regiones con ruido experimental, sombras o variaciones abruptas de iluminación.

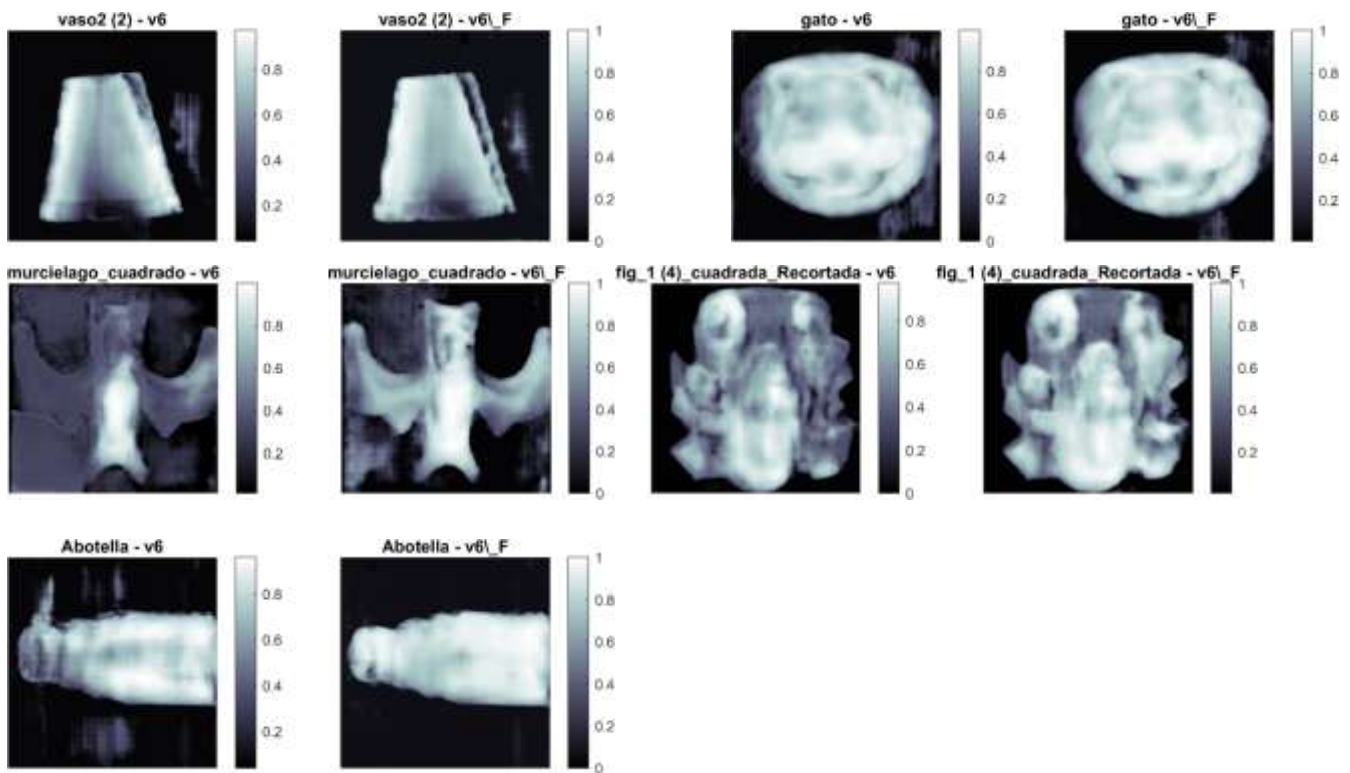


Figura 54. Inferencias V6 vs V6 con fine-tuning con aumento de datos.

Tras aplicar el fine-tuning, las predicciones del modelo V6_F mostraron mejoras claras en varios de los objetos evaluados. En particular, se observó:

- Mayor coherencia espacial, con superficies más uniformes y menos artefactos estructurados.

- Corrección parcial del problema de escala, logrando predicciones con un contraste más cercano al mapa de fase real.
- Reducción de texturas periódicas, que provenían del entrenamiento con datos sintéticos ideales.
- Transiciones más limpias, especialmente en objetos con geometría suave o bordes amplios.

Estas mejoras no aparecieron en todos los objetos con la misma magnitud, pero sí en un número suficiente de casos para considerar que el fine-tuning aporta un beneficio real, complementando las métricas cuantitativas. Por ello, el modelo afinado V6_F se considera más adecuado para operar sobre mapas de fase experimentales en el escenario low-wrap.

6.4 CNN vs métodos tradicionales

La siguiente sección presenta un análisis comparativo entre los métodos tradicionales de desdoblamiento de fase espacial y la mejor arquitectura basada en aprendizaje profundo entrenada en este trabajo. Para ello se evaluaron objetos sintéticos, donde existe un ground truth que permite una comparación cuantitativa, y objetos reales adquiridos mediante perfilometría por proyección de franjas (FPP), cuya complejidad experimental permite valorar la robustez práctica de los algoritmos.

6.4.1 Objetos sintéticos

La comparación entre los métodos tradicionales de desdoblamiento de fase y el mejor modelo de aprendizaje profundo se realizó inicialmente sobre objetos sintéticos, donde se dispone de un ground truth que permite evaluar de manera precisa la calidad del desenrollado. Este análisis integró tanto evaluación visual en 2D y 3D como métricas cuantitativas derivadas de la reconstrucción.

Las Figuras 55 a 58 ilustran diversos ejemplos representativos del comportamiento de los algoritmos tradicionales frente a objetos sintéticos con distintas complejidades geométricas. En objetos con variaciones suaves, algunos métodos clásicos, como Itoh, Constantini o Zhao,

recuperan parcialmente la estructura global del mapa de fase. Sin embargo, en cuanto el objeto presenta gradientes pronunciados, zonas delgadas, regiones multivaluadas o cavidades internas, estos algoritmos comienzan a mostrar limitaciones claras. En la Figura 55, por ejemplo, donde se evaluó un objeto sintético altamente irregular, la mayoría de los métodos tradicionales generaron mapas fragmentados, con discontinuidades abruptas, pérdidas de información y artefactos asociados al orden de propagación. En varios casos, el mapa desenvuelto no conserva la topología original del ground truth y su reconstrucción 3D resulta severamente distorsionada.

Este comportamiento se repite en los objetos de tipo esférico (Figura 55 y Figura 56). Aunque la forma global es más simple, los métodos tradicionales vuelven a presentar escalonamientos, ruido estructural y zonas planas mal reconstruidas, especialmente en métodos basados en gradiente (Ghiglia) o propagación unidireccional (Itoh). Al comparar estos resultados con la reconstrucción tridimensional del ground truth, se observa que la falta de continuidad espacial se refleja en superficies con aristas falsas o discontinuidades no físicas. Asimismo, algunos métodos como PEARLS y STIEPU presentan acumulación de error y deformaciones volumétricas que no corresponden a la geometría real del objeto.

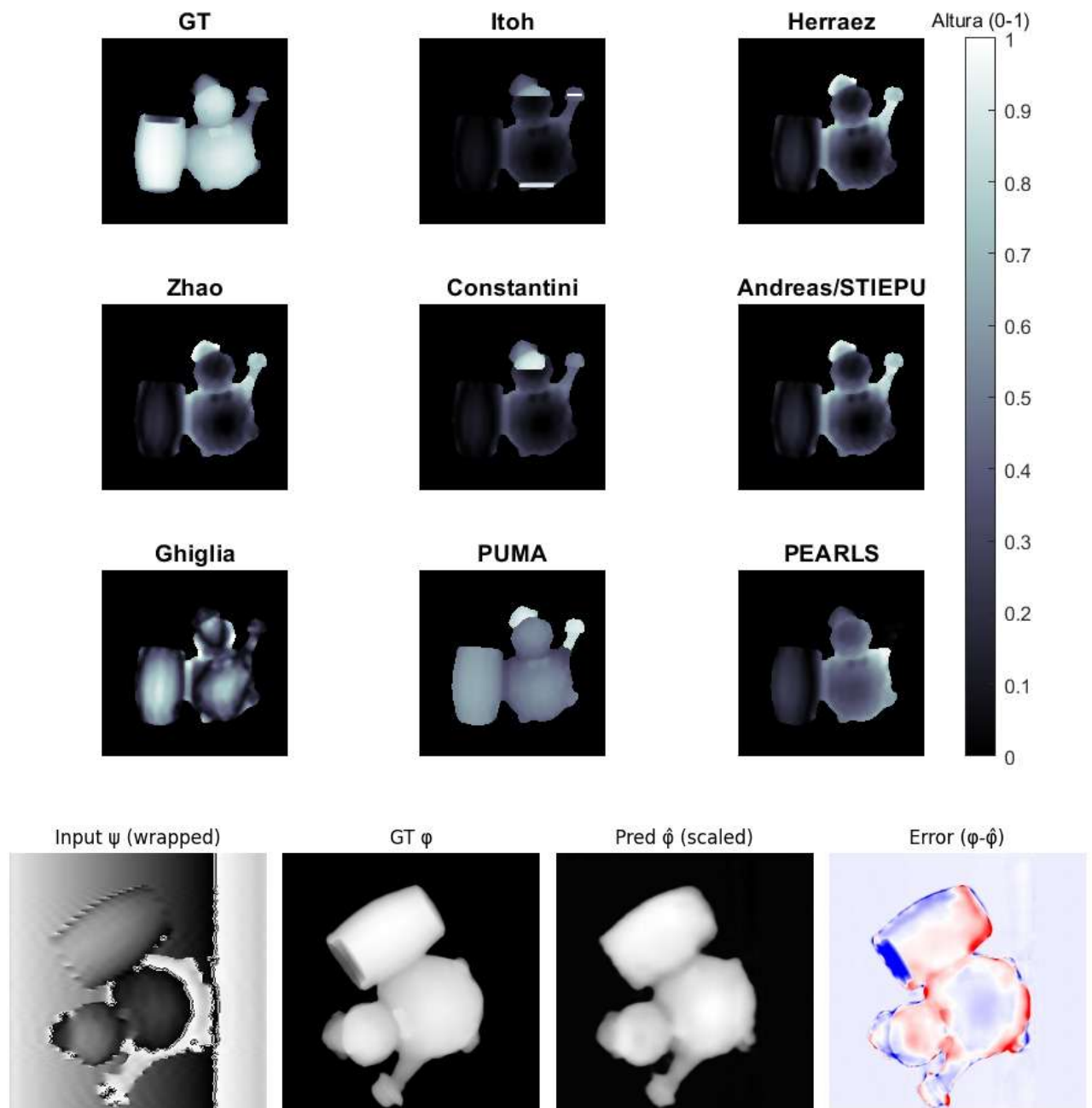
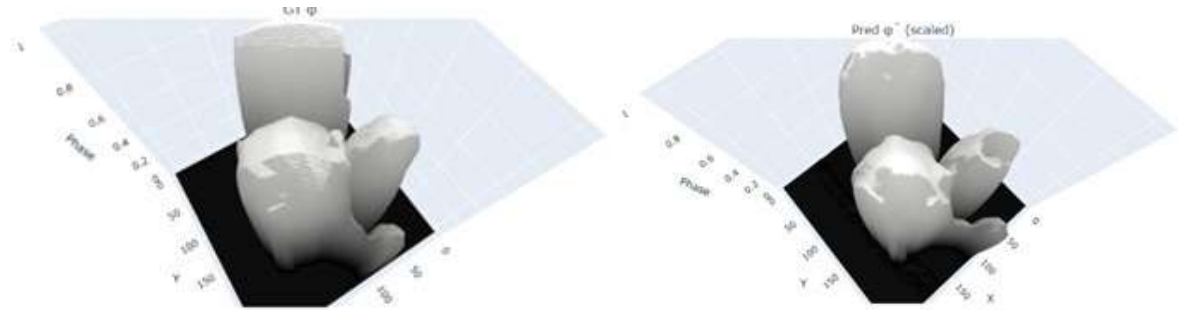


Figura 55. Desdoblamiento de fase métodos tradicionales vs predicción CNN.

Modelo CNN V3

GT

Predicción



Métodos tradicionales

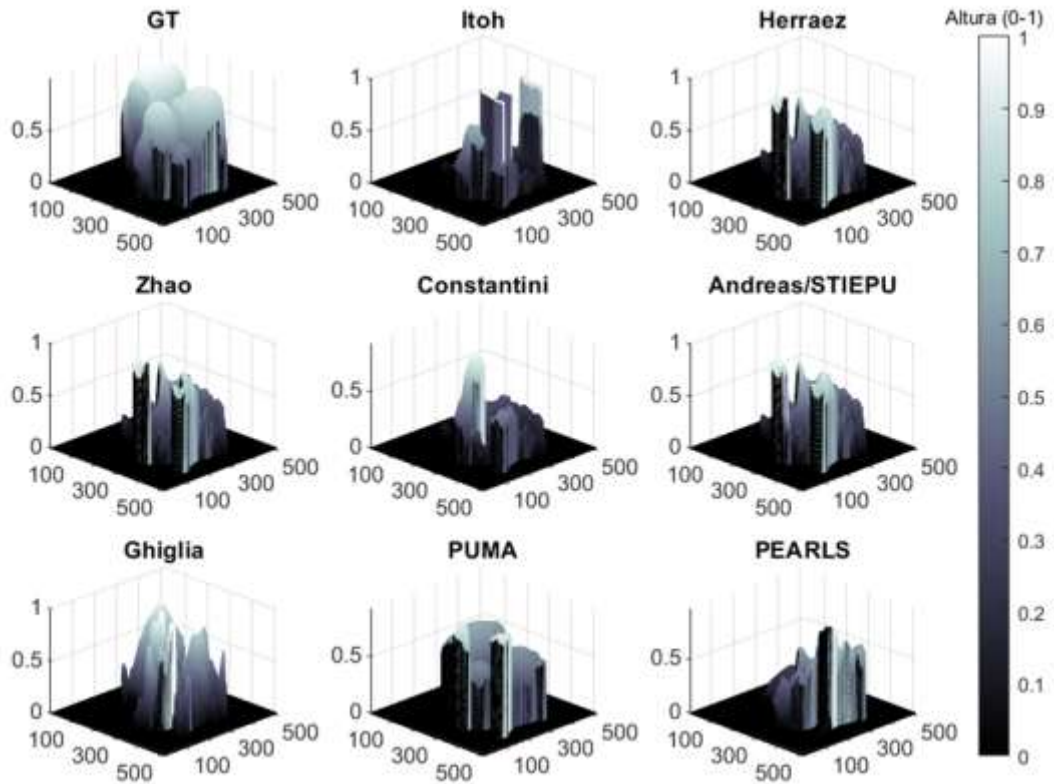
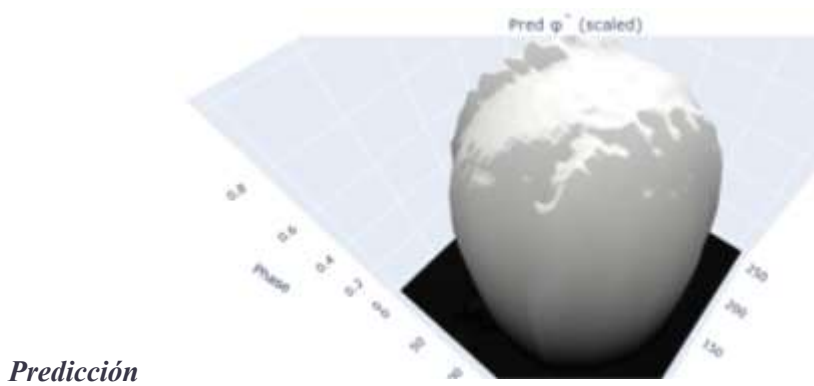
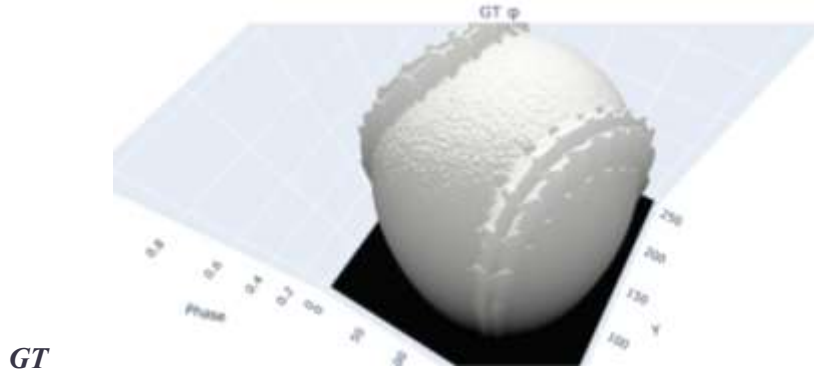


Figura 56. Desdoblamiento de fase métodos tradicionales vs predicción CNN mesh.

En la Figura 58 se aprecia el caso de un objeto sintético tipo pelota de béisbol. A pesar de tratarse de una superficie lisa con patrones periódicos, los algoritmos tradicionales generan reconstrucciones con ruido en las costuras, errores en la curvatura global y distorsiones en la topología circular. En contraste, el modelo de red neuronal convolucional reproduce tanto la forma tridimensional global como detalles locales, preservando continuidad y suavidad incluso en zonas donde el mapa envuelto presenta discontinuidades visibles.

Modelo CNN V3



Métodos tradicionales

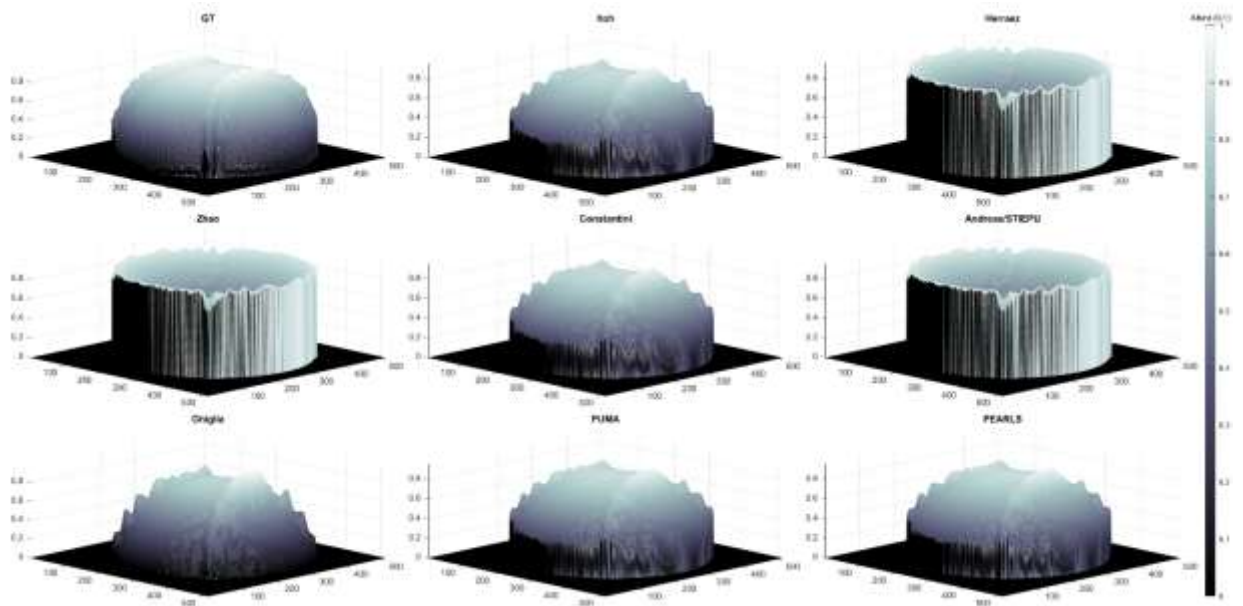
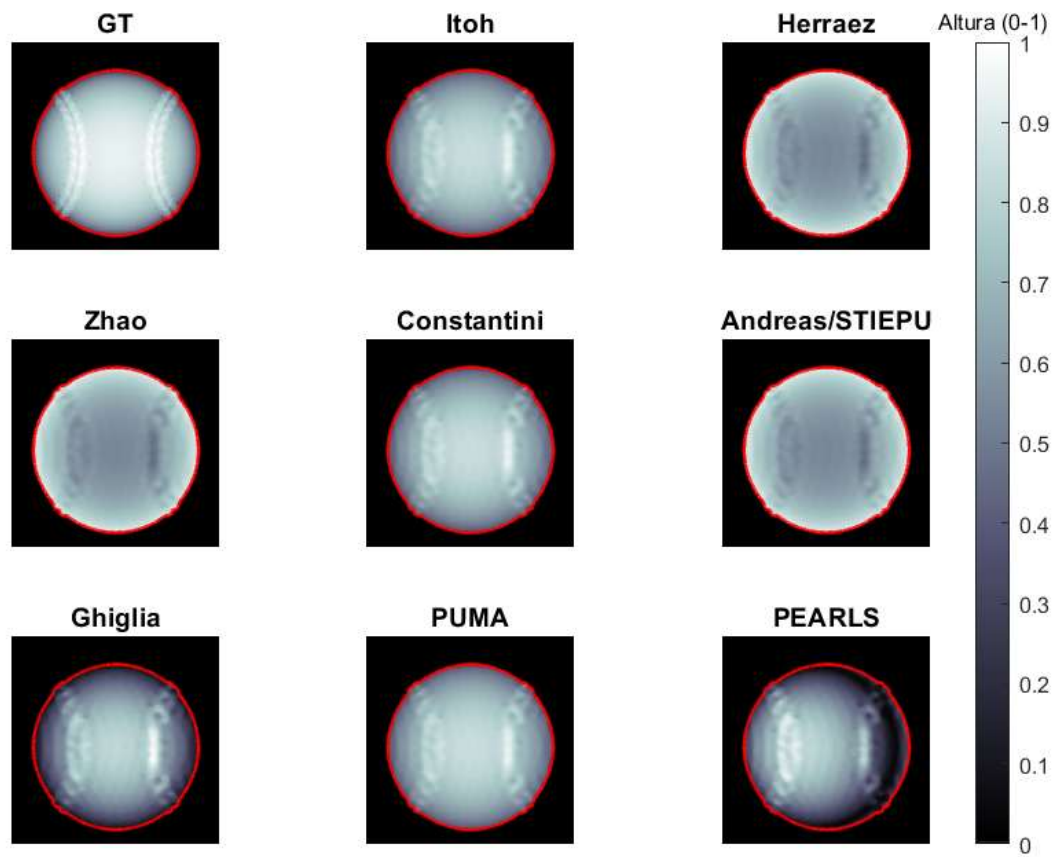


Figura 57. Desdoblamiento de fase CCN vs Métodos tradicionales 3D

Métodos

tradicionales



Modelo

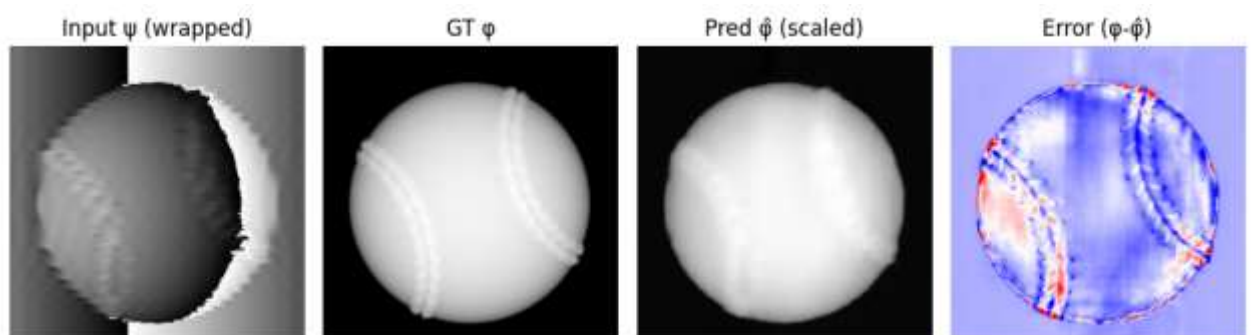


Figura 58. Desdoblamiento de objeto CNN vs métodos tradicionales 2D

Este análisis cualitativo se ve reforzado por los resultados cuantitativos sintetizados en la tabla de desempeño de métodos tradicionales, Tabla 3. Valores promedio Métodos tradicionales. . El mejor resultado obtenido entre estos métodos corresponde a PUMA, con un PSNR promedio de 18.78 dB, un SSIM de 0.886 y un IMMSE cercano a 0.0146. El resto de los algoritmos mostró valores de PSNR entre 14 y 17 dB y errores IMMSE que van desde 0.02 hasta 0.16, reflejando una discrepancia importante respecto al ground truth. Por el contrario, los modelos entrenados con aprendizaje profundo alcanzaron valores notablemente superiores, Tabla 5. En particular, el modelo V17 (U-Net estándar, 256 px, partición mixed-object split) obtuvo un PSNR de 31.88 dB, un SSIM de 0.959 y un IMMSE de 0.0009, superando por un margen amplio todos los métodos tradicionales.

Las reconstrucciones tridimensionales corroboran estas métricas: mientras que los métodos clásicos producen superficies escalonadas o con regiones mal definidas, la CNN genera modelos continuos, suaves y coherentes con la geometría sintética. Los mapas de error ($\phi - \hat{\phi}$) muestran que la discrepancia entre la predicción y el ground truth se concentra únicamente en bordes y zonas con cambios bruscos de fase, manteniendo una correspondencia casi perfecta en el resto del volumen.

En términos globales, los resultados evidencian que los métodos tradicionales dependen fuertemente del tipo de variación del mapa de fase y se ven afectados por ruido, envolventes múltiples y discontinuidades espaciales. En cambio, la CNN es capaz de aprender relaciones no lineales entre el mapa envuelto y el mapa real, generando desenrollados más exactos y consistentes. Esta capacidad se refleja tanto en las métricas cuantitativas, que muestran reducciones del IMMSE superiores a un orden de magnitud, como en la calidad visual y estructural de las reconstrucciones tridimensionales.

La tabla comparativa final resume de manera clara esta diferencia:

Método	PSNR (dB)	SSIM	IMMSE
PUMA (mejor método tradicional)	18.785	0.886	0.0146
U-Net V17 (mejor modelo CNN)	31.88	0.959	0.0009

Tabla 6. Comparación cuantitativa entre mejor método tradicional y modelo CNN

La CNN mejora el PSNR en más de 13 dB, aumenta el SSIM en 0.073 y reduce el IMMSE aproximadamente por un factor de 16, consolidándose como la alternativa más precisa y robusta para el desdoblamiento de fase sintético.

En conjunto, la evaluación realizada demuestra que, para objetos sintéticos, la CNN no solo supera cuantitativamente a los métodos tradicionales, sino que además ofrece reconstrucciones tridimensionales más fieles al ground truth y libres de artefactos estructurales, evidenciando su eficacia para resolver las ambigüedades de fase incluso en escenarios altamente complejos.

6.4.2 Objetos reales

Para los objetos reales adquiridos mediante perfilometría por proyección de franjas, se evaluó el desempeño de los métodos tradicionales y del modelo basado en CNN utilizando las Figuras 55 a 58. Cada una de estas figuras incluye: el objeto con franjas, el mapa de fase envuelto y las reconstrucciones obtenidas por los algoritmos clásicos comparadas con la mejor predicción generada por el modelo entrenado. Este análisis permite valorar la capacidad de generalización de ambas aproximaciones ante ruido físico, variaciones de iluminación y distorsiones inherentes al sistema óptico.

En la Figura 55, correspondiente a un objeto tubular alargado, botella, los métodos tradicionales presentan errores notorios, como discontinuidades longitudinales, zonas incompletas y pérdida de estructura en regiones donde el mapa envuelto contiene gradientes suaves. Itoh y Herraes generan rupturas claras a lo largo del eje del objeto, mientras que Zhao y Constantini muestran regiones donde el desdoblamiento falla parcialmente. Ghiglia y PUMA conservan mejor la continuidad, pero siguen presentando ruido estructural. En comparación, la predicción del modelo basado en CNN representa una superficie más uniforme, continua y con menos artefactos, evidenciando su mayor tolerancia ante las imperfecciones presentes en los datos reales.

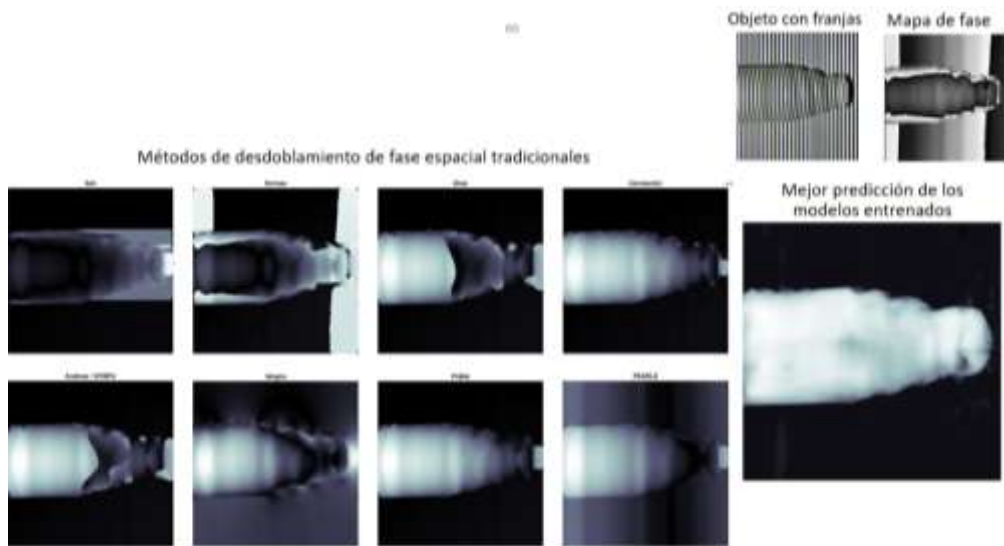


Figura 59. Desdoblamiento de fase métodos tradicionales vs mejor modelo.

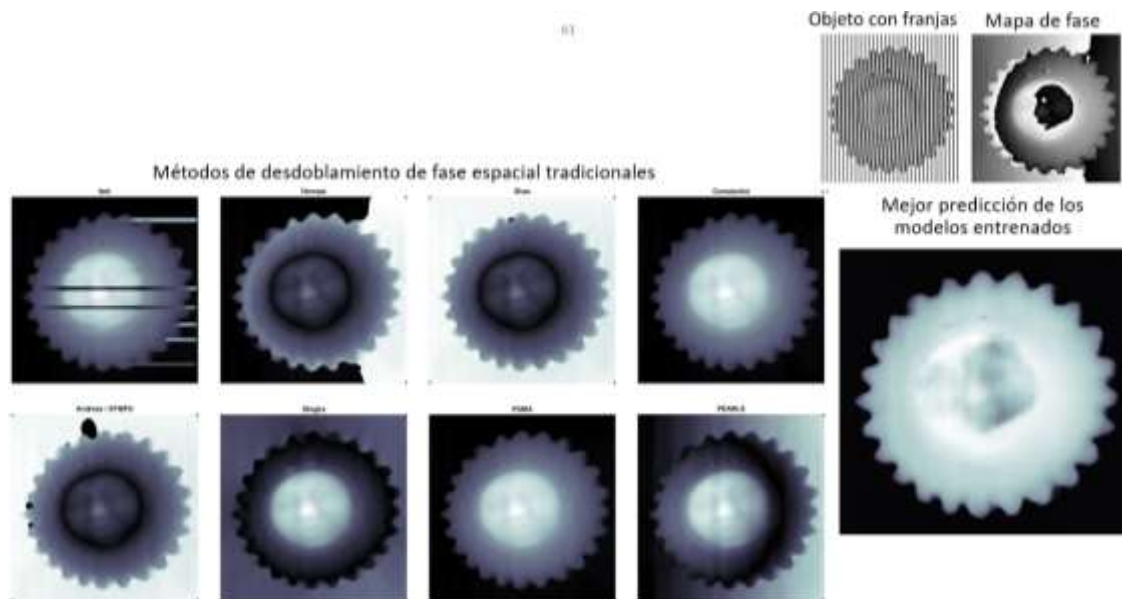


Figura 60. Desdoblamiento de fase métodos tradicionales vs mejor modelo.

La Figura 56 muestra una figura de una escultura que semeja la forma de un sol, un objeto cuya simetría y bordes dentados representan un desafío adicional. En este caso, los métodos tradicionales tienden a suavizar en exceso el contorno o a deformarlo, perdiendo la nitidez de la estructura circular e introduciendo errores en la zona central. Itoh presenta bandas

residuales; Herraez suaviza parte de la estructura; PUMA y PEARLS pierden detalles en los dientes del engrane. La CNN, aunque también suaviza ciertos bordes, preserva de manera más estable la circularidad general y reduce considerablemente el ruido interno, ofreciendo una representación más coherente del objeto.

En la Figura 57, que corresponde a un objeto de geometría irregular con cavidades internas y variaciones bruscas de fase, los métodos tradicionales exhiben errores aún más severos. Se observan regiones completamente deformadas, discontinuidades marcadas y artefactos que distorsionan por completo la estructura real. Algunos algoritmos, como PEARLS y Zhao, fallan al desdoblarse áreas amplias del mapa de fase, mientras que Ghiglia y PUMA presentan oscilaciones intensas y pérdida de volumen. La predicción del modelo entrenado muestra una reconstrucción más estable, recupera en mayor medida las cavidades internas y elimina gran parte del ruido estructural. Aunque se observa un leve suavizado de detalles finos, la forma global del objeto es más fiel en comparación con los métodos clásicos.

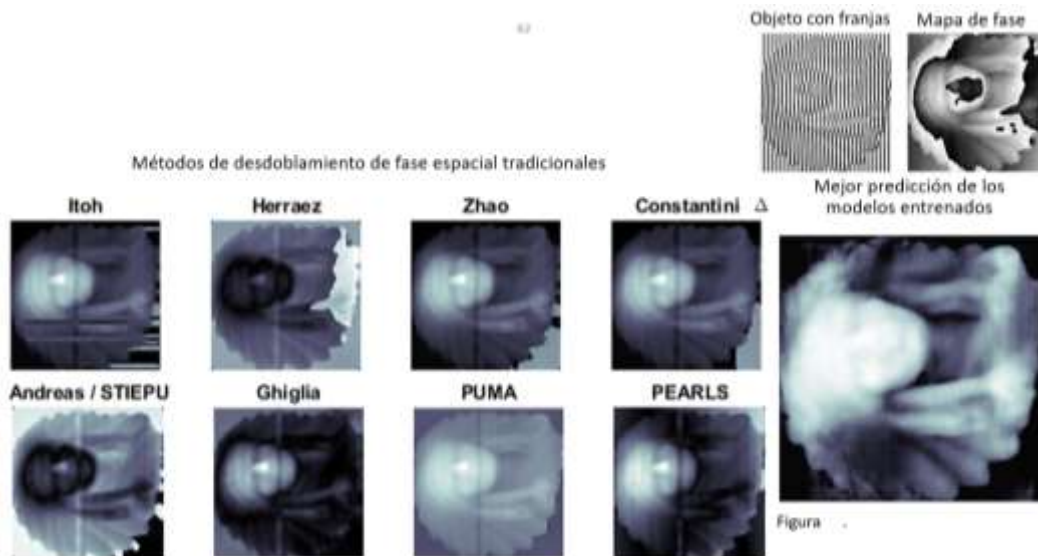


Figura 61. Desdoblamiento de fase métodos tradicionales vs mejor modelo

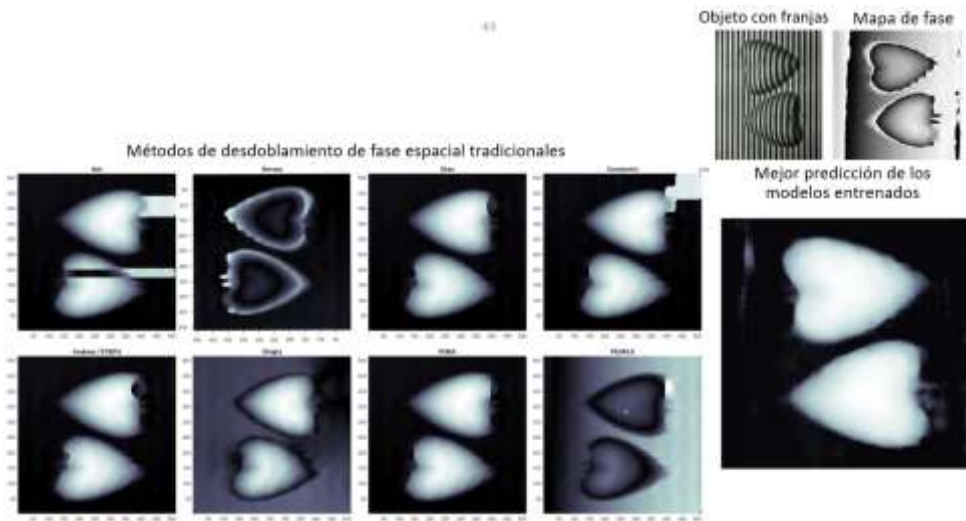


Figura 62. Desdoblamiento de fase métodos tradicionales vs mejor modelo.

La Figura 58 presenta un objeto compuesto por dos lóbulos. En esta geometría, los métodos tradicionales generan discontinuidades internas, regiones mal desenvueltas y pérdida de simetría entre los dos lóbulos; Ghiglia y PEARLS presentan deformaciones volumétricas pronunciadas, mientras que PUMA mantiene parcialmente la forma, aunque con artefactos visibles. El modelo basado en CNN produce un mapa continuo y homogéneo, con mejor conservación de la relación entre los dos lóbulos y una reducción notable del ruido presente en los métodos tradicionales.

Las observaciones derivadas de las Figuras 55 a 58 permiten concluir que, en condiciones reales, los métodos tradicionales presentan una sensibilidad considerable a variaciones experimentales tales como sombras, saturación, interferencias y patrones irregulares de iluminación. Como resultado, sus reconstrucciones presentan artefactos, rupturas y deformaciones que comprometen la estructura del objeto. Por el contrario, el modelo basado en CNN demuestra una mayor capacidad de generalización ante estas condiciones no ideales, logrando reconstrucciones más limpias, continuas y coherentes. Aunque la red suaviza algunos detalles finos, preserva de mejor manera la forma global de los objetos y reduce la presencia de errores estructurales, lo que confirma su mayor robustez frente a los desafíos inherentes a los datos reales.

VII. Conclusiones

El trabajo desarrollado en esta tesis presenta una evaluación integral del desdoblamiento de fase espacial mediante métodos tradicionales y arquitecturas de aprendizaje profundo aplicadas a objetos sintéticos y reales generados a través de perfilometría por proyección de franjas. Los resultados permiten afirmar que la implementación de una red neuronal convolucional mejora de manera significativa el proceso de desdoblamiento, reduciendo el ruido presente en los mapas y disminuyendo el error acumulado durante la reconstrucción tridimensional, lo que confirma la hipótesis planteada desde el inicio. A lo largo de los experimentos realizados se evidenció que los métodos tradicionales, ampliamente utilizados en interferometría y procesamiento de fase, presentan un desempeño adecuado únicamente bajo condiciones ideales. Algoritmos como Herraez, Constantini, Robust2D, PUMA, STIEPU, Ghiglia-Romero o PEARLS mostraron limitaciones claras frente a escenarios con ruido estructurado, gradientes pronunciados, sombras, cambios en iluminación y mapas multi-wrap. En los objetos sintéticos, estas técnicas produjeron artefactos, discontinuidades residuales, suavizado excesivo, pérdida de detalles en zonas de alto gradiente, propagación acumulada del error y una marcada dependencia de la calidad del mapa envuelto. En contraste, los modelos de aprendizaje profundo mostraron una capacidad notable para generalizar y capturar variaciones espaciales complejas incluso cuando los datos presentaban discontinuidades o ruido significativo. Entre las arquitecturas evaluadas, la U-Net estándar alcanzó el mejor desempeño cuantitativo en datos sintéticos, con valores de PSNR y SSIM superiores a todos los métodos tradicionales y con un error IMMSE de orden significativamente menor. Por otro lado, U-Net con SQD-LSTM y U-Net Res-Net demostraron ser más robustas en la evaluación sobre objetos reales, donde las condiciones ópticas imperfectas requieren modelos con mayor capacidad representativa. La arquitectura Multiresolution ofreció un desempeño intermedio cuya estabilidad dependió fuertemente del tipo de mapa utilizado y del particionado del dataset.

El análisis detallado de los datos demostró también que el tamaño espacial de entrada tiene un efecto directo en el desempeño, ya que las imágenes de 512×512 permitieron conservar más detalles y obtener métricas superiores, especialmente en los casos low-wrap. Asimismo, el particionado del dataset se identificó como un factor decisivo: el esquema object-wise split permitió evaluar la verdadera capacidad de generalización hacia objetos no vistos, mostrando resultados más estables que mixed-object split, aun cuando este último consiguiera una pérdida de entrenamiento aparentemente menor.

La evaluación en objetos reales, apoyada en las Figuras 55 a 58, permitió extender el análisis hacia condiciones experimentales no controladas, donde el impacto del ruido físico y de las imperfecciones de adquisición se vuelve evidente. En estos resultados se observa que los métodos tradicionales tienden a romper la continuidad del objeto, generar artefactos, perder información de bordes y producir mapas incompletos. Las reconstrucciones del engrane, del objeto tubular, del objeto irregular y del objeto de lóbulos múltiples muestran deformaciones características en casi todos los métodos clásicos, mientras que la CNN produce mapas más suaves, continuos y estructuralmente coherentes. Aunque el modelo suaviza algunos detalles finos, conserva la forma global de los objetos con mayor fidelidad y es menos afectado por sombras, saturación y patrones de iluminación no uniformes. Esto indica que el modelo aprende representaciones más robustas del campo de fase que los algoritmos basados en gradientes locales, lo que resulta en mejores reconstrucciones en escenarios reales.

Durante el desarrollo de esta investigación surgieron retos importantes que influyeron en los resultados obtenidos. El ruido físico presente en las adquisiciones reales, incluyendo saturación, sombras y variaciones de intensidad, afectó la coherencia de los mapas de fase. La complejidad asociada al desdoblamiento multi-wrap generó discontinuidades que incluso los modelos profundos encontraron difíciles de resolver. El dataset real disponible fue limitado tanto en cantidad como en variedad, lo que restringe el potencial de generalización del modelo y hace más evidente la diferencia entre el comportamiento sintético y el real. La normalización de fase se identificó como un factor altamente sensible: pequeñas variaciones afectaron el patrón de convergencia del modelo. Además, algunas arquitecturas complejas mostraron inestabilidad en la validación cuando la distribución de entrenamiento difería de

la de prueba, y ciertos modelos requirieron tiempos de entrenamiento prolongados y recursos computacionales considerables debido al tamaño de las imágenes y la profundidad de las redes.

El estudio presenta también limitaciones importantes. El conjunto de datos reales fue reducido y con variaciones de ruido considerables, lo que limita la capacidad del modelo para aprender de manera robusta sobre condiciones experimentales diversas. El uso de datos sintéticos, aunque útil para obtener ground truth confiable, no reproduce completamente las imperfecciones de los sistemas ópticos, lo que crea una brecha entre los escenarios sintéticos y reales. El enfoque supervisado exige un ground truth preciso, lo cual no siempre es posible en aplicaciones reales. Además, la evaluación se centró exclusivamente en FPP; técnicas como holografía digital, OCT o interferometría multicanal podrían ofrecer perspectivas adicionales. Tampoco se exploraron arquitecturas más recientes como los transformers o modelos con mecanismos avanzados de atención, que actualmente muestran mejor desempeño en tareas de visión. Finalmente, no se abordó la reconstrucción en tiempo real, que es necesaria para aplicaciones industriales y de inspección continua.

A partir de estos resultados, se identifican diversas áreas de oportunidad para trabajos futuros. La ampliación del dataset real aparece como una necesidad inmediata, incorporando mayor variedad geométrica, diferentes niveles de ruido y condiciones experimentales más heterogéneas. Integrar conocimiento físico mediante modelos physics-informed permitiría obtener redes más interpretables y robustas. El uso de técnicas auto-supervisadas y semi-supervisadas reduciría la dependencia del ground truth sintético, permitiendo aprovechar mejor los datos reales disponibles. Las arquitecturas basadas en transformers, atención espacial o modelos híbridos podrían mejorar la capacidad del modelo para captar dependencias de largo alcance. También es necesario desarrollar arquitecturas especializadas para el desdoblamiento multi-wrap, así como explorar redes ligeras optimizadas para operación en tiempo real. La integración multimodal, combinando información de intensidad, mapas de calidad o imágenes RGB, podría mejorar aún más la reconstrucción. Finalmente, una validación experimental completa que integre el pipeline físico completo

(proyección, captura, reconstrucción y análisis 3D) permitiría evaluar el sistema en un escenario plenamente operativo.

En conjunto, esta tesis permite concluir que las redes neuronales convolucionales representan una alternativa sólida y eficaz para el desdoblamiento de fase espacial, superando a los métodos tradicionales tanto en precisión como en estabilidad. La evidencia obtenida indica que los modelos basados en aprendizaje profundo no solo reducen el ruido y mejoran la continuidad del mapa de fase, sino que también ofrecen reconstrucciones tridimensionales más fieles, incluso en condiciones experimentales adversas. Si bien persisten desafíos relacionados con la disponibilidad de datos reales y la complejidad inherente al problema, los resultados obtenidos sientan una base sólida para el desarrollo de algoritmos más avanzados que permitan mejorar la exactitud, robustez y aplicabilidad del desdoblamiento de fase en aplicaciones científicas e industriales.

VIII. Bibliografía

- Abdul-Rahman, H. S., Gdeisat, M. A., Burton, D. R., Lalor, M. J., Lilley, F., & Moore, C. J. (2007). *Fast and robust three-dimensional best path phase unwrapping algorithm*.
- Bioucas-Dias, J., Katkovnik, V., Astola, J., & Egiazarian, K. (2008). Absolute phase estimation: adaptive local denoising and global unwrapping. *Applied Optics*, 47(29), 5358. <https://doi.org/10.1364/AO.47.005358>
- Bioucas-Dias, J. M., & Valadão, G. (2007). Phase unwrapping via graph cuts. *IEEE Transactions on Image Processing*, 16(3), 698–709. <https://doi.org/10.1109/TIP.2006.888351>
- Chollet, F. (2021). *Deep Learning with Python* (Second Edition). Manning Publications Co. LLC.
- Çiçek, Ø., Abdulkadir, A., Lienkamp, S. S., Brox, T., & Ronneberger, O. (2016). 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 9901, 424–432. https://doi.org/10.1007/978-3-319-46723-8_49
- Costantini, M. (1998). A novel phase unwrapping method based on network programming. *IEEE Transactions on Geoscience and Remote Sensing*, 36(3), 813–821. <https://doi.org/10.1109/36.673674>
- Curioso, W., & Brunette, M. (2020). Inteligencia artificial e innovación para optimizar el proceso de diagnóstico de la tuberculosis. *Revista Peruana de Medicina Experimental y Salud Pública*, 37(3), 554–558. <https://doi.org/10.17843/rpmesp.2020.373.5585>
- de Santos García, R., Gómez Silva, M. J., Domínguez Monferrer, C., Miguélez Garrido, M. H., & Cantero Guisández, J. L. (2022, noviembre). Aplicación de redes neuronales convolucionales a la monitorización del nivel de desgaste de herramientas en sistemas industriales de taladrado. *XV Congreso Iberoamericano en Ingeniería Mecánica*. <https://doi.org/10.5944/bicim2022.217>

- Díaz, J., Lagunas, J., Vieyra, A., Jacuinde, J., & Olivares-Rojas, J. (2022). Dialog modeling of domain specific cognitive assistants. *Avances en Interacción Humano-Computadora*, 7(1), 17–20. <https://doi.org/10.47756/aih.c.y7i1.121>
- Estrada, J. C., Servin, M., & Quiroga, J. A. (2011). Noise robust linear dynamic system for phase unwrapping and smoothing. *Optics Express*, 19(6), 5126. <https://doi.org/10.1364/OE.19.005126>
- Feng, S., Chen, Q., Gu, G., Tao, T., Zhang, L., Hu, Y., Yin, W., & Zuo, C. (2019). Fringe pattern analysis using deep learning. *Advanced Photonics*, 1(02), 1. <https://doi.org/10.1117/1.ap.1.2.025001>
- Ghiglia, D. C., & Romero, L. A. (1994). Robust two-dimensional weighted and unweighted phase unwrapping that uses fast transforms and iterative methods. *Journal of the Optical Society of America A*, 11(1), 107–117. <https://doi.org/10.1364/JOSAA.11.000107>
- Gontarz, M., Dutta, V., & Kujawinska, M. (2023). Deep learning-based phase unwrapping for holographic tomography. En E. Stella, D. Ceglarek, Q. Kemao, & F. Soldovieri (Eds.), *Multimodal Sensing and Artificial Intelligence: Technologies and Applications III* (p. 25). SPIE. <https://doi.org/10.1117/12.2673675>
- Gorthi, S. S., & Rastogi, P. (2010a). Fringe Projection Techniques: Whither we are? *Optics and Lasers in Engineering*, 48(2), 133–140. <https://doi.org/10.1016/j.optlaseng.2009.03.008>
- Gorthi, S. S., & Rastogi, P. (2010b). Fringe Projection Techniques: Whither we are? scale of measurement (mi. En *Optics and Lasers in Engiering* (Vol. 106, Número 123).
- Hanspach, J., Bollmann, S., Grigo, J., Karius, A., Uder, M., & Laun, F. B. (2022). Deep learning-based quantitative susceptibility mapping (QSM) in the presence of fat using synthetically generated multi-echo phase training data. *Magnetic Resonance in Medicine*, 88(4), 1548–1560. <https://doi.org/10.1002/mrm.29265>
- Hartley, R., & Zisserman, A. (2004). *Multiple View Geometry in Computer Vision*. Cambridge University Press.

- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778. <https://doi.org/10.1109/CVPR.2016.90>
- Herráez, M. A., Burton, D. R., Lalor, M. J., & Gdeisat, M. A. (2002). Fast two-dimensional phase-unwrapping algorithm based on sorting by reliability following a noncontinuous path. *Applied Optics*, 41(35), 7437. <https://doi.org/10.1364/AO.41.007437>
- Huntley, J. M., & Saldner, H. O. (1997). Shape measurement by temporal phase unwrapping: comparison of unwrapping algorithms. En *Meas. Sci. Technol* (Vol. 8). <http://iopscience.iop.org/0957-0233/8/9/005>
- Isensee, F., Jaeger, P. F., Kohl, S., Petersen, J., & Maier-Hein, K. H. (2021). nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, 18, 203–211. <https://doi.org/10.1038/s41592-020-01008-z>
- Itoh, K. (1982). Analysis of the phase unwrapping algorithm. *Applied Optics*, 21(14), 2470. <https://doi.org/10.1364/AO.21.002470>
- Izaurieta, F., & Saavedra, C. (s/f). *Redes Neuronales Artificiales*.
- Jabir, B., & Falih, N. (2021). Dropout, a basic and effective regularization method for a deep learning model: A case study. *Indonesian Journal of Electrical Engineering and Computer Science*, 24(2), 1009–1016. <https://doi.org/10.11591/ijeecs.v24.i2.pp1009-1016>
- Ji, S., Xu, W., Yang, M., & Yu, K. (2013). 3D Convolutional neural networks for human action recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1), 221–231. <https://doi.org/10.1109/TPAMI.2012.59>
- Jiménez-Murillo, D. (2023). Segmentación de tejido cerebral mediante redes neuronales convolucionales basadas en U-Net. *RISTI - Revista Ibérica de Sistemas e Tecnologias de Informação*, 52, 5–24. <https://doi.org/10.17013/risti.52.5-24>

- Judge, T. R., & Bryanston-Cross, & P. J. (1994a). A Review of Phase Unwrapping Techniques in Fringe Analysis. En *Optics and Lasers in Engineering* (Vol. 21).
- Judge, T. R., & Bryanston-Cross, P. J. (1994b). A Review of Phase Unwrapping Techniques in Fringe Analysis. *Optics and Lasers in Engineering*, 21(3), 199–239. [https://doi.org/10.1016/0143-8166\(94\)90087-6](https://doi.org/10.1016/0143-8166(94)90087-6)
- Kandel, E. R., Koester, J., Mack, S., & Siegelbaum, S. (2021). *Principles of neural science* (E. R. Kandel, KOESTER JOHN D., MACK SARAH H., & SIEGELBAUM STEVEN A., Eds.; Sixth). McGraw Hill.
- Kemao, Q. (2007). Windowed Fourier transform for fringe pattern analysis. *Applied Optics*, 46(26), 6644–6651. <https://doi.org/10.1364/AO.46.006644>
- Köl, S. (2016). *Computer Vision Metrics: Survey, Taxonomy, and Analysis*. Springer.
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (s/f). *ImageNet Classification with Deep Convolutional Neural Networks*. <http://code.google.com/p/cuda-convnet/>
- Kurian, D., Hagberg, G. E., Scheffler, K., & Paul, J. S. (2024). A predictor–corrector phase unwrapping algorithm for temporally undersampled gradient-echo <sc>MRI</sc>. *Magnetic Resonance in Medicine*, 91(4), 1707–1722. <https://doi.org/10.1002/mrm.29964>
- Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. En *Nature* (Vol. 521, Número 7553, pp. 436–444). Nature Publishing Group. <https://doi.org/10.1038/nature14539>
- Li, W., Avram, A. V., Wu, B., Xiao, X., & Liu, C. (2014). Integrated Laplacian-based phase unwrapping and background phase removal for quantitative susceptibility mapping. *NMR in Biomedicine*, 27(2), 219–227. <https://doi.org/10.1002/nbm.3056>
- Liu, T., Wisnieff, C., Lou, M., Chen, W., Spincemaille, P., & Wang, Y. (2013). Nonlinear formulation of the magnetic field to source relationship for robust quantitative susceptibility mapping. *Magnetic Resonance in Medicine*, 69(2), 467–476. <https://doi.org/10.1002/mrm.24272>

- Malacara, D., Malacara, Z., & Servn, Manuel. (2005). *Interferogram analysis for optical testing*. Taylor & Francis.
- Martinez-Carranza, J., Falaggis, K., & Kozacki, T. (2017). Fast and accurate phase-unwrapping algorithm based on the transport of intensity equation. *Applied Optics*, 56(25), 7079. <https://doi.org/10.1364/AO.56.007079>
- Niño-Carmona, C., Chero, M., Ortiz, E., Julca, J., & Ipanaqué, C. (2021). Red neuronal convolucional usando VHDL para entrenar un clasificador de objetos en una imagen. *Revista de la Universidad del Zulia*, 12(32), 240–260. <https://doi.org/10.46925//rdluz.32.16>
- Olmedo, S., Ortega, P., Hurtado, G., Morales, M., & Arreguin, R. (s/f). *IMPLEMENTACIÓN DE ALGORITMOS DE DESDOBLAMIENTO DE FASE PARA RECONSTRUCCIÓN TRIDIMENSIONAL DE OBJETOS UTILIZANDO MATLAB*.
- Olmedo, S., Ortega, P., Hurtado, G., Morales, M., & Arreguín, R. (2020). *Implementación de algoritmos de desdoblamiento de fase para reconstrucción tridimensional de objetos utilizando MATLAB*.
- Perera, M. V., & De Silva, A. (2020). Phase Unwrapping Using Spatial Quad-Directional LSTM Networks. *IEEE Transactions on Geoscience and Remote Sensing*, 58(7), 5111–5124. <https://doi.org/10.1109/TGRS.2019.2958924>
- Perera, M. V., & de Silva, A. (2021). A joint convolutional and spatial quad-directional LSTM network for phase unwrapping. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, 2021-June*, 4055–4059. <https://doi.org/10.1109/ICASSP39728.2021.9414748>
- Purves, D., & others (Eds.). (2004). *NEUROSCIENCE: Third Edition* (3rd ed.). Sinauer Associates, Inc.
- Quiroga, F. (s/f). *Medidas de invarianza y equivarianza a transformaciones en redes neuronales convolucionales*. <https://doi.org/10.35537/10915/90903>

- Robinson, S., Schödl, H., & Trattnig, S. (2014). A method for unwrapping highly wrapped multi-echo phase images at very high field: UMPIRE. *Magnetic Resonance in Medicine*, 72(1), 80–92. <https://doi.org/10.1002/mrm.24897>
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 9351, 234–241. https://doi.org/10.1007/978-3-319-24574-4_28
- Schmidhuber, J. (2015). Deep Learning in neural networks: An overview. En *Neural Networks* (Vol. 61, pp. 85–117). Elsevier Ltd. <https://doi.org/10.1016/j.neunet.2014.09.003>
- Seitz, S. M., Curless, B., Diebel, J., Scharstein, D., & Szeliski, R. (2006). *A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms*. <http://vision.middlebury.edu/mview>.
- Servin, M., Quiroga, J. A., & Padilla, J. M. (Eds.). (2014). *Fringe Pattern Analysis for Optical Metrology: Theory, Algorithms, and Applications* (1st ed.). Wiley-VCH. <https://doi.org/10.1002/9783527681075>
- Su, X., & Chen, W. (2004). Reliability-guided phase unwrapping algorithm: A review. *Optics and Lasers in Engineering*, 42(3), 245–261. <https://doi.org/10.1016/j.optlaseng.2003.11.002>
- Sun, J., Wauthier, C., Stephens, K., Gervais, M., Cervone, G., La Femina, P., & Higgins, M. (2020). Automatic Detection of Volcanic Surface Deformation Using Deep Learning. *Journal of Geophysical Research: Solid Earth*, 125(9). <https://doi.org/10.1029/2020JB019840>
- Sun, Y., Xue, B., Zhang, M., & Yen, G. G. (2020a). Completely Automated CNN Architecture Design Based on Blocks. *IEEE Transactions on Neural Networks and Learning Systems*, 31(4), 1242–1254. <https://doi.org/10.1109/TNNLS.2019.2919608>

- Sun, Y., Xue, B., Zhang, M., & Yen, G. G. (2020b). Evolving Deep Convolutional Neural Networks for Image Classification. *IEEE Transactions on Evolutionary Computation*, 24(2), 394–407. <https://doi.org/10.1109/TEVC.2019.2916183>
- Sun, Y., Yu, Y., & Wang, W. (2018). Moiré Photo Restoration Using Multiresolution Convolutional Neural Networks. *IEEE Transactions on Image Processing*, 27(8), 4160–4172. <https://doi.org/10.1109/TIP.2018.2829898>
- Takeda, M., Ina, H., & Kobayashi, S. (1982). Fourier-transform method of fringe-pattern analysis for computer-based topography and interferometry. *Journal of the Optical Society of America*, 72(1), 156–160. <https://doi.org/10.1364/JOSA.72.000156>
- Takeda, M., & Mutoh, K. (1983). Fourier transform profilometry for the automatic measurement of 3-D object shapes. *Applied Optics*, 22(24), 3977–3982. <https://doi.org/10.1364/AO.22.003977>
- Tran, D., Bourdev, L., Fergus, R., Torresani, L., & Paluri, M. (2015). Learning Spatiotemporal Features with 3D Convolutional Networks. *2015 IEEE International Conference on Computer Vision (ICCV)*, 4489–4497. <https://doi.org/10.1109/ICCV.2015.510>
- Vidal, A., Oviol, R., & Ferrando, M. (1999). *Método Multigrad para Phase Unwrapping en Interferometría SAR*.
- Wang, K., Kemaq, Q., Di, J., & Zhao, J. (2022). Deep learning spatial phase unwrapping: a comparative review. *Advanced Photonics Nexus*, 1(01). <https://doi.org/10.1117/1.apn.1.1.014001>
- Wang, K., Li, Y., Kemaq, Q., Di, J., & Zhao, J. (2019). One-step robust deep learning phase unwrapping. *Optics Express*, 27(10), 15100. <https://doi.org/10.1364/oe.27.015100>
- Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Transactions on Image Processing*, 13(4), 600–612. <https://doi.org/10.1109/TIP.2003.819861>

- Zhang, Y., Cao, Y., Wu, H., & Li, C. (2024). Spatial phase unwrapping algorithm based on multi-anchors bidirectional location and suppression. *Optical Engineering*, 63(01). <https://doi.org/10.1117/1.OE.63.1.014101>
- Zhao, H., Chen, W., & Tan, Y. (1994). Phase-unwrapping algorithm for the measurement of three-dimensional object shapes. *Applied Optics*, 33(20), 4497. <https://doi.org/10.1364/AO.33.004497>
- Zhao, Z., Zhou, M., Du, Y., Li, J., Fan, C., Zhang, X., Wei, X., & Zhao, H. (2022). Robust phase unwrapping algorithm based on Zernike polynomial fitting and Swin-Transformer network. *Measurement Science and Technology*, 33(5), 055002. <https://doi.org/10.1088/1361-6501/ac4ac2>
- Zhou, D. X. (2020). Universality of deep convolutional neural networks. En *Applied and Computational Harmonic Analysis* (Vol. 48, Número 2, pp. 787–794). Academic Press Inc. <https://doi.org/10.1016/j.acha.2019.06.004>
- ZHOU, S., BAI, L., YANG, Y., WANG, H., & FU, K. (2019). A Depthwise Separable Network for Action Recognition. *DEStech Transactions on Computer Science and Engineering, cisnrc*. <https://doi.org/10.12783/dtcse/cisnrc2019/33352>
- Zhu, S., Zang, Z., Wang, X., Wang, Y., Wang, X., & Liu, D. (2021). Phase unwrapping in ICF target interferometric measurement via deep learning. *Applied Optics*, 60(1), 10. <https://doi.org/10.1364/AO.405893>

IX. Anexos

9.1 Registro del protocolo de investigación



DIPFI
POSGRADO
INGENIERÍA

UNIVERSIDAD AUTÓNOMA DE QUERÉTARO
FACULTAD DE INGENIERÍA
DIVISIÓN DE INVESTIGACIÓN Y POSGRADO

**REGISTRO DEL PROTOCOLO DE INVESTIGACIÓN
DEL ESTUDIANTE DE POSGRADO**

Ley 2 Expediente electrónico en línea para la División	No. Registro de Proyecto*	15033
	Fecha de Registro*	12 - NOV - 2024
	Fecha de inicio de proyecto:	10/12/2024
	Fecha de término de proyecto:	01/12/2025
1. DATOS DEL SOLICITANTE		
No. de expediente:	32593	
Apellido Paterno	Apellido Materno	Nombre(s)
VILLAFUERTE	RUZ	KARLA LIZETH
Dirección:		
Calle y número	Colonia	C.P.
Yura #200-7	Arboledas	76140
Estado	Teléfono (incluirla)	Correo Electrónico
22	4661157376	lvillafuerte@gmail.com



2. DATOS DEL PROYECTO		
Facultad:	INGENIERÍA	
Programa:	MAESTRIA EN CIENCIAS EN INTELIGENCIA ARTIFICIAL	
Tema específico del proyecto:	Evaluación del estado de arte en el proceso de implementación de una especialización sobre neuronales	
 PEDRAZA ORTEGA JESUS CARLOS Director de tesis	 TOVAR ARRAGA SAUL Coordinador de programa	 VILLAFUERTE RUZ KARLA LIZETH Alumno
 Dr. Juan Carlos Hernández Jefe de División de Investigación y Posgrado de la Fac. de Ing.	 Dra. María de la Luz Prietz Rectora de Fac. Ing.	 Dr. Manuel Toledano Ayala Director de Investigación y Posgrado UAQ



C. U., 21 de octubre de 2024

Karla Lizeth Villafuerte Ruiz
Estudiante de Maestría en Ciencias en Inteligencia Artificial
Expediente 328949

Presente

El Comité de Ética Aplicada a la Investigación de la Facultad de Ingeniería ha revisado el protocolo del trabajo de tesis:

CEAIFI-148-2024-TP

**Reducción del ruido en el proceso de desdoblamiento
de fase espacial mediante redes neuronales**


Con apego a los lineamientos éticos de beneficencia, no maleficencia, justicia y autonomía, este comité ha dado el siguiente dictamen:

Aprobado éticamente




El presente dictamen tiene vigencia de un año a partir de su fecha de emisión.

Sirva esta carta para los fines académicos que al interesado convengan.

Atentamente
"El ingenio para crear, no para destruir"


Dra. Aurora Fermat Díaz
Presidente del CEAIFI
afemat@uaq.mx

9.2 Inglés

		UNIVERSIDAD AUTÓNOMA DE QUERÉTARO FACULTAD DE LENGUAS Y LETRAS	
<p>A QUIEN CORRESPONDA:</p>			
<p>La que suscribe, Directora de la Facultad de Lenguas y Letras, hace C O N S T A R que</p>			
<p>VILLAFUERTE RUIZ KARLA LIZETH</p>			
<p>Presentó el Examen de Manejo de la Lengua efectuado el día veinticinco de agosto de dos mil veinticinco, en el cual obtuvo la siguiente calificación:</p>			
<p>8+</p>			
<p>Se extiende la presente a petición de la parte interesada, para los fines escolares y legales que le convengan, en el Campus Aeropuerto de la Universidad Autónoma de Querétaro, el día diecisiete de septiembre de dos mil veinticinco.</p>			
	<p>Atentamente, "Enlazar Culturas por la Palabra"</p>		
			
	<p>DRA. MA. DE LOURDES RICO CRUZ</p>		
<p>MLRC/mgoa*CL*FLL-C.-1842</p>			
	<p>CRECER EN LA DIVERSIDAD</p>	<p>📞 442 192 12 00 EXT. 61010</p>	<p>📍 Campus Aeropuerto, Avda. Vía Fray Agustín Serra s/n, Santiago de Querétaro, Qro. México. C.P. 76145.</p>
			<p>FOLIO: 8206</p>

9.3 Retribución social



Gobierno de México
Secretaría de Educación Pública

Educación
Subsecretaría de Educación Media Superior
Dirección General de Educación Tecnológica
Agropecuaria y Ciencias del Mar
Centro de Bachillerato Tecnológico Agropecuario No. 123



Cerritos, San Luis Potosí, 7 de mayo del 2025

MESA: DIRECCION
OFICIO: 64-A/2025
ASUNTO: RECONOCIMIENTO

A QUIEN CORRESPONDA
PRESENTE:

El **CBTA 123** y su comunidad educativa a través del director del plantel otorgan el presente:

RECONOCIMIENTO

a la **Ing. Biomédica Karla Lizeth Villafuerte Ruiz** por su valiosa participación como conferencista en la clase de Tutoría donde a través de una reunión virtual utilizando la app de Google Meet compartió su experiencia sobre la elección de su profesión, así como el impartir la **Conferencia: Introducción a la inteligencia artificial**, buscando acercar a nuestros estudiantes a esta reciente tecnología.

La Ing. Biomédica Karla Lizeth Villafuerte Ruiz, a través de esta conferencia, inspira a nuestros estudiantes a tomar decisiones informadas sobre su futuro profesional y a valorar las nuevas tecnologías en el ámbito educativo.

Por su tiempo, esfuerzo y la valiosa aportación que hizo a nuestra comunidad educativa, le expresamos nuestro más sincero agradecimiento y reconocimiento.

ATENTAMENTE





M.E. JOSE PABLO CARRIZALES GALVAN

NOMBRE Y FIRMA DEL DIRECTOR



2025
Año de
La Mujer
Indígena

Carretera Cerritos - Villa Juárez km 0.8 Carretera San Luis Potosí, Cerritos C.P. 78440 San Luis Potosí.
Tel: 489 95 5 20 55 y 489 95 5 42 27 www.cbta123.edu.mx

9.4 Artículo

