Universidad Autónoma de Querétaro
Facultad de ingeniería
Maestría en ciencias (Inteligencia Artificial)

Reducción del error en la reconstrucción 3D mediante perfilometría de fourier por Redes NeuronalesConvolucionales

Tesis

Que como parte de los requisitos para obtener el Grado de

Maestra en ciencias de la Inteligencia Artificial

Presenta

Mercy Yureymi García Echevarria

Dirigido por

Dr. Jesús Carlos Pedraza Ortega

Dr. Jesus Carlos Pedraza Ortega
Presidente

MCIA. Francisco Emiliano Aguayo Serrano
Secretario

Dr. Juan Manuel Ramos Arreguín
Vocal

Dr. Saul Tovar Arriaga
Suplente

Dr. Andras Takacs
Suplente

REPOSITORIO INSTITUCIONAL DGBSDI-UAQ

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.

Centro Universitario, Querétaro, Qro. México
Fecha de aprobación por el Consejo Universitario (mes y año)

A mis papás, por ser mi mayor ejemplo a seguir, por su amor incondicional, su esfuerzo constante y por creer en mi en cada paso de este camino. Este logro es tan suyo como mío.

A mi hermano, por su compañía incondicional, su apoyo sincero y por ser siempre un motivo para seguir adelante.

A mis amigos, por los momentos compartidos, los ánimos en los días difíciles y por recordarme la importancia de preservar.

A mi asesor, por su guía experta, su tiempo, su paciencia y por compartir sus conocimientos para hacer posible la culminación de este proyecto.

A la Secretaria de Educación, Ciencia, Tecnología e Innovación por el respaldo y la oportunidad brindada para el desarrollo de esta investigación.

Agradecimientos

Quiero agradecer a todas las personas que han estado apoyándome en toda esta etapa de mi formación académica

A mis papás por su apoyo constante, su amor incondicional, porque creen en mi siempre, por su comprensión y respaldo en cada momento del camino, por su esfuerzo constante son el cimiento de cada uno de mis logros, también quiero agradecer a mi hermano por estar presente en cada uno de esos logros.

A mi asesor, el Dr. Jesus Carlos Pedraza por su confianza, orientación, paciencia y acompañamiento a lo largo de esta investigación, también por motivarme a dar lo mejor de mi.

Al cuerpo de sinodales por sus invaluables aportes y comentarios, los cuales enriquecieron de manera significativa el alcance del proyecto

A mis maestros de la maestría por compartirme sus conocimientos, experiencias y tiempo colaborando de manera significativa a mi formación académica.

A mis amigos por el apoyo mutuo, compañía, comprensión, por el intercambio cotidiano, haciendo mas llevadero este camino dándome ánimos en los momentos que mas necesitaba

Por ultimo a la Secretaria de Ciencias, Humanidades, Tecnología e innovación por el apoyo económico y la beca de posgrado otorgada durante la realización de este programa académico.

Abreviaturas y siglas

Inteligencia artificial

- **CNN** - Convolutional Neural Network
- **UHRNet** - U-shaped High-Resolution Network
- **SFNet** - Symmetric Fusion Network
- **Hformer** - Hierarchical Transformer Network
- **FPN**- Feature Pyramid Network
- **FAM** - Feature Aggregation Module
- **CAT** - Cross-Attention Transformer
- **ReLU** - Rectified Linear Unit
- **GELU** - Gaussian Error Linear Unit
- **GPU** - Graphics Processing Unit

Técnicas ópticas y metrología 3D

- **FTP**- Fourier Transform Profilometry
- **FPP**- Fringe projection Profilometry
- **PU**- Phase Unwrapping
- **FT2D**- Two-Dimensional Fourier Transform
- **IFT**- Inverse Fourier Transform

Entrenamiento

- **GT**- Ground Truth
- **MSE**- Mean Squared Error
- **RMSE**- Root Mean Squared Error
- **MAE**- Mean Absolute Error
- **SSIM**- Structural Similarity Index Measure
- **IMMSE**- Image Mean Squared Error
- **PSNR**- Peak Signal-to-Noise Ratio

Resumen

La reconstrucción tridimensional por perfilometria por Transformada de Fourier, es una técnica óptica sin contacto es fundamental en el área de la metrología debido a su rapidez de adquisición el cual va a partir de un único patrón de franjas proyectadas, sin embargo el flujo analítico se ve restringido por la atenuación de los detalles finos durante el filtrado espectral por otro lado los algoritmos de desenvolvimiento de fase sufre limitaciones ante las sombras, oclusiones o discontinuidades abruptas. Con el fin de mitigar dichas restricciones se desarrollo e implemento la FTP utilizando CNN para poder reducir el error en la reconstrucción tridimensional y elevar tanto la precisión como la eficiencia del proceso. De manera metodológica, se implemento el algoritmo tradicional FTP en MATLAB en donde se evaluaron distintas geometrías de filtrado en frecuencia junto con los algoritmos de desenvolvimiento de fase. También se genero una base de datos sintética en Blender la cual esta compuesta por objetos con franjas proyectadas y Ground Truth sin sesgos experimentales. Con estos datos se entrenaron tres arquitecturas de redes neuronales profundas (UHRNet, SFNet y Hformer) esto se hizo con ayuda de google colab con una tarjeta gráfica NVIDIA A100 mediante una función de perdida compuesta, esto formulo un mapeo directo de franja a profundidad sorteando la ambigüedad periódica modular de 2π. Los resultados cuantitativos en el conjunto frente al método tradicional reduciendo el error cuadrático medio hasta tres ordenes de magnitud. La red UHRNet consolidó el desempeño mas sobresaliente al alcanzar un IMMSE de 0.000224, un PSNR de 37.38 dB y una similitud estructural, preservando con alta fidelidad, cavidades y geometrías.

Palabras clave: Perfilometría de Fourier, Reconstrucción tridimensional, Redes Neuronales Convolucionales, Proyección de franjas, Desenvolvimiento de fase.

Abstract

Three-dimensional reconstruction by Fourier Transform profilometry, a non-contact optical technique, is fundamental in the area of metrology due to its speed of acquisition, which is based on a single pattern of projected fringes. However, the analytical flow is restricted by the attenuation of fine details during spectral filtering. On the other hand, phase development algorithms suffer limitations due to shadows, occlusions or abrupt discontinuities. In order to mitigate these restrictions, FTP was developed and implemented using CNN to reduce the error in the three-dimensional reconstruction and increase both the precision and efficiency of the process. Methodologically, the traditional FTP algorithm was implemented in MATLAB where different frequency filtering geometries were evaluated along with the phase unwrapping algorithms. A synthetic database was also generated in Blender which is composed of objects with projected stripes and Ground Truth without experimental biases. With this data, three deep neural network architectures were trained (UHRNet, SFNet and Hformer). This was done with the help of Google Colab with an NVIDIA A100 graphics card using a compound loss function, this formulated a direct mapping from stripe to depth, avoiding the modular periodic ambiguity of 2π. The quantitative results in the set compared to the traditional method reducing the mean square error up to three orders of magnitude.

The UHRNet network consolidated the most outstanding performance by achieving an IMMSE of 0.000224, a PSNR of 37.38 dB and a structural similarity, preserving cavities and geometries with high fidelity.

Keywords: Fourier Profilometry, 3D Reconstruction, Convolutional Neural Networks, Fringe Projection, Phase Unwrapping.

Índice

Índice de cuadros **III**

Índice de figuras **III**

I Introducción **2**
1.1 Antecedentes . . . 3
1.2 Planteamiento del problema . . . 5
1.3 Justificación . . . 6
1.4 Hipótesis . . . 7
1.5 Objetivos . . . 7
1.6 Alcances y limitaciones . . . 8
1.7 Organización de la tesis . . . 9

II Reconstrucción Tridimensional **11**
2.1 Transformada de Fourier . . . 13
2.2 Transformada Discreta de Fourier Bidimensional . . . 14
2.3 Teorema de la Convolución . . . 15

III Perfilometría por Transformada de Fourier (FTP) **16**
3.1 Modelado Espectral del Patrón de Franjas . . . 16
3.2 Proceso Analítico y Extracción de Fase . . . 17

IV Algoritmos de desenvolvimiento de fase (Phase Unwrapping) **19**
4.1 Clasificación de Algoritmos de desenvolvimiento . . . 19
4.1.1 Algoritmos Locales (Seguimiento de Trayectoria) . . . 20
4.1.2 Algoritmos Globales (Optimización o norma mínima): . . . 21
4.1.3 Algoritmos Especiales . . . 22

V Inteligencia Artificial en Reconstrucción 3D **24**
5.1 Demodulación de la Fase Espacial: El enfoque Híbrido Físico-Convolucional 24
5.1.1 Arquitectura de Dos Etapas (CNN1 y CNN2) . . . 24
5.2 Metodologías de Desenvolvimiento de Fase mediante Redes Neuronales 26
5.2.1 Método de Regresión Directa (dRG) . . . 26
5.2.2 Método de predicción de orden de franja (dWC) . . . 27
5.2.3 Método Asistido por Aprendizaje (dAS) . . . 28
5.2.4 Desenvolvimiento Temporal (DL-TPU) y Geométrico de Fase (DL-GPU) . . . 28
5.2.5 Compensación de Errores Sistemáticos y Distorsiones No Lineales . . . 28
5.3 Paradigmas de vanguardia en la Metrología Asistida por IA . . . 29
5.3.1 Codiseño óptico-Algorítmico (Joint Optimization) . . . 29

5.3.2 Cuantificación de incertidumbre en Fase basada en Redes Bayesianas (BNN) . . . 30
5.4 Extracción de la Fase mediante Redes convolucionales (CNN) . . . 31
5.4.1 Arquitectura U-Net y Funciones de Pérdida . . . 32
5.5 Desenvolvimiento de Fase y Modelos Híbridos (Hformer) . . . 34
5.5.1 Matemática del Hformer: Bloque CAT . . . 36
5.6 Mapeo Directo Extremo a Extremo: Arquitectura UHRNet . . . 37
5.6.1 Motor Matemático: La función de pérdida compuesta . . . 39
5.7 Arquitecturas complementarias en el procesamiento de franjas y mapas de profundidad . . . 41
5.7.1 Redes Residuales Profundas (ResNet) y Bloques Residuales 41
5.7.2 Redes Neuronales Tripletas (Triplet Networks) . . . 42
5.8 Redes Generativas Antagónicas (GANs y Cycle-GAN) . . . 44

VI Material y Métodos o Metodología 48
6.1 Base de datos . . . 49
6.1.1 Base de datos sintética . . . 49
6.1.2 Composición y Generación de cada objeto . . . 50
6.1.3 Especificaciones Técnicas y Adquisición de los datos . . . 52
6.1.4 Mapas de fase envuelta . . . 52
6.1.5 Base de datos con imágenes reales . . . 53
6.2 Método Tradicional . . . 54
6.2.1 Proyección de franjas . . . 55
6.2.2 Espectro de Fourier (2D-FFT) . . . 57
6.2.3 Filtrado Espectral . . . 58
6.2.4 Transformada Inversa y Extracción de la fase envuelta . . . 62
6.2.5 Desenvolvimiento de Fase espacial (Phase Unwrapping) . . 63
6.3 Métricas . . . 65
6.4 Redes Neuronales convolucionales . . . 66
6.4.1 Configuración de cada una de las redes usadas . . . 66
6.4.2 UHRNet (U-shaped High-Resolution Network) . . . 67
6.4.3 SFNet (Symmetric fusion Network) . . . 68
6.4.4 Hformer (Hierarchical Vision Tranformer) . . . 69
6.5 Entorno experimental y protocolo de entrenamiento . . . 70
6.5.1 Optimización Paramétrica . . . 71
6.5.2 Supervisión y función de perdida . . . 71

VII Resultados y discusión 73
7.1 Base de datos sintética . . . 73
7.2 Evaluación del Desempeño del Método Tradicional . . . 74
7.3 Entrenamiento, Dinámica de Pérdida y Convergencia de las Redes Neuronales . . . 78
7.3.1 Resultados UHRNet . . . 78

7.3.2 Resultados SFNet 84
7.3.3 Resultados de la Red Hformer 87
7.4 Comparativo de las redes 91

VIII Conclusiones **95**

Referencias bibliográficas **97**

Índice de cuadros

1 Comparación de algoritmos de Phase Unwrapping (PU) con Máscara Diamante. Fuente propia 76
2 Comparación de métricas según la arquitectura. Fuente propia 78
3 Resultados del conjunto de prueba (Test) para UHRNet. 81
4 Resultados del conjunto de prueba (Test) para SFNet. 86
5 Resultados del conjunto de prueba (Test) para Hformer. 90
6 Comparación global de métricas por modelo evaluado. 92
7 Comparación de MAE por modelo en los objetos evaluados (Mickey, Dinosaurio y Pegaso). 93

Índice de figuras

1 Descomposición de señales. Fuente propia 13
2 Imagen original vs Espectro de magnitud. Fuente propia 14
3 Anatomía del Espectro de Franjas. Fuente propia 17
4 Proceso de la perfilometría de Fourier. Fuente propia 17
5 Transición de Fase Envuelta a Desenvuelta. Fuente propia 19
6 Clasificación de los Algoritmos de desenvolvimiento. Fuente propia . . . 20
7 Impacto del Ruido en el Unwrapping. Fuente propia 22
8 Arquitectura de dos etapas. Fuente propia 25
9 Predicción de orden de franja (dWC). Fuente propia 27
10 Sistema de Joint Optimization. Fuente propia 29
11 Diagrama de la Arquitectura U-Net con Conexiones de Salto (Skip Connections). Fuente propia 32
12 Arquitectura híbrida Hformer para el desenvolvimiento de fase. Fuente propia 35
13 Arquitectura UHRNet para el mapeo directo de franja a profundidad. Fuente propia 38
14 Arquitectura ResNet-34. Fuente propia 42
15 Bloque Residual Básico (Identify Block). Fuente propia 42
16 Arquitectura de las Redes Neuronales Tripletas. Fuente propia 44
17 Arquitectura GAN Convencional. Fuente propia 45

18 Arquitectura de Red Cycle-GAN para Mapeo de dominios. Fuente propia 46
19 Diagrama de la metodología usada. Fuente propia 49
20 Conjunto de muestras del GT. Fuente propia 50
21 Composición del conjunto de imágenes por Objeto. Fuente propia . . . 51
22 Ejemplo de datos reales. Fuente propia 54
23 Método tradicional perfilometria de fourier. Fuente propia 55
24 Proyección de franjas para objetos reales. Fuente propia 56
25 Máscara circular. Fuente propia 59
26 Máscara elipse. Fuente propia 60
27 Máscara rectangular. Fuente propia 61
28 Máscara de diamante. Fuente propia 62
29 Caracteristicas de cada una de las redes. Fuente propia 67
30 Arquitectura de la UHRNet modificada. Fuente propia 67
31 Arquitectura SFNet. Fuente propia 68
32 Arquitectura Hformer. Fuente propia 69
33 Base de datos sintética en el entrenamiento. Fuente propia 73
34 Objeto con franjas y mascara de diamante. Fuente propia 75
35 Reconstrucción 3D del objetos sintético con distinto algoritmos de desenvolvimiento de fase. Fuente propia 77
36 Reconstrucción 3D de objetos reales con distinto algoritmos de desenvolvimiento de fase. Fuente propia 77
37 Curva de entrenamiento de la red UHRNet. Fuente propia 79
38 Función de perdida de la red UHRNet. Fuente propia 79
39 Métricas de validación durante el entrenamiento. Fuente propia 80
40 Muestras de prueba. Fuente propia 83
41 Entrenamiento de la red SFNet. Fuente propia 84
42 Función de perdida SFNet. Fuente propia 84
43 Curvas de entrenamiento de las métricas SFNet. Fuente propia 85
44 Muestras de prueba para la red SFNet. Fuente propia 87
45 Entrenamiento de la red Hformer. Fuente propia 88
46 Validación de perdida Hformer. Fuente propia 88
47 curvas de entrenamiento de las métricas Hformer. Fuente propia 89
48 Muestras de prueba para la red Hformer. Fuente propia 91
49 Comparativa entre las tres redes. Fuente propia 92
50 Comparativa entre las tres redes. Fuente propia 93
51 Comparativa entre las tres redes 3D. Fuente propia 94

I. Introducción

En el área de la ingeniería, la metrologia óptica y la manufactura avanzada la reconstrucción tridimensional de las superficies sin contacto físico se han convertido en herramientas indispensables, la capacidad de recuperar geometrías complejas con alta resolución métrica y tiempos reducidos de adquisición resulta fundamental. Entre las tecnologías de luz estructuradas, la Perfilometria por Proyección de franjas (FPP) la cual se destaca por su naturaleza no destructiva y su versatilidad operativa.

Dentro de la FPP, la perfilometria por transformada de fourier (FTP), la cual fue propuesta por Takeda y Mutoh en 1983, su principio se basa en proyectar un patrón de franjas periódicas sobre el objeto, dicha topografía modula de manera espacial la fase de la señal luminosa, cuando se transforma en el dominio de la frecuencia mediante la transformada de fourier bidimensional es posible aislar la componente portadora fundamental a tráves de un filtrado pasabanda y con ayuda de la Transformada inversa de fourier (IFT) ayuda a recuperar un mapa de fase envuelta.

A pesar de su rapidez este método presenta limitaciones tanto matemáticas como operativas, en la etapa de filtrado espectral tiene como limitación el uso ventanas estrecha debido a que atenúa el ruido de fondo pero suprime las componente de alta frecuencia espacial, mientras que las ventanas que son excesivas introducen solapamiento espectral con la componente continua (DC) y los armónicos superiores. El mapa de fase desenvuelto queda acotado exigiendo un proceso de desenvolvimiento de fase con el fin de recuperar una distribución continua de alturas.

la metrología óptica contemporánea ha comenzado una transición fundamental hacia el Aprendizaje Profundo (Deep Learning) y el empleo de Redes Neuronales Convolucionales (CNN). Esta evolución tecnológica permite reformular la reconstrucción 3D no como un encadenamiento de etapas heurísticas y filtros dependientes del operador, sino como una tarea de regresión directa no lineal de franja a profundidad (fringe-to-depth). Mediante arquitecturas convolucionales avanzadas (como esquemas piramidales con módulos de agregación, redes híbridas con mecanismos de autoatención tipo Transformers o topologías multirresolución en paralelo que preservan rutas de alta definición continua), las redes son capaces de internalizar la física de la deformación óptica de franjas. De este modo, logran sortear la ambigüedad periódica modular de la fase y mitigar el ruido ambiental, alcanzando niveles de exactitud métrica propios de métodos multi-captura pero conservando la inmediatez de una única imagen de entrada.

La presente investigación tiene como propósito poder desarrollar la perfilometria de fourier mediante redes neuronales convolucionales con el fin de poder reducir el error del método tradicional, abordando el diseño y evaluación del flujo de la FTP, la generación de una base de datos sintética con patrones de iluminación modulada y el Ground Truth, así como la propuesta, entrenamiento y validación, comparando tres modelos de redes. A través de este enfoque híbrido entre el rigor físico óptico y la capacidad adaptativa de la inteligencia artificial, se busca sentar las bases para una metrología 3D automatizada, robusta y de alta fidelidad.

1.1. Antecedentes

La Perfilometría por Transformada de Fourier es una técnica óptica no invasiva la cual fue propuesta en el año de 1983 por Takeda y Mutoh, dicha técnica se ha consolidado como un método importante en la metrología óptica con el fin de hacer la reconstrucción tridimensional a partir de un único patrón de franjas [1][2] [3]. Su función principal es tratar que el patrón de franjas sea capturado como una señal portadora la cual ha sido modulada por la topografía del objeto [2]. Cuando se transforma la imagen al dominio de la frecuencia se aplica la transformada de Fourier bidimensional o unidimensional, el propósito principal es aislar la frecuencia que contiene la información de la fase con la ayuda de un filtro, como siguiente paso se aplica la transformada inversa de Fourier para pasar al dominio de espacial [2][4]. Aunque la técnica de FTP ha sido usada durante varios años, esta presenta limitaciones matemáticas severas y márgenes de error significativos, estos se derivan debido a su procesamiento global de la señal y su alta sensibilidad a la iluminación de fondo [4]. Estas limitaciones algorítmicas provocan que existan distorsiones en la fase, así como una gran pérdida de precisión al escanear perfiles complejos, texturas aisladas o superficies que tengan discontinuidades abruptas, a esto se le conoce como fuga espectral y superposiciones de frecuencias [4].

Con el paso del tiempo se ha optado por una transición desde un modelo físico rígido que es basado en la óptica geométrica hacia el uso de Redes Neuronales Convolucionales (CNN) y Aprendizaje Profundo (Deep Learning) [4][5], con el fin de poder terminar con dichas limitaciones logrando resolver el tema de la eficiencia en la medición, así como la precisión en la reconstrucción [6][7].

El primero gran hito para la reducción del error se enfocó en la etapa de extracción de la fase. Los métodos que son impulsados por inteligencia artificial evitan que se pueda predecir la fase de forma directa, esto se debe a que las redes neuronales tienen como limitación que tienen imprecisión al querer rastrear los continuos saltos espaciales y ambigüedades de 2π que están presentes en los mapas de fase envueltos [8]. Feng et al. propusieron un enfoque de una arquitectura CNN la cual tiene dos capas de entrenamiento para poder predecir con una alta fidelidad los componentes matemáticos intermedio de la imagen: se basa en la intensidad del fondo, así como el numerador y el denominador de la función arcotangente [8] [9]. Cuando se calcula la fase final la red logra evitar la filtraciones del dominio de la frecuencia propias de la FTP, ayudando a reducir el Error Absoluto Medio (MAE) reconstruyendo bordes finos con ciertos niveles de exactitud que son comparables con los algoritmos clásicos de múltiples disparos, esto se logra hacer de manera automatizada y sin intervención humana para ajustar filtros [7][10][11].

Cuando ya se extrajo la fase envuelta se requiere de un paso importante que es el desenvolvimiento de la fase envuelta (phase unwrapping) con el fin de obtener un mapa continuo de alturas. Dicha etapa es de suma importancia, así como es la más crítica esto se debe a que el ruido residual puede provocar que se generen ambigüedades las cuales se propagan de manera global, haciendo que no se haga una buena reconstrucción [12]. Se han empleado arquitecturas de CNN clásicas como lo son la U-Net y DCNN, sin

embargo, no se obtuvieron buenos resultados ya que las redes que son puramente convolucionales carecen de un campo receptivo global masivo, lo cual provoca que tengan la incapacidad de modelar de manera adecuada las dependencias a larga distancia de una imagen compleja [13][14]. Para poder solucionar esta deficiencia las investigaciones más recientes han desarrollado modelos híbridos en cual se combinan las CNNs con ciertos mecanismo de atención. La arquitectura que destaca hasta este momento en la Hformer, la cual se basa en integrar una red base de alta resolución mejor conocida como (HRNet) la cual contiene módulos de codificadores y decodificadores que son construidos sobre el bloque CAT (Cross Atention Transformer) [15][16][17]. Dicho modelo incluye una Red de Pirámide de Características (FPN) mejorada con el propósito de fusionar mapas a múltiples escalas [18][19]. El Hformer puede analizar el objeto de manera global integrando la información de toda la superficie, superando las CNNs normales generando predicciones de franjas visualmente continuas, suaves y casi libres de errores en áreas que son topográficamente agresivas [20][21].

En otras investigaciones presentan otra alternativa la cual es el mapeo directo que va desde franja hasta la profundidad (Fringe-to-Depth) esta se enfoca en evadir el riesgo de propagación de errores que son provenientes de las etapas de extracción, filtrado y desenvolvimiento [22]. Una de las mejores arquitecturas avanzadas que ha tenido buenos resultados en este enfoque es la UHRNet (U-shaped High-Resolution Network) [23]. Esta arquitectura utiliza una base de codificación-decodificación tipo U-Net a la cual se le incorporan Bloques de Convolución Multinivel y Bloques de Fusión de Alta Resolución. Estas características permiten transmitir de manera eficiente la información desde el patrón inicial evitando la perdida de los detalles finos [24][25]. También emplea una función de perdida compuesta la cual penaliza de manera rigurosa el Error Cuadrático por bloques (chunked L2) y el índice de Similitud Estructural (SSIMLoss) [23][26]. La UHRNet ha cumplido con el objetivo que es disminuir el Error Cuadrático Medio (RMSE) de la reconstrucción en un 58.87 % en comparación con los métodos basado en la U-Net pura logrando que se pueda restaurar con una alta precisión la reconstrucción 3D a partir de un solo patrón de entrada [27].

Una vanguardia nueva que se ha implementado se basa en el aprendizaje profundo informado por la física (Physics-Informed Deep) la cual se enfoca en no tener sobreajuste ante los ruidos inusuales, ya que esto es inaceptable en la metrología industria rigurosa [28][29]. Los modelos recientes como lo son el marco PI-FPA (Physics-Informed Fringe Pattern Analysis) los cuales integran una red neuronal ligera con un módulo de transformada de Fourier la cual es mejorada por el propio aprendizaje (LeFTP) [30]. Las redes PIDL se basan en principios físicos que van enfocados en la óptica, así como en las ecuaciones de interferencia en su función de perdida, la cual induce de manera matemática a la predicción de mantener una estricta consistencia espacial y en el espectro de frecuencia [30][31]. Esto provoca se pueda acortar de manera definitiva la brecha que hay entre la inmensa capacidad predictiva algorítmica con el rigor de los métodos físicos eliminando los errores impredecibles [32].

Los métodos actuales de reconstrucción 3D se enfocan en abordar el aislamiento del ruido casi-periódico de Moire basándose en la innovación en las operaciones ma-

temáticas: las convoluciones en el dominio de Fourier (FFC -Fast Fourier Convolution) [33][3][34]. Estas fueron remplazando las convoluciones espaciales estándar, arquitecturas como lo son la Unidad de Fourier (Fourier Unit) las cuales tienen como propósito procesar los tensores aplicando aplicando una Transformada Discreta de Fourier, realizando el filtrado enfocándose en la separación de las partes reales como las imaginarias la cual luego se va a retomar mediante la Transformada Inversa [35][36]. Las Redes Neuronales Convolucionales de multiresolución obtienen un campo receptivo matemáticamente infinito que es altamente eficiente e ideal para atenuar los ruidos periódicos que son introducidos por los proyectores digitales de franjas esto garantiza que la reconstrucción tridimensional alcance una buena fidelidad óptica, esto sucede cuando se opera en el dominio de la frecuencia [34][37].

1.2. Planteamiento del problema

La reconstrucción tridimensional de objetos mediante la perfilometría de Fourier enfrenta varios desafíos significativos los cuales afectan la precisión y la calidad de los modelos generados. Los métodos tradicionales, como lo son el análisis mediante la transformada de Fourier y el desplazamiento de fase, presentan limitaciones para el manejo de una manera adecuada del ruido ambiental, las discontinuidades en la superficie y las variaciones en la reflectividad de los materiales. Dichos factores limitan la capacidad de los sistemas actuales de poder producir reconstrucciones precisas, especialmente cuando se aplican a objetos con geometrías complejas o superficies irregulares.

El ruido ambiental puede distorsionar los patrones de franjas que son proyectadas sobre los objetos introduciendo errores en la fase recuperada y en la reconstrucción tridimensional final. Por otro lado, la discontinuidad en la superficie, como lo son los bordes pronunciados o cambios abruptos de profundidad, otra limitación que tiene es cuando se quiere hacer una reconstrucción tridimensional de un objeto que no tenga color este no se puede hacer hasta que la superficie se cubra y se vuelva mate, estas son difíciles de representar con precisión mediante las técnicas convencionales, las cuales no están diseñadas para capturar variaciones sin introducir artefactos significativos en el modelo. Las variaciones en la reflectividad de los materiales también representan una desventaja, esto se debe a que afectan la intensidad y claridad de las franjas proyectadas, teniendo como consecuencia la dificultad de la extracción precisa de la información de fase.

Otro factor importante es la adquisición del dataset mediante la perfilometría de Fourier, tradicionalmente se hace el uso de proyectores para la generación de los patrones lo cual provoca que se introduzcan errores al proceso, como lo son las distorsiones geométricas, no linealidades en la proyección, desenfoque, perdida de contrate a altas frecuencias y ruido óptico. Dichos factores en conjunto con las condiciones que tiene el entorno, así como las propiedades del objeto, provoca que el dataset tenga variabilidades no controladas complicando el entrenamiento al hacer el uso de modelos más robustos.

Debido a dichas limitaciones, el uso de un dataset sintético es una alternativa viable la cual tiene varias ventajas ya que permiten generan datos que son controlados ya que

se tiene un conocimiento de la fase, así como la geometría real (ground truth), lo cual facilita el entrenamiento supervisado de los modelos de aprendizaje profundo. También permite simular las condiciones adversas como lo que es el ruido, las variaciones de iluminación, discontinuidades y errores en los sistemas de proyección, contribuyendo a mejorar la capacidad de la generalización de los modelos.

Estos problemas además de reducir la precisión de la reconstrucción también limitan la aplicabilidad de la perfilometría de Fourier en los ámbitos industriales, médicos y científicos, en donde la calidad de la reconstrucción tridimensional es fundamental. La necesidad de querer superar estas limitaciones ha llevado a explorar enfoques más avanzados los cuales puedan mejorar la capacidad de las técnicas de perfilometría con el fin de mejorar la capacidad de las técnicas de perfilometría para poder producir reconstrucciones más detalladas y precisas.

La implementación de Redes Neuronales Convolucionales (CNN) en conjunto con la perfilometría de Fourier ofrece una solución para poder abordar estos desafíos. Las CNN cuentan con la capacidad de aprender patrones complejos a partir de grandes volúmenes de datos, permitiéndonos poder identificar y corregir las distorsiones que son introducidas por el ruido, las discontinuidades y las variaciones en la reflectividad. Para la integración efectiva de estas tecnología aun requiere investigación para desarrollar modelos que se puedan generalizar bien en distintas condiciones garantizando una mejora en la calidad de la reconstrucción tridimensional.

¿Cómo puede la integración de la perfilometría de Fourier con Redes Neuronales Convolucionales mejorar la precisión y reducir el error en la reconstrucción tridimensional, especialmente en condiciones adversas con ruido y discontinuidades?

1.3. Justificación

La reconstrucción de imágenes tridimensionales es fundamental en las áreas como la ingeniería, la medicina y la manufactura, en donde es necesario obtener información precisa sobre la forma y estructura de objetos complejos. Basándonos en este contexto, la perfilometría de Fourier ha sido una metodología eficaz para este tipo de reconstrucción mediante en análisis de patrones de las franjas proyectadas sobre la superficie de un objeto. Sin embargo, los métodos que se usan convencionalmente de la perfilometría de Fourier pueden presentar limitaciones como lo son el ruido ambiental, discontinuidades en la superficie y variaciones en la reflectividad de los materiales, esto complica la obtención de las reconstrucciones precisas y detalladas.

La incorporación de Redes Neuronales Convolucionales (CNN) en este proceso es de suma importancia para la solución de estos desafíos. Las CNN cuentan con la capacidad comprender y modelar patrones complejos que ocurren a través de datos, los cuales permiten mejorar de manera significativa la calidad de las reconstrucciones tridimensionales al identificar y corregir distorsiones que son inducidas por el ruido, así como las discontinuidades de la superficie. De igual forma, las CNN pueden ser entrenadas de tal forma que estimen de una forma precisa la fase en regiones que tiene datos incompletos contribuyendo a una buena definición de las características geométricas de los objetos

reconstruidos.

La perfilometría de Fourier asistida por CNN no solo incrementa la precisión en la reconstrucción tridimensional de los objetos, también permite una mayor automatización del proceso, provocando que se vuelva más eficiente. Esto resulta relevante en la industria en donde la velocidad y la precisión son importantes para garantizar la calidad del producto. Por otro lado, en el ámbito médico, una reconstrucción precisa de las estructuras anatómicas ayuda a optimizar el diagnóstico y el tratamiento, sin embargo, en la manufactura la reconstrucción tridimensional precisa es fundamental para el control de calidad, así como la fabricación de componentes complejos.

Esta justificación radica en la necesidad de poder desarrollar un enfoque avanzado en el cual se combine la capacidad analítica de la perfilometría de Fourier con las Redes Neuronales Convolucionales, con el fin de superar las limitaciones de los métodos tradicionales ofreciendo una herramienta más robusta y precisa la reconstrucción tridimensional. Esta implementación tiene como potencial contribuir de manera significativa en el ámbito académico como en el industrial, proporcionando una solución práctica con el objetivo de mejorar la precisión y eficiencia en la reconstrucción de estas imágenes en una amplia variedad de contextos.

1.4. Hipótesis

La implementación de un modelo en cual se integren la Perfilometría de Fourier con Redes Neuronales Convolucionales permitirá mejorar la precisión, así como reducir los errores en la reconstrucción tridimensional de objetos en un 5 % proporcionando una solución más robusta antes las condiciones adversas de ruido y discontinuidades en la superficie.

1.5. Objetivos

Objetivo General: Desarrollar e implementar la perfilometría de Fourier utilizando Redes Neuronales Convolucionales, con el fin de poder reducir el error en la reconstrucción tridimensional de objetos, mejorando la precisión y la eficiencia del proceso.
Objetivos específicos:

- Implementar un algoritmo tradicional de perfilometría de Fourier
- Desarrollar una base de datos sintética con objetos que tengan franjas proyectadas
- Proponer una red neuronal convolucional que emule el comportamiento de la perfilometría de Fourier para la reconstrucción tridimensional de objetos complejos.
- Evaluar el error en la reconstrucción tridimensional usando un modelo propuesto comparándolo con métodos tradicionales de perfilometría de Fourier.
- Realizar pruebas experimentales para poder validar la eficacia del modelo y ajustar parámetros para minimizas la precisión.

1.6. Alcances y limitaciones

Los alcances de dicho proyecto se enfocan en el cumplimiento de los objetivos, para poder llevarlos a cabo se hizo el uso de tres software MATLAB para la implementación del algoritmo tradicional de la perfilometría de Fourier, BLENDER para el desarrollo de la base de datos sintética y COLAB para la realización de las redes neuronales convolucionales comparando los resultados que se obtuvieron. Por otro lado, dichos alcances toman un enfoque en el desarrollo e implementación de una alternativa de la reconstrucción tridimensional combinando el método tradicional de la perfilometría de Fourier con el aprendizaje profundo abarcando lo siguiente:

- Desarrollo de la red neuronal emulando el comportamiento de la perfilometría de Fourier: se diseñaron y entrenaron dos arquitecturas basadas en Redes Neuronales Convolucionales (CNN) las cuales se orientaron a la reconstrucción 3D a partir de un único patrón de franjas, manteniendo la ventaja del "disparo único" (single-shot) que usa el método de Fourier.

- Extracción de la fase de manera automatizada y guiada: uno de los propósitos de la red neuronal es evitar la complejidad de la extracción de la fase envuelta. La red estimo componentes intermediarios como lo son la intensidad de fondo y el numerados/denominador de la función arcotangente evitando el uso de filtros pasabanda manuales.

- Eliminación del ruido: el uso de las arquitecturas convolucionales ayudo a atenuar las distorsiones logrando actuar como un filtro para poder eliminar el ruido ambiental, así como el ruido cuasi-periódico de Moiré.

- Uso de dataset sintético: para poder obtener buenos resultados en el entrenamiento de las redes, se creó un conjunto de datos simulados por computadora con ayuda del software Blender, permitiendo que la red pueda tener un modelo de referencia exacto (ground truth) libre de errores métricos, así como imperfecciones de calibración.

- Validación de resultados: se hizo una evaluación de la viabilidad, así como el rendimiento de la red comparando de manera estadística sus reconstrucciones tridimensionales.

Las limitaciones que se tuvieron en su realización se encuentran ligadas en las deficiencias matemáticas del algoritmo tradicional de la Perfilometría por Transformada de Fourier (FTP), restringiendo la calidad de los datos:

- Alta sensibilidad a la iluminación y ruido: la técnica de propuso Takeda se enfoca en el procesamiento de la señal de manera global, esto provoca que esta sea sensible a la iluminación de fondo, así como las variaciones de baja frecuencia ambientales, lo cual introduce ruido causando fuertes distorsiones en el momento de la captura.

- Fuga espectral y superposición de frecuencias: cuando se tienen objetos con geometrías complejas, la FTP sufre consecuencias las cuales son conocidas como fugas espectrales y las superposición de las frecuencias, causando fuertes distorsiones en la fase extraída.

- Perdida de detalles de alta frecuencia: para el uso de la FTP tradicional es necesario aplicar filtros en el dominio tanto espacial como frecuencial provocando que se tenga una pérdida de información de alta frecuencia, haciendo que los detalles finos que se tienen en la superficie, bordes afilados o texturas complejas no se puedan reconstruir con precisión.

- Dificultades en las discontinuidades abruptas: la transformada de Fourier puede tener limitaciones practicas cuando el objeto que se está analizando presenta cambios abruptos en la profundidad. Dichos cambios provocan que se tengan ambigüedades y saltos en el desenvolvimiento de fase (phase unwrapping) haciendo que se arruine la topografía final, esto se debe a que el método analítico tradicional es incapaz de distinguir si la discontinuidad pertenece a la forma que tiene realmente el objeto o a un error del algoritmo.

1.7. Organización de la tesis

Esta presente tesis se encuentra estructurada por ocho capítulos los cuales son organizados de manera secuencial con el fin de poder guiar al lector desde la fundamentación teórica hasta la validación experimental y las conclusiones finales:

- **Capitulo I: Introducción.** Se redacta la introducción en donde se exponen los antecedentes de la metrología óptica sin contacto, de igual manera se formaliza el planteamiento del problema ante las limitaciones que se tienen con la técnicas tradicionales, también se define la justificación del problema. Por otro lado, se establece la hipótesis general de la investigación, el objetivo general así como los objetivos específicos, delimitando los alcances y las restricciones de dicha investigación.

- **Capitulo II: Reconstrucción Tridimensional.** Para esta sección se desarrolla de manera matemática el marco teórico el cual sustenta el análisis de las señales ópticas, profundizando en las propiedades de la transformada de Fourier, la formulación de la Transformada Discreta de Fourier Bidimensional (2D-DFT) y el Teorema de la convolución, los cuales son esenciales para el procesamiento bidimensional en el dominio de la frecuencia.

- **Capitulo III: Perfilometría por Transformada de Fourier (FTP).** Se describen los principios físicos y ópticos del método clásico de FTP. En donde se aborda el modelo espectral el cual es derivado de las proyecciones de franjas detallándose el procedimiento analítico necesario para poder aislar la portadora fundamental y así poder obtener la distribución de la fase espacial envuelta.

- **Capitulo IV: Algoritmos de desenvolvimiento de fase (Phase Unwrapping).** En este apartado se presenta una clasificación de los métodos analíticos tradicionales los cuales son encargados de resolver la ambigüedad modula de 2π. También se desglosa la teoría, ventajas y restricciones que tienen los algoritmos locales los cuales son basados en seguimiento de trayectoria, los algoritmos globales de optimización o también conocidos de norma mínima, así como los esquemas especiales de desdoblamiento.

- **Capitulo V: Inteligencia Artificial en Reconstrucción 3D.** En esta parte se analiza la integración de la herramientas de Aprendizaje profundo en la metrología óptica. Se explora el enfoque físico-convolucional, las distintas metodologías de desenvolvimiento de fase asistidas o directas, así como los modelos de vanguardia de mapeo de extremo a extremo (fringe-to-depth). De igual manera, se formalizan las arquitecturas neuronales empleadas (U-Net, Hformer y UHRNet), el diseño de las funciones de pérdida compuestas y las redes complementarias para la preservación de características.

- **Capítulo VI: Material y métodos o Metodología.** Acá se detalla la formulación metodológica así como el protocolo experimental implementado incluyendo el proceso de generación de la Base de datos sintética en Blender y la adquisición de imágenes reales, la implementación sistemática del pipeline tradicional (evaluando filtros y algoritmos de desenvolvimiento), la configuración de las arquitecturas de red neuronal (UHRNet, SFNet y Hformer), y la definición de las condiciones de entrenamiento y métricas estadísticas de evaluación cuantitativa.

- **Capitulo VII: Resultados y discusión.** Se presenta el análisis comparativo del desempeño de los modelos evaluados. En este apartado se discuten los resultados obtenidos mediante el método tradicional exponiendo la dinámica de convergencia, curvas de pérdida y métricas de error de las arquitecturas neuronales, culminando con una discusión técnica comparativa que evidencia las capacidades y precisiones de cada red frente al ground truth.

- **Capítulo VIII: Conclusiones.** Se sintetizan las aportaciones científicas y el nuevo conocimiento generado a partir del cumplimiento de los objetivos, contrastando los hallazgos frente al estado del arte y proponiendo directrices para futuros trabajos de investigación.

II. Reconstrucción Tridimensional

La reconstrucción tridimensional (3D) es un proceso mediante el cual se pueden adquirir y almacenar en la memoria de una computadora las características físicas de un objeto real tales como sus dimensiones, forma de la altura y topografía [1]. Este proceso tiene como objetivo generar un modelo virtual exacto el cual produzca con alta fidelidad el volumen, la textura y el perfil que tiene el objeto de estudio [2]. Las metodologías que ayudan a lograr dicha reconstrucción se dividen en dos categorías: técnicas de contacto y técnicas de no contacto [4][5].

Las técnicas que son de contacto, como lo son las máquinas de medición por coordenadas (CMM) o los palpadores físicos los cuales son montados sobre brazos robóticos, tienen como fin explorar la superficie que tiene el objeto tocándola de manera directa para poder obtener las coordenadas de miles de punto [5][6].

Esta técnicas ofrecen una alta precisión y son utilizadas en la industria manufacturera, sin embargo, presentan ciertas desventajas que son muy importantes: son procesos considerablemente lentos los cuales dependen de la fricción física, son fáciles de causar deformaciones o daños que son irreversibles en piezas frágiles, flexibles o que tienen valor histórico [7][8][9][10].

Las técnicas de no contacto surgieron con el propósito de poder superar las barreras de las técnicas que son de contacto, estas son de carácter no invasivo las cuales se subdividen en métodos pasivos (como la estereovisión y la fotogrametría) y activos (como el tiempo de vuelo y la proyección de luz estructurada)[10][12][13].

Dentro de las técnicas activas se encuentra la Perfilometría por Proyección de Franjas (FPP) la cual se ha destacado de manera significativa. Dicho método consiste en iluminar la escena o imagen proyectando un patrón estructurado de manera bidimensional, comúnmente en un perfil sinusoidal, este es proyectado sobre el objeto [14][15]. La topografía y los cambios de profundidad de la superficie son de suma importancia ya que distorsionan el patrón de luz, este es capturado por un sistema de visión (cámara) [14][16]. Dicho proceso computacional funciona para extraer la información de dicha imagen siguiendo un flujo de trabajo riguroso [17][18]:

- Análisis de franjas y extracción de la fase: este paso se hace mediante algoritmos matemáticos como lo es el desplazamiento de fase o la Transformada Fourier (FTP), el propósito de este análisis es extraer la información de la profundidad la cual quedo codificada en el cambio de fase de la imagen [19][20].

- Desenvolvimiento de fase (Phase Unwrapping): en el cálculo que hace inicialmente se obtiene la fase envuelta la cual va de un rango de $-\pi$ a π, eso produce saltos o ambigüedades, para poder corregir estos cortes para poder obtener un mapa de fase continuo y suave es necesario aplicar un algoritmo [21][22][23].

- Mapeo de fase a altura: como punto final, la fase continua se convierte en coordenadas tridimensionales métricas utilizando los parámetros de calibración de la cámara y el proyector [17][24].

Ventajas de la perfilometría por proyección de franjas:

El uso de la proyecciones de luz estructurada tiene varias ventajas las cuales hace que esta metodología sea de las mas populares:

- Naturaleza no invasiva: esto es de suma importancia ya que se protege su integridad física de la muestras, debido a que se requiere que exista una distancia entre el sistema y el objeto.
- Alta velocidad y resolución: los escáneres de línea barren el objeto punto por punto capturando toda el área iluminada a la vez. Por otro lado, en el caso de la transformada de Fourier (FTP), permite que se obtenga el modelo 3D con un solo disparo, lo cual la hace ideal para capturas rápidas o entornos que son dinámicos [26][27].
- Bajo costo: la perfilometría de franjas requiere una configuración de hardware más sencilla ya que solo se ocupa un proyector y una imagen estándar, los cuales nos ayudan a obtener resultados con alta resolución micrométrica [16][28].
- Versatilidad: esta técnica permite que se pueda adaptar a objeto con distintas dimensiones, ya que solo basta con modificar la frecuencia, así como la densidad de las franjas proyectadas para poder ajustar el cambio de la medición [15].

Aplicaciones

La digitalización 3D no destructiva tiene varias aplicaciones en el control de calidad industrial, ingeniería inversa, arqueología y en la medicina [29]. En el área de biomédica, esta técnica permite la reconstrucción desde la anatomía humana para lo que es la fabricación de prótesis personalizadas de extremidades que son amputadas, hasta la creación de modelos volumétricos que sirven de base para la impresión 3d o manufactura aditiva [30][31][32].

A pesar de tener una alta aplicación, basado en ambientes prácticos dicha técnica tradicional se puede ver severamente afectada por interferencias, sombras las cuales son causadas por las geometrías complejas, altos niveles de reflectividad y ruidos periódicos como lo es el ruido de Moire [28][33][3]. El ruido de Moire es uno de los desafíos que más persiste en la reconstrucción tridimensional con la ayuda de proyección de franjas, este se manifiesta como un ruido cuasi-periodico el cual contamina las imágenes capturadas [1][2]. Dicho fenómeno óptico surge debido que se produce cierta interferencia física entre la frecuencia espacial que tienen los patrones de luz que son proyectados sobre el objeto y la estructura la cuadricula formada por los pixeles del sensor de la cámara utilizada para su adquisición [4]. Durante el procesamiento digital, el efecto de Moire se hace presente de manera crítica en la etapa de análisis, esto se debe a que introduce alteraciones, así como artefactos que son severos en lo algoritmos que son los encargados en las los pasos de extracción y desenvolvimiento de fase [2][5].

Este ruido trae como consecuencia que las distorsiones en la señal portadora lo cual impide la recuperación de la profundidad real que tiene la superficie, degradando de manera drástica las exactitud topográfica obteniendo que el modelo 3D tenga una

baja fidelidad tanto métrica como visual [1][2]. La integración de Redes Neuronales Convolucionales (CNN) y técnica de Deep Learning ha sido de gran avance en esta disciplina; las arquitecturas de CNN tienen como fin reparar las información faltante o filtrar el ruido de manera inteligente, aumentando su robustez permitiendo que se generen modelos 3D que son impecables los cuales superan las barreras de los métodos ópticos clásicos [3][34][35].

2.1. Transformada de Fourier

La Transformada de Fourier (FT) es una herramienta matemática que funciona de manera analítica en el procesamiento digital de las imágenes. Su funcionamiento principal esta basado en la capacidad de poder transformar la información de una señal que bidimensional en el dominio espacial en donde los pixeles se van organizando mediante coordenadas cartesianas (x,y) hacia el dominio de la frecuencia en donde la imagen se representa mediante sus componentes sinusoidales (seno y coseno) estructuradas en coordenadas de frecuencia espacial (u,v) [1][2][4][5].

En la figura 1, se puede mostrar una señal que es compleja la cual esta separada en múltiples ondas sinusoidales que son simples, ayudándonos a visualizar los conceptos de los armónicos.

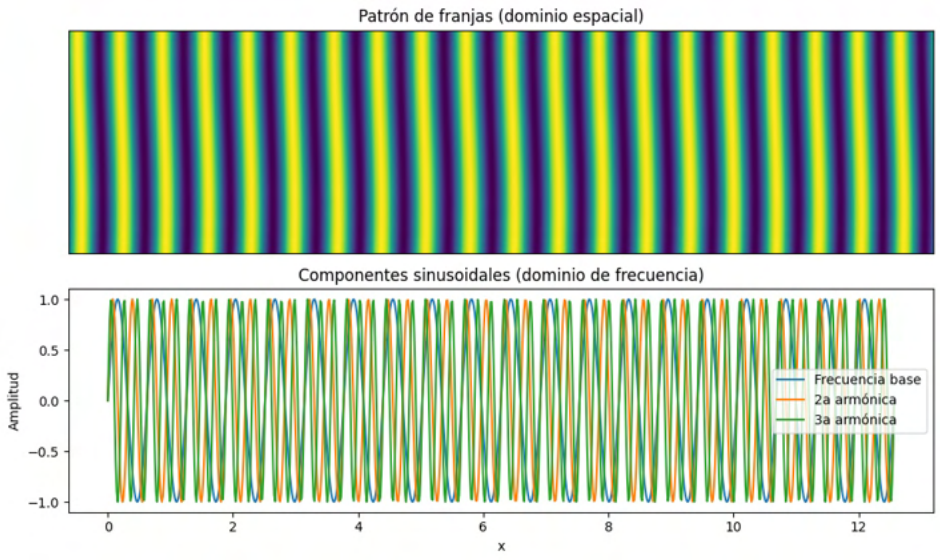


Figura 1: Descomposición de señales. Fuente propia

Tomando en cuenta una perspectiva física, cualquier imagen se puede entender como una superposición de ondas sinusoidales las cuales tienen distintas frecuencias, amplitudes y orientaciones. Cuando se tienen frecuencias bajas significa que se tienen variaciones lentas de intensidad como lo es la iluminación general o el fondo de la escena. Por otro lado, las frecuencias altas representan los cambios abruptos que se tienen como lo son los bordes, detalles finos o el ruido aleatorio. [5]

2.2. Transformada Discreta de Fourier Bidimensional

En el procesamiento de imágenes digitales, se trabajan con señales discretas las cuales son capturadas por los componentes. En estos casos se emplea la versión discreta de la transformada. Para una imagen $f(x, y)$ de dimensiones $M \times N$, cuya definición matemática es 1 [1][2][5]:

$$F(u, v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y)\, e^{-j2\pi\left(\frac{ux}{M}+\frac{vy}{N}\right)} \tag{1}$$

Donde $f(x, y)$ representa la intensidad del pixel de la imagen original, $F(u, v)$ hace referencia al espectro resultante, este contiene la amplitud y la fase de cada frecuencia y $e^{-j2\pi\left(\frac{ux}{M}+\frac{vy}{N}\right)}$ equivale la base compleja la cual nos permite la proyección de la imagen original hacia el espacio de frecuencia. [1][5]. Para retornar al dominio espacial original, es necesario aplicar la Transformadad inversa de Fourier (IFFT), que en su forma matematica se muestra de la siguiente manera [5][7]:

$$f(x, y) = \frac{1}{MN} \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} F(u, v)\, e^{j2\pi\left(\frac{ux}{M}+\frac{vy}{N}\right)} \tag{2}$$

Esto nos da como resultado un espectro en donde el módulo $F(u, v)$ nos indica prominencia de cada frecuencia, y la fase indica la posición relativa que tienen dichas ondas en la imagen.

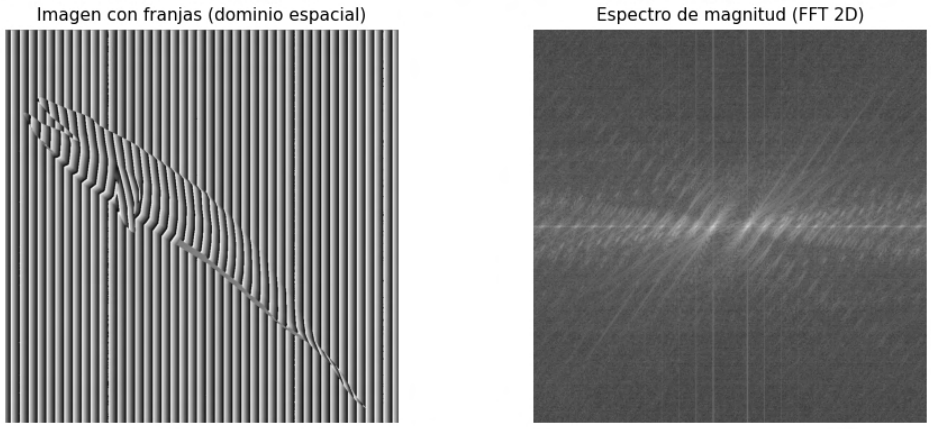


Figura 2: Imagen original vs Espectro de magnitud. Fuente propia

En la figura 2 se muestra una comparación entre el dominio espacial y el dominio de frecuencia. En la de la izquierda se muestra la imagen capturada en donde se puede observar la deformación de las franjas sinusoidales esto se debe a la topografía que tiene el objeto, por otro lado en la derecha se muestra el espectro de magnitud que se obtiene mediante la FFT 2D, en donde los lóbulos laterales contienen la información de la fase modulada por la altura del objeto.

2.3. Teorema de la Convolución

Una ventaja analítica que tiene mayor peso para este procesamiento es el conocido como Teorema de la Convolución. dicho teorema es de suma importancia en la perfilometría esto se debe a la eficiencia computacional ya que establece que una operación de convolución compleja entre una imagen y un filtro que esta en el dominio espacial equivale a una multiplicación que va de punto a punto en el dominio de la frecuencia, como se muestra en la ecuación (3)[10][11][12]. Esto facilita que se apliquen de manera efectiva los filtros pasabanda con el propósito de poder aislar la información que tiene la fase de manera instantánea, esto se logra multiplicando el espectro que tiene la imagen por una máscara de filtrado la cual esta diseñada para que solo pueda pasar la frecuencia de las franjas proyectadas [13][14][15][16].

$$f(x,y) * h(x,y) \iff F(u,v) \cdot H(u,v) \tag{3}$$

III. Perfilometría por Transformada de Fourier (FTP)

La perfilometria por Transformada de Fourier fue propuesta en 1983 por M. Takeda y K. Mutoh, esta técnica se ha consolidado como un método óptico no invasivo para la reconstrucción tridimensional [3][17][18][19][20]. Una de sus principales ventajas es extraer la topografía completa del objeto a partir de la captura de un único patrón de franjas [18][19][21].

3.1. Modelado Espectral del Patrón de Franjas

El principal proceso de este método es que se proyecte un patrón de luz estructurada sobre el objeto como una señal portadora la cual ha sido modulada en fase por las variaciones topográficas de la superficie [18][22][23]. La intensidad de la imagen que esta deformada la cual es capturada por la cámara, $g(x, y)$, se modela de manera matemática como se muestra en la ecuación (4) [19][24][25][26].

$$g(x, y) = a(x, y) + b(x, y)cos[2\pi f_0 x + \phi(x, y)] \tag{4}$$

Donde $a(x, y)$ representa a la intensidad de la iluminación que tiene de fondo, por otro lado, $b(x, y)$ hace referencia a la amplitud de contraste de modulación que tienen las franjas, f_0 es la frecuencia espacial fundamental de la rejilla portadora y por ultimo, $\phi(x, y)$ es la fase de interés, es en donde se contiene la información directa de la altura y distorsión geométrica del objeto [19][25].

Para poder hacer un análisis en el dominio de la frecuencia, se expande el término del coseno mediante la identidad de Euler, en la cual se obtienen tres componentes los cuales se manifiestan en el espectro de fourier [22][27][28], como se muestra en la ecuación (5).

$$g(x, y) = a(x, y) + c(x, y)e^{j2\pi f_0 x} + c^*(x, y)e^{-j2\pi f_0 x} \tag{5}$$

Donde $c(x, y) = \frac{1}{2}b(x, y)e^{j\phi(x,y)}$, y $C^*(x, y)$ denotan a su complejo conjugado [27][29].

Al expandir el término, se obtienen tres componentes los cuales se manifiestan en el espectro de Fourier [27][29]:

- **Componente DC (Lóbulo Central):** nos representa la intensidad de fondo $a(x, y)$

- **Lóbulos laterales (Fundamentales):** estos están ubicados en $+f_0$ y $-f_0$. El lóbulo lateral es el que contiene la información de la fase $\phi(x, y)$ modulada por la altura del objeto.

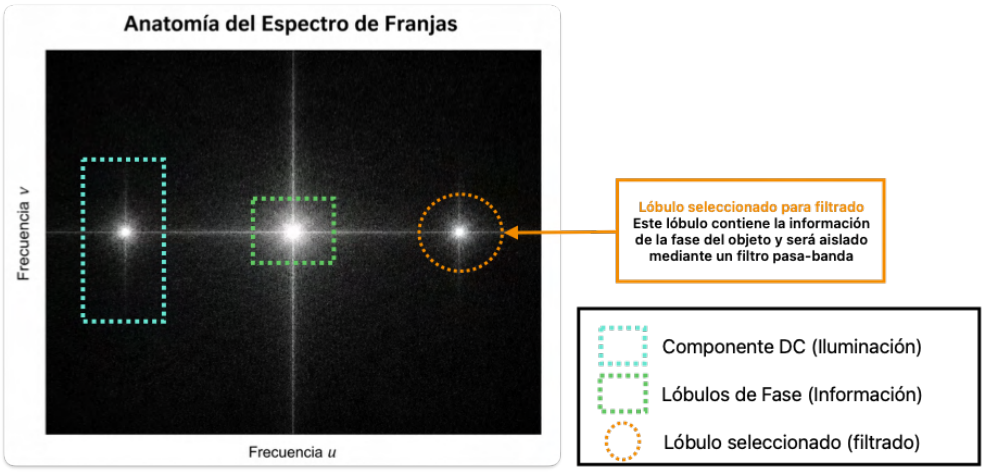


Figura 3: Anatomía del Espectro de Franjas. Fuente propia

3.2. Proceso Analítico y Extracción de Fase

El proceso del método que propuso Takeda se enfoca en aislar el término de la fase, tomando en cuenta las siguientes etapas como se muestra en la figura 4.

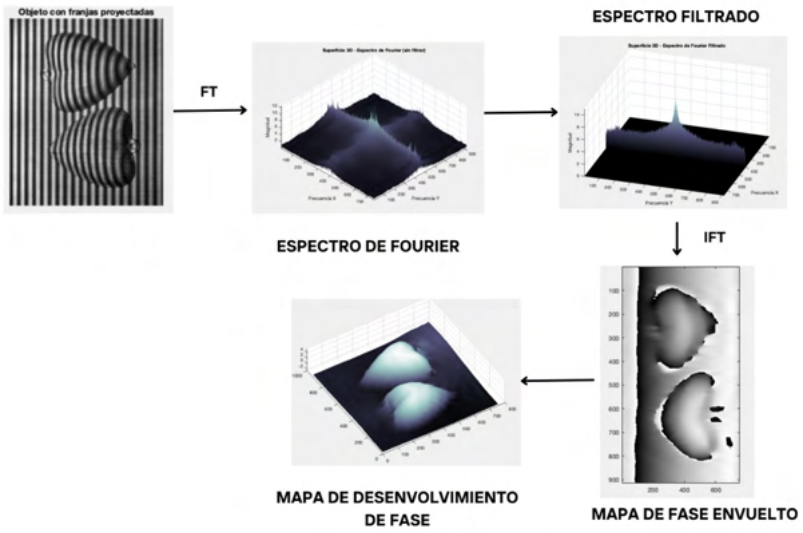


Figura 4: Proceso de la perfilometría de Fourier. Fuente propia

a. **Espectro de Fourier:** como primer paso se aplica la Transformada de Fourier (FFT) unidimensional o bidimensional a la imagen $g(x, y)$, generando una espectro $G(f, y)$ en los tres lóbulos principales, como se presenta en la ecuación (6)

[29][30][31][32].

$$G(f,y) = A(f,y) + C(f - f_0, y) + C^*(f + f_0, y) \tag{6}$$

b. **Filtrado y traslación:** Las variaciones en las topografías van cambiando de manera lenta en comparación con la frecuencia portadora, para poder aislar el lóbulo lateral $C(f - f_0, y)$ y descartar la componente central $A(f, y)$ la cual contiene el ruido de la iluminación se aplica un filtro pasa-banda centrado en f_0[18][31][33][3]. Posteriormente, dicho lóbulo ya filtrado se traslada al origen del espectro de frecuencia para poder eliminar la frecuencia portadora f_0, obteniendo así $C(f, y)$ [16][3][34][35].

c. **Transformada Inversa (IFFT):** Para poder retornar la señal centrada $C(f, y)$ al dominio espacial se debe aplicar la Transformada Inversa de Fourier, obteniendo como resultado la señal compleja $c(x, y)$[35][36][37].

d. **Calculo de la fase envuelta:** como punto final, la distribución que tiene la fase $\phi(x, y)$ se recupera con ayuda de las partes imaginaria ($\Im$) y real ($\Re$) de $c(x, y)$ mediante la función del logaritmo complejo o el arcotangente [34][37][38]. En la ecuación (7)

$$\phi(x,y) = \arctan\left(\frac{\Im\{c(x,y)\}}{\Re\{c(x,y)\}}\right) \tag{7}$$

Debido a la naturaleza que tiene la función arcotangente, la fase extraída se le conoce como fase envuelta, la cual esta limitada a un rango continuo que va de $-\pi$ a π, después de que se obtiene se debe someter a los algoritmos de desenvolvimiento de fase (phase unwrapping) los cuales sirven para generar un mapa continuo [39][40][41]. Esto hace que la fase se vuelva continua convirtiéndose a unidades métricas de altura $h(x, y)$ con ayuda de los parámetros geométricos de triangulación del sistema óptico [42][43].

IV. Algoritmos de desenvolvimiento de fase (Phase Unwrapping)

Para las técnicas de perfilometría por proyección de franjas, la extracción de una fase es de suma importancia para la recuperación de la información tridimensional, al ser calculada esta fase de manera matemática mediante la función de arcontangente se encuentra acotada en el intervalo $(-\pi, \pi]$ o $[-\pi, \pi]$ [2][44][45]. La condición mencionada produce un mapa de fase envuelto, dicho mapa presenta falsas discontinuidades o saltos que son abruptos de 2π los cuales no reflejan la topografía real del objeto [36][44].

El desenvolvimiento de fase, es una parte importante ya que este proceso busca recuperar la fase continua absoluta, esta es proporcional a la altura del objeto, se hace mediante la eliminación los saltos abruptos [44][20]. Este proceso se rige por la relación fundamental, la cual se describe como en la ecuación (8).

$$\phi(x, y) = \psi(x, y) + 2\pi k(x, y) \tag{8}$$

donde $\phi(x, y)$ hace referencia a la fase continua ya desenvuelta, $\psi(x, y)$ es la fase envuelta medida y $k(x, y)$ se basa en un campo de números enteros (orden de la franja) estos números deben determinarse para poder corregir la fase [36][44].

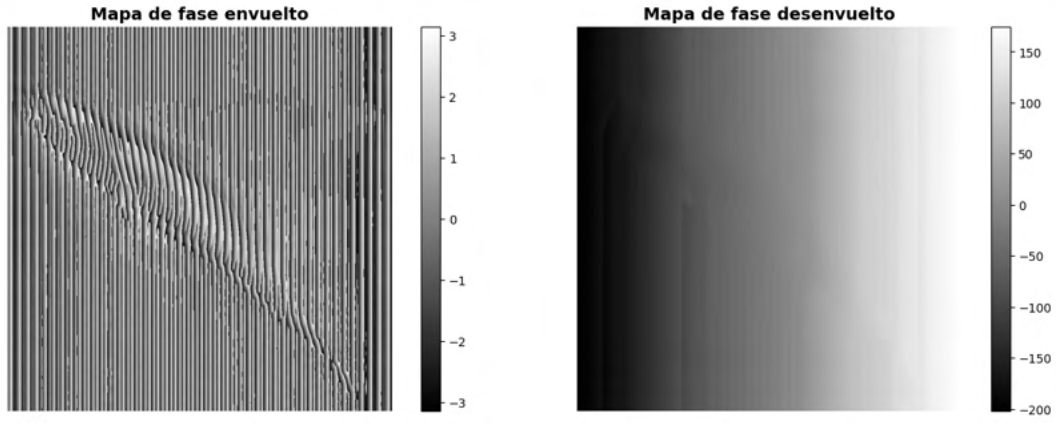


Figura 5: Transición de Fase Envuelta a Desenvuelta. Fuente propia

4.1. Clasificación de Algoritmos de desenvolvimiento

Este proceso tiene un gran desafió el cual es la presencia de ruido y atenuaciones en la imagen complicando la identificación de los saltos. Si se hace un mal calculo en las regiones de ruido el error de la fase se propagara hacia el resto de la imagen [20][44][45]. Para poder combatir este desafió, la literatura divide a los algoritmos en dos enfoques: Algoritmos locales (o de seguimiento de trayectoria) y Algoritmos globales también conocidos como de norma mínima [45].

En la figura 6 se muestra un diagrama en donde se presenta la clasificación de estos algoritmos los cuales se explicaran a continuación:

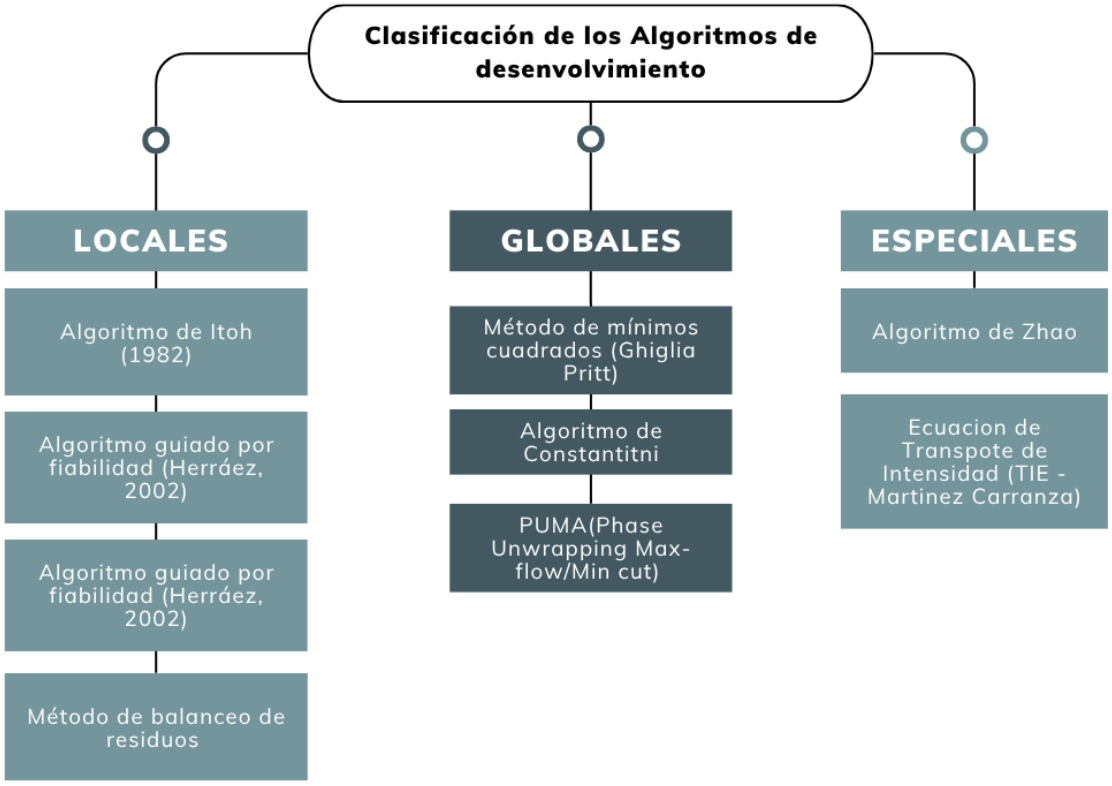


Figura 6: Clasificación de los Algoritmos de desenvolvimiento. Fuente propia

4.1.1. Algoritmos Locales (Seguimiento de Trayectoria)

Este tipo de algoritmos calculan el desenvolvimiento integrando el valor de la fase a lo largo de un camino especifico, estos suelen ser rápidos pero estos enfrentan un gran reto de no propagar errores si cruzan un píxel ruidoso [45]

- **Algoritmo de Itoh (1982):** este algoritmo fue pionero y creado en 1982. Este se basa en el supuesto de que las diferencias de fase real entre píxeles adyacentes deben ser menores a π. Dicho algoritmo escanea la imagen linea por linea, en caso de que este detecte una diferencia que exceda de π, esto asume que es un salto artificial, para corregirlo le suma o resta 2π. Aunque este método es rápido tiene como deficiencia que es vulnerable al ruido: un solo error que pase en un pixel este se propaga a lo largo de todo el reglón [44][45].

- **Algoritmo Guiado por Fiabilidad (Herráez, 2002):** Este algoritmo fue propuesto en 2002 por Herráez con el fin de mitigar la propagación de errores, este se enfoca en la clasificación de cada pixel otorgándole una métrica de calidad o fiabilidad, esta basada en las segundas diferencias de la fase local. En lugar de que este algoritmo siga un camino lineal o continuo, procesa y desenvuelve las regiones que son mas fiables, dejando la áreas que tienen ruido al final. Esto ayuda a reducir de manera drástica la propagación del error brindando una gran robustez ante las discontinuidades [44][45].

- **Método de Balanceo de Residuos:** este es un enfoque que busca identificar lo errores locales, los cuales son conocidos como residuos positivos y negativos, estos analizan la fase en pequeñas máscaras de $2x2$ píxeles [24]. Una vez que se identificaron estos aspectos, el algoritmo traza ciertas "brechas"las cuales conectan y aíslan los residuos, creando barreras las cuales son de suma importancia la integración teniendo prohibido cruzar. Esto ayuda a aislar el ruido previniendo su propagación por la malla [24].

4.1.2. Algoritmos Globales (Optimización o norma mínima):

Estos tipos de algoritmos plantean el proceso de desenvolvimiento como un problema matemático de optimización global, esta busca reducir el error en toda la matriz simultanea [45].

- **Método de mínimos cuadrados (Ghiglia & Pritt):** dicho método se encarga de formular el problema intentando minimizar la discrepancia que es el error cuadrático entre el gradiente de la fase envuelta y el de la fase desenvuelta en toda la imagen [44][45]. Esto lo hace con ayuda de la Transformada Discreta del Coseno (DCT) bajo condiciones de frontera de Neumann con el fin de asegurar que el resultado global sea continuo. Aunque este es muy robusto contra el ruido, tiene una naturaleza matemática la cual tiende a suavizar de manera excesiva las regiones del objeto las cuales presentan pendientes abruptas o bordes finos. Una de las variantes es el método de Andreas, el cual añade un término de regularización con el fin de optimizar la suavidad de la fase, aunque también puede perder los detalles finos [44][45].

- **Algoritmo de Constantini:** El desenvolvimiento se modela como un problema de optimización global con variables que son puramente entera. Para poder resolverlo de manera eficiente sin un alto costo computacional, este lo transforma en un problema que es de flujo de costo mínimo dentro de una red. Contantini se encarga de evitar la dispersión de los errores, teniendo resultados exactos y una mayor resistencia ante el ruido [44].

- **PUMA (Phase Unwrapping Max-flow/Min cut):** este algoritmo fue desarrollado por Bioucas-Dias y Valadão, este se encarga de formular el desenvolvimiento como un problema que es de minimización de energía dentro de un grafo

[12][45]. Evalúa los potenciales convexos y no convexos de la fase mediante las técnicas de cortes de grafos con el fin de minimizar las discrepancias. Con la ayuda de distintas evaluaciones estadísticas se ha demostrado que este algoritmo es uno de los mas estables y robustos que se ha usado en la literatura, combinando un bajo nivel en el error con buenas preservaciones estructurales del objeto 3D. Existe un derivado de este método el cual se conoce como PEARLS, este se encarga de aplicar primero un suavizado local adaptativo polinómico con el propósito de atenuar el ruido antes de usar PUMA, aunque este aumenta de manera drástica el costo computacional [45].

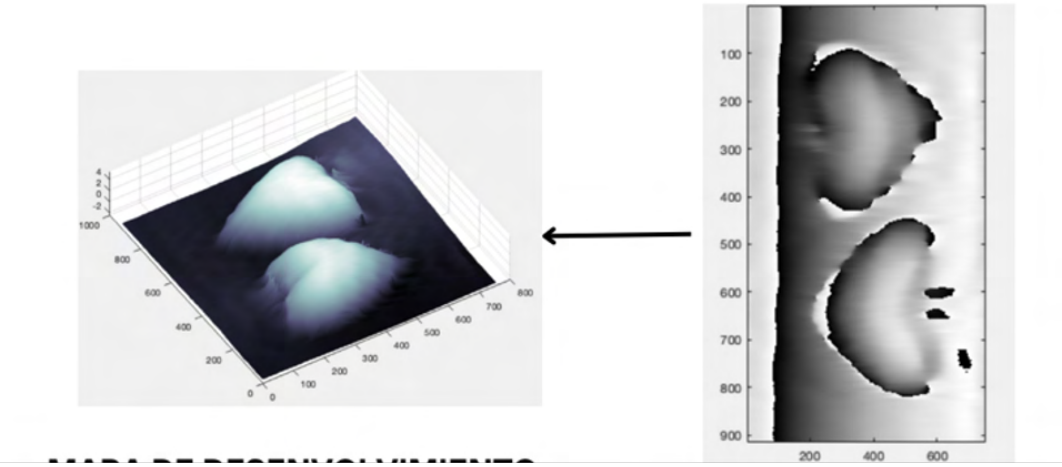


Figura 7: Impacto del Ruido en el Unwrapping. Fuente propia

En la figura 7 se muestra el objeto con franjas, en donde se pueden apreciar las deformaciones reales, a la derecha se presenta el espectro de magnitud de este mismo. En dicha figura se presenta el ruido que es visible en el espectro sucio, para resolver estos problemas se requiere de ciertos algoritmos globales como lo es PUMA o Mínimos Cuadrados con el fin de evitar fallos en la reconstrucción.

4.1.3. Algoritmos Especiales

- **Algoritmo de Zhao:** este fue propuesto para se utilizado en formas 3D complejas. Estas emplean dos distribuciones en la fase que tienen distintas precisiones. Esta hace uso de la fase que tienen menor resolución con el fin de estimar de manera robusta las órdenes de las franjas y después corrige punto por punto el valor de la fase mas exacta [45]. Esta permite preservar de mejor manera los detalles en zonas que tienen alto gradiente que tienen cambios abruptos, pero a pesar de estas ventajas tiene como principal desventaja que presenta una alta sensibilidad al ruido si este es excesivo en la distribución inicial [44][45].

- **Ecuación de Transporte de Intensidad (TIE - Martínez Carranza):** Este enfoque resuelve el desenvolvimiento evitando que se haga una integración convencional. Dicho algoritmo calcula la derivada axial de manea directa desde la fase envuelta aplicando el Laplaciano con Transformadas de Fourier y mirror-padding. Este es independiente a las trayectorias que son ruidosas, sin embargo cuando se tienen objetos con mucho ruido se requiere la implementación de manera iterativa, provocando que se aumente el costo del cálculo [45].

V. Inteligencia Artificial en Reconstrucción 3D

La metrología óptica que esta basada en proyección de franjas ha evolucionado con la incorporación de Aprendizaje Profundo (Deep Learning - DL), enfocándose en metodologías que son gobernadas estrictamente por modelos físicos rígidos a esquemas que son impulsados por datos los cuales son conocidos como (data-driven) [46].

En los métodos tradiciones de perfilometría por proyección de franjas, existe una relación matemática entre las imágenes de intensidad registradas por la cámara, $I \in \mathbb{R}^m$, y las coordenadas topográficas del objeto, $P \in \mathbb{R}^m$ (conocido como el mapa de fase o de profundidad), este se modela a través de un operador físico directo A [46], como se muestra en la ecuación (9).

$$I = A(P) \tag{9}$$

Existe una resolución analítica para dicho problema inverso ($P = A^{-1}(I)$), este a representado un problema histórico en la metrología esto es debido a la existencia de un ruido experimental, la reflectividad variable y las sombras de oclusión [46]. Basándose en los enfoques de Aprendizaje Profundo se suprime la necesidad de formular de manera explicita el operador analítico inverso [46]. Se recolecta un conjunto de datos el cual se asocia con los parámetros topográficos de referencia conocido como el Ground Truth (GT), p, con sus respectivos patrones de intensidad medidos (I), con el fin de poder entrenar un modelo el cual asimile una regla de reconstrucción paramétrica R_θ la cual es controlada por los pesos θ de la red convolucional [46], como se muestra en la siguiente ecuación (10):

$$\arg\min_{\theta} \sum \|p - R_\theta(I)\|^2 \tag{10}$$

5.1. Demodulación de la Fase Espacial: El enfoque Híbrido Físico-Convolucional

Feng y su grupo de investigación introdujeron el primer marco convolucional híbrido para la extracción de fase en el 2019 con el fin de resolver la contradicción existente entre los métodos espaciales de un solo disparo, el cual es caracterizado por su alta velocidad pero tiene una baja precisión y lo métodos temporales de múltiples capturas, estos son altamente precisos sin embargo son vulnerables al movimiento dinámico [10][46] [40].

5.1.1. Arquitectura de Dos Etapas (CNN1 y CNN2)

Una de las principales fallas que redes convolucionales es que a la hora de querer mapear de manera directa de la intensidad del patrón de franjas $I(x, y)$, a la fase envuelta ($\psi(x, y)$) fallan en la parte de la convergencia durante el entrenamiento, esto se debe a la incapacidad que tiene en las operaciones de las convolucionales espaciales para modelar de forma directa las discontinuidades de salto que es de 2π inherente a

la función de arcotangente [10][46]. Una solución que se dio para poder resolver esta limitación analítica es que el sistema asocia las ecuaciones del desplazamiento de fase con la flexibilidad de la redes neuronales profundas en un flujo dividido en dos bloques convolucionales secuenciales [10][46]:

a. **Estimación de fondo (CNN1):** Hace referencia al primer módulo convolucional el cual se basa en tomar el patrón sinusoidal deformado y ayuda a predecir la intensidad de fondo promedio o también conocido como componente de baja frecuencia $a(x, y)$ [10][46]. A este paso se le conoce como un filtrado espectral adaptativo en escala de grises, eliminando la modulación de las franjas con alta fidelidad [10][46]

b. **Generación de componentes Tridimensionales (CNN2):** esta segunda red toma como entradas de forma conjunta el patrón original $I(x, y)$ y el fondo estimado por la CNN1 $a(x, y)$ [10][46]. Cuando ya se tiene dicha información, la red estima con precisión nanométrica las funciones continuas correspondientes al numerador $M(x, y)$ y al denominador $D(x, y)$ de la ecuación trigonométrica de demodulación temporal [10][46].

A continuación en la figura 8 se presenta un esquema de la arquitectura de dos etapas [10][46].

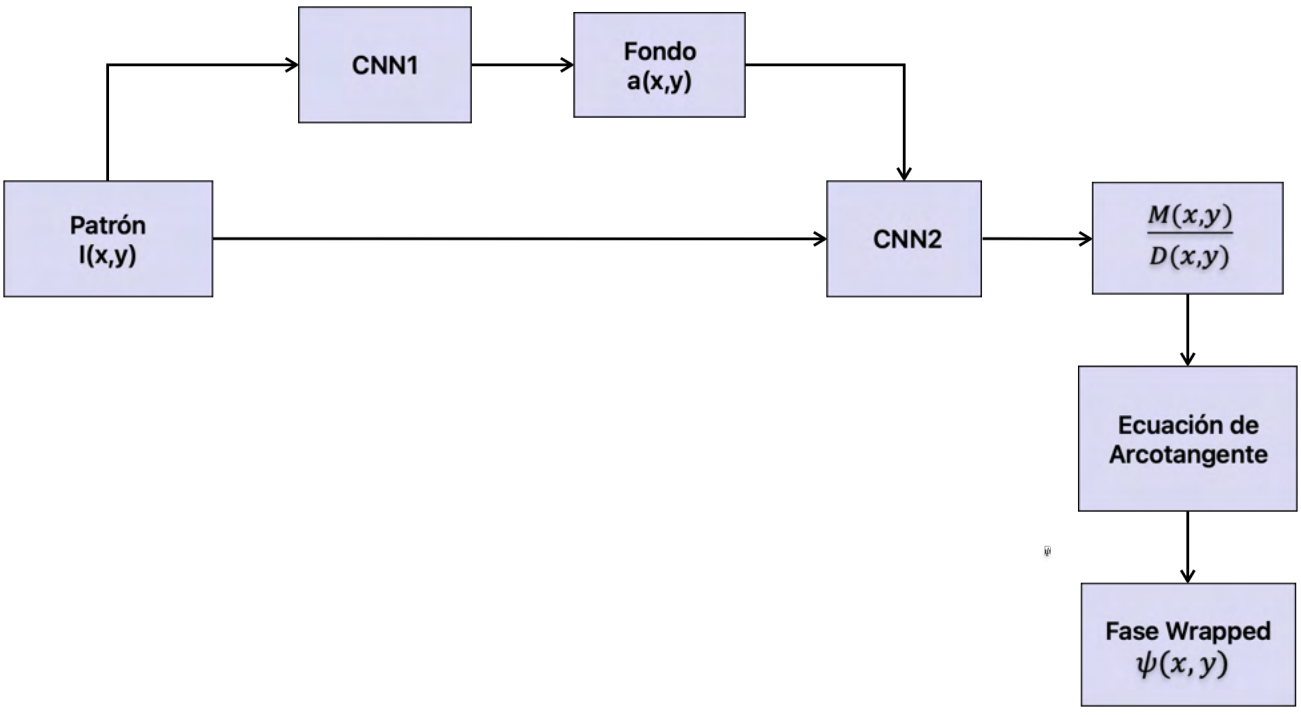


Figura 8: Arquitectura de dos etapas. Fuente propia

En su forma matemática, la red como principal funcionamiento predice las siguientes proyecciones continuas mostradas en las ecuaciones (11,12) [10]:

$$M(x,y) = \sum_{n=1}^{N-1} I_n(x,y) sin\delta_n = \frac{N}{2} B(x,y) sin\phi(x,y) \tag{11}$$

$$D(x,y) = \sum_{n=1}^{N-1} I_n(x,y) cos\delta_n = \frac{N}{2} B(x,y) cos\phi(x,y) \tag{12}$$

Donde $B(x,y)$ hace referencia al contraste local de las franjas y $\phi(x,y)$ representa a la fase de interés [10]. Cuando la CNN2 estima de forma nítida y libre de ruido a ambos tensores, la fase envuelta final $\psi(x,y)$, se calcula de manera analítica aplicando la función de arcotangente determinista, como se presenta en la ecuación (13) [46][10]:

$$\psi(x,y) = arctan(\frac{M(x,y)}{D(x,y)}) \tag{13}$$

Dicho modelo híbrido en donde se usa la física y red ayuda a evitar que exista una fuga espectral así como el solapamiento en el dominio de la frecuencia los cuales son típicos en la Perfilometría por Transformada de Fourier (FTP) [46][10].

5.2. Metodologías de Desenvolvimiento de Fase mediante Redes Neuronales

El método de desenvolvimiento de fase (Phase Unwrapping) espacial el cual representa la etapa mas crítica y matematicamente compleja de la perfilometría. Al ser calculada mediante la arcotangente la fase envuelta se obtiene en el rango que va de $(-\pi, \pi]$ [37]. El principal funcionamiento de los algoritmos de desenvolvimiento de fase es determinar de manera espacial el campo de números enteros de orden de franja $k(x,y)$ con el fin de poder recuperar la fase absoluta continua [37][47], en la ecuación (14) se presenta en su forma matemática:

$$\phi(x,y) = \psi(x,y) + 2\pi k(x,y) \tag{14}$$

Wang analizó la taxonomía que integro Wang, en donde clasifican los enfoque de desenvolvimiento de fase basado en IA en tres metodologías fundamentales dependiendo del uso y la formulación matemática de la red neuronal [37]:

5.2.1. Método de Regresión Directa (dRG)

Dicha técnica realiza una aproximación directa que va de extremo a extremo (End-to-End) en donde la red se alimenta con la fase envuelta ruidosa $\psi(x,y)$ e infiere de forma directa e inmediata el mapa continuo de la fase absoluta $\phi(x,y)$ en un solo paso de regresión [37][46]. Este método tiene como ventaja que su velocidad de inferencia

es extremadamente alta [46]. Por otro lado, tiene como limitación que al entrenarse con datos numéricos simulados las redes muestran problemas severos de generalización fallando al procesar los objetos que son complejos que tienen pendientes topográficas que son agresivas o geometrías discontinuas aisladas en donde la ambigüedad de 2π no se puede resolver localmente, esto sucede debido a la extrema dificultad de poder recolectar fases absolutas reales en entornos prácticos [46][37].

5.2.2. Método de predicción de orden de franja (dWC)

Dicho enfoque formula a desenvolvimiento como una tarea de clasificación semántica multiclase pixel por pixel [37][46]. En este método la red neuronal analiza el contexto de las franjas prediciendo estadísticamente cual es la clase entera que es el valor entero discreto de $k(x, y)$ que es correspondiente de cada pixel [37][46].

- **PhaseNet**: este es un modelo pionero en 2019 por Spoorthi el cual implemento una estructura de codificador-decodificador con el fin de predecir los órdenes de franja de forma discreta a partir de la fase envuelta [46].

- **PhaseNet 2.0:** se mejoró la precisión de segmentación mediante la incorporación de bloques convolucionales densos (DenseNet), logrando que se puedan mapear las fases envueltas ruidosas directamente hacia fases absolutas robustas y predecibles [46].

En la figura 9, se muestra el proceso de dicho método:

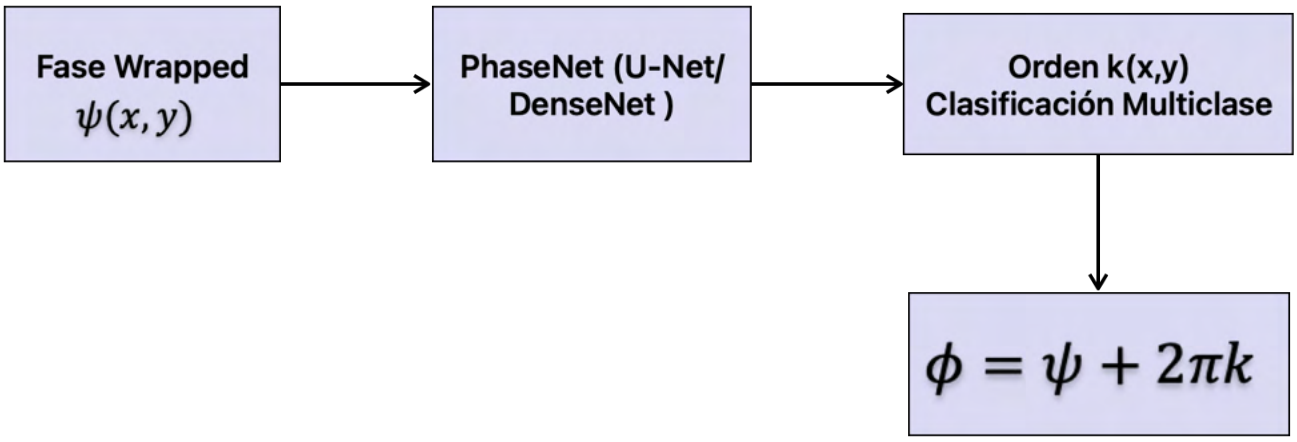


Figura 9: Predicción de orden de franja (dWC). Fuente propia

En cuanto se resuelven los problemas de desenvolvimiento a través del cálculo exacto del número de franja entero k mediante la clasificación, esto ayuda a que se eliminen los errores de redondeo numérico típicos de la regresión pura [37][47].

5.2.3. Método Asistido por Aprendizaje (dAS)

La red neuronal no calcula el orden k ni la fase absoluta directa. En su lugar entrena para poder predecir las variables intermedias de alta fidelidad como lo son el gradiente del orden de la franja o las lineas de discontinuidad de fase [37]. Después dichas lineas de frontera que fueron predichas por la IA se inyectan como restricciones prioritarias y barreras de exclusión en un algoritmo analítico de optimización clásico como lo es el método de mínimos cuadrados o cortes de grafos de flujo de costo mínimo [37]. Dicho enfoque híbrido fusiona la robustez de la IA con el fin de aislar el ruido local con la estabilidad matemática rigurosa de los algoritmos clásicos [37].

5.2.4. Desenvolvimiento Temporal (DL-TPU) y Geométrico de Fase (DL-GPU)

Para poder solventar de manera definitiva la ambigüedad física de 2π sin recurrir a un largo tiempo de captura de los métodos analíticos, se propusieron dos variantes profundas avanzadas [46]:

- **DL-based TPU:** en esta variante la red convolucional se puede alimentar de manera simultanea con un solo mapa de fase de un solo período libre de saltos el cual proporciona una guía gruesa de baja frecuencia, tambien lo hace con una fase envuelta ruidosa de akta frecuencia [46]. La red ayuda a asimilar geométricamente la relación temporal que existe entre ambas señales prediciendo de forma directa las órdenes de las franjas del mapa de alta frecuencia, esto ofrece una notable estabilidad en perfilometría dinámica tridimensional ante los ruidos de movimiento y baja relación señal-ruido [46].

- **DL-based GPU:** esta diseñada para sistemas ópticos de estereovisión binocular [46]. Dicho modelo procesa de manera simultanea los patrones de franja estéreo de doble vista así como los mapas de un plano de referencia para poder mapear de manera directa los órdenes de franja en un volumen de medición tridimensional ampliado, evitando los fallos que son provocados por las oclusiones [46].

5.2.5. Compensación de Errores Sistemáticos y Distorsiones No Lineales

Los componentes físicos de un sistema de perfilometría introducen perturbaciones las cuales corrompen la señal sinusoidal limpia. Entre dichos errores destacan el efecto Gamma no lineal el cual es causado por el proyector, la saturación de los píxeles de la cámara así como las distorsiones armónicas de alto orden que son generadas por el desenfoque binario de rejillas [46].

El uso de las redes convolucionales han ayudado a neutralizar dichas distorsiones de fase de acoplamiento sin tener que recurrir a inestables y complejos modelos de calibración analítica [46]:

- **Tree-Net (Yang et al., 2019):** Esta arquitectura cuenta con tres canales basado en U-Net los cuales ayudan a procesar directamente las intensidades de entrada

para poder aprender de forma robusta la función de distorsión Gamma no lineal, dicha arquitectura logro corregir los artefactos ópticos disminuyecndo en error de reconstrucción residual en un 90 % [46]

- **Compensación Multifactor (Zuo et al.):** Este es un modelo el cual esta basado en U-Net el cual es entrenado con patrones imperfectos reales y mapas de fase óptimos los cuales son obtenidos mediante el desplazamiento de fase de 12 pasos demostrando que es capaz de atenuar simultáneamente los efectos de saturación de iluminación y los armónicos residuales del proyector con solo introducir tres patrones ruidosos como entrada [46].

5.3. Paradigmas de vanguardia en la Metrología Asistida por IA

5.3.1. Codiseño óptico-Algorítmico (Joint Optimization)

De manera tradicional, el diseño del hardware óptico que se basa en la frecuencia o el desfase del patrón a proyectar, este se decidía de forma aislada e heurística con ayuda de aproximaciones como lo es la multiplexación de color o frecuencia [46]. El codiseño óptico-algorítmico plantea la integración del modelo físico matemático de la formación de la imagen así como de la propagación de la luz de forma diferenciable dentro de las capas de la red convolucional [46]. En la figura 10, se muestra como es el sistema de Joint Optimización.

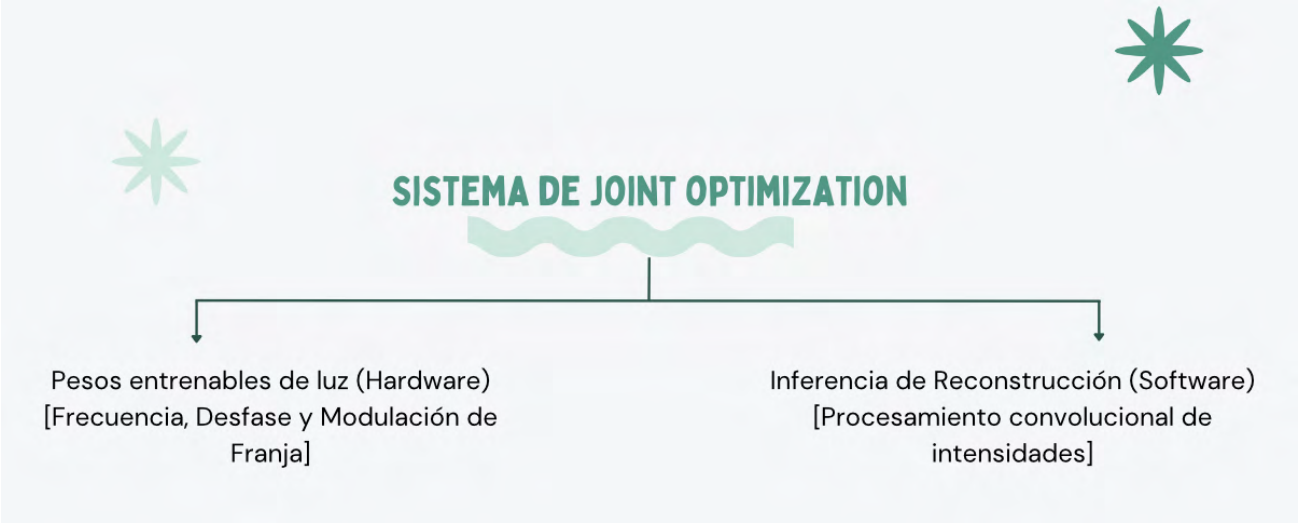


Figura 10: Sistema de Joint Optimization. Fuente propia

En dicho paradigma, los parámetros que ayudan a controlar la proyección de la luz estructurada actúan como pesos los cuales son entrenables en el hardware [46]. Al

optimizar de forma conjunto la proyección física asi como el algoritmo matemático de inferencia tridimensional, esta red aprende de forma autónoma cual es el mejor patrón de luz que debe emitirse con el fin de maximizar la calidad geométrica del objeto y como se puede procesar de forma nítida las capturas específicas, logrando que se obtenga una precisión superior con el mínimo númer de imágenes [46].

5.3.2. Cuantificación de incertidumbre en Fase basada en Redes Bayesianas (BNN)

Se implento el uso de Redes Neuronales Bayesianas (BNN) con el fin de poder mitigar la inaceptable naturaleza conocida como Çaja Negra"de la IA en los entornos industriales de alta exigencia [46]. Al contrario de las redes tradicionales las cuales devuelven predicciones puntuales de la fase las BNN asignan las distribuciones de probabilidad a sus pesos internos, permitiendo inferir para cada pixel un mapa cuantitativo de incertidumbre espacial de la fase [46]. El mapa de incertidumbre ayuda a proporcionar información directa en tiempo real sobre la magnitud exacta del error de fase local, permitiendo al sistema identificar y detener el flujo en caso de que se detecten anomalías de reconstrucción severas antes de procesar el volumen 3D definitivo [46].

Tradicionalmente los métodos de perfilometría óptica fueron basados en modelos físicos y matemáticos rígidos [1]. A pesar de que estos métodos fueron eficaces, técnicas como la Perfilometría por Transformada de fourier (FTP) tiene una deficiencia las cual es que tiene una alta sensibilidad frente a la iluminación de fondo, el ruido, así como variaciones abruptas en la topografía del objeto, generando fugas espectrales así como superposiciones de las frecuencias [2].

El estado del arte ha llegado a un paradigma fundamental en los datos: el Aprendizaje Profundo (Deep Learning) con el fin de poder terminar con estas limitaciones. Dicho enfoque tiene como propósito sustituir los ajustes manuales de filtrado por modelos que enfocan en reconocer los patrones a una gran escala, en donde la Redes Neuronales Convolucionales (CNN) juegan un papel muy importante debido a que asimilan una relación no lineal entre un patrón de franjas deformado y su respectiva topografía [1][4].

A diferencia de los distintos métodos analíticos los cuales se basan en despejar la fase a partir de ciertas ecuaciones trigonometricas explicitas [10], las Redes Neuronales modelan dicho problema como una aproximación de funciones complejas [1]. El modelo matemático general tiene como propósito buscar la aproximación de una función objetivo $f(x) = y$ con ayuda de una función parametrizada $g(x, W) = \hat{y}$, donde W representa a los parámetros (pesos) que la red debe aprender durante la etapa del entrenamiento para la predicción $\hat{y}$ sea lo mas cercana posible a la topografia real y [10][23].

Existen algunas arquitecturas y aportes de dicha transición tecnologica, las cuales son las siguientes:

5.4. Extracción de la Fase mediante Redes convolucionales (CNN)

Uno de los primeros hitos en la aplicación de la Inteligencia Artificial fue la automatización de la fase. Al intentar predecir la fase envuelta de manera directa mediante la implementación de redes neuronales clásicas, no se tuvieron buenos resultados debido a que presentaban altos margenes de errores ya que tienen la incapacidad de rastrear y predecir los saltos espaciales discontinuos así como las ambigüedades matemáticas de 2π [1][2].

Feng et al. impulso un enfoque basado en arquitecturas avanzadas que son impulsadas por CNN con el fin de poder evadir estas limitaciones analíticas, estas se estructuraron en dos etapas de entrenamiento para poder predecir componentes matemáticos que tienen alta fidelidad como lo son: la intensidad de fondo y los términos del numerador y el denominador de la función arcotangente [48][6]. Cuando se calcula la fase final a partir de estos términos indirectos, la red neuronal evita las filtraciones del dominio de la frecuencia de la FTP ayudando a reducir el Error Absoluto Medio (MAE) permitiendo que la reconstrucción sea precisa de bordes finos de forma automatizada [6].

La parte importante de este tipo de arquitecturas es la capa convolucional [32], esta tiene el propósito de sustituir las multiplicaciones matriciales densas con la ayuda de una operación de convolución discreta [4]. Esto permite que se puedan extraer características locales preservando la topología bidimensional que tiene el patrón de franjas [4][2].

Este tipo de red cuenta con tres partes matematicas importantes, las cuales se describiran a continuación:

- **Operación de Convolución Discreta:** Se tiene una imagen dada de entrada $I(x, y)$ y un filtro o núcleo el cual se conoce como Kernel (K) el cual tiene tamaño $k \times k$, el mapa de caracteristicas resultante en este caso se representa como $C(x, y)$ el cual se calcula de la siguiente manera como se muestra en la ecuación (15) [4][24]

$$C(x, y) = \sum_{j=1}^{k} \sum_{i=1}^{k} I(x + i, y + j) \cdot k(i, j) \tag{15}$$

  Mediante la optimización de los pesos de **k**, la red juega un papel importante ya que aprende a detectar de manera automática las modulaciones de franjas y bordes finos [4][12].

- **Funciones de Activación:** esta parte es de suma importancia ya que el modelo aprende a resolver el mapeo no lineal que va desde la franja hasta la profundidad aplicando funciones de activación a la salida de las convoluciones [32][23]. Una de las mas utilizadas de estas funciones en la **ReLU (Rectified Linear Unit)** [12][23]:

$$\text{ReLU}(x) = \text{máx}\{0, x\} \tag{16}$$

Dicha función ayuda a mitigar el problema que se tiene de desvanecimiento del gradiente, permitiendo que se tenga un aprendizaje mas profundo [23].

- **Operaciones de Submuestreo (pooling):** Para que se pueda otorgar a la red un campo receptivo mas amplio, es necesario aplicar capas de *Max Pooling*, las cuales tienen el propósito de reducir la dimensionalidad espacial, esto se logra seleccionando el valor maximo en una venta que es de $r \times r$, como se muestra en la ecuación (17)[12][23]:

$$P(x, y) = \max_{i,j \in r \times r} I(x + i, y + j) \tag{17}$$

5.4.1. Arquitectura U-Net y Funciones de Pérdida

Una de las arquitecturas que se ha consolidado como el estándar base para este tipo de tareas de regresión es la U-Net, esto se debe a su topología simétrica que tiene de codificador-decodificador [12][8]. Esta arquitectura tiene como ventaja las skip connections (conexiones de salto), estas se basan en concatenar mapas de características de alta resolución de la fase de contracción directamente en la fase expansión. Esto ayuda a garantizar que la información semántica de alto nivel se pueda fusionar con los detalles que tienen bajo nivel que son los bordes y texturas finas ayudando a mitigar la propagación de ruido [4].

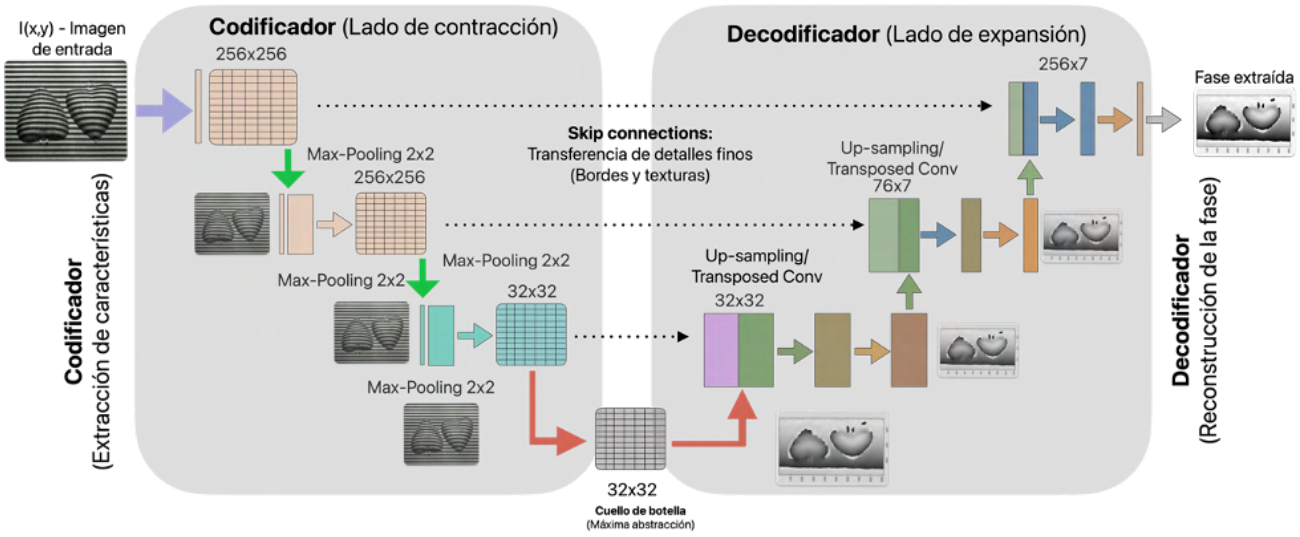


Figura 11: Diagrama de la Arquitectura U-Net con Conexiones de Salto (Skip Connections). Fuente propia

En la figura 11, se muestra la arquitectura U-net, en donde se puede observar de manera detallada el proceso de contracción (codificador) para la extracción de las ca-

racterísticas globales del objeto así como el proceso de expansión (Decodificador) para la reconstrucción de la fase extraída. Las skip connections son de suma importancia ya que garantizan la transferencia de los detalles finos desde la etapa inicial, ayudando a mitigar la perdida de la información espacial durante el proceso de submuestreo.

La red U-Net es una arquitectura estándar para la extracción de fase esto se debe a su tipo de topología simétrica de codificador-decodificador como se muestra en la figura 11 [12]. Una de las ventajas que radica en esta arquitectura radica en las conexiones de salto la cuales son conocidas como skip connections, el principal objetivo de estas es concatenar los mapas de alta resolución de la fase de contracción de manera directa en la parte de la expansión [12][4]. Dichas propiedades garantizan que el contexto global se pueda fusionar con los detalles microscópicos.

Para el entrenamiento de dicha red se basa en la minimización de una función de pérdida con la ayuda de la retropropagación o mejor conocida como Backpropagation [12][23]. Para tareas especificas como la reconstrucción topografica, se emplean métricas analiticas como:

- **Error Cuadrático Medio en Imágenes (IMMSE):** Dicha métrica ayuda a medir la discrepancia numérica que se tiene entre la fase/altura predicha $K(x, y)$ y la real que denota por $I(x, y)$, esto se muestra en la ecuación (18)[45]:

$$IMMSE = \frac{1}{mn} \sum_{y=1}^{m} \sum_{x=1}^{n} [I(x, y) - K(x, y)]^2 \tag{18}$$

  Donde $I(x, y)$ representa a la altura que tiene el pixel de la imagen de referencia (ground truth), por otro lado $K(x, y)$ también equivale al valor del pixel dentro de la misma coordenada pero esta ya es dentro de la imagen reconstruida por el algoritmo y por ultimo m y n simboliza el número total de renglones y columnas respectivamente de la imagen.

  Esta métrica se utiliza estrictamente con el fin de evaluar la calidad que tiene la reconstrucción de los objetos, comparándolos de pixel a pixel la imagen original real con la imagen ya reconstruida en este caso puede ser después de aplicar un algoritmo de desdoblamiento de fase [45]

  Para poder interpretar el IMMSE, es necesario saber que la formula resta el valor del real en cada pixel elevando el error al cuadrado $[I - K]^2$ [45]. El propósito de elevar al cuadrado es para cumplir los dos siguientes propósitos:

  - Asegurar que los errores negativos y positivos no se cancelen entre si.
  - Penaliza fuertemente a la red neuronal cuando esta comete errores de gran magnitud

  Cuando los valores del IMMSE son mas pequeño significa es que se hizo una buena reconstrucción, ya que demuestran que las alturas o fases estimadas por

la red tienen muy poca discrepancia numérica frente a la real del objeto que fue evaluado.

- **Función de coste:** Esta es utilizada con el fin de garantizar similitudes estructurales globales con el fin de poder evitar la inestabilidad [12]. Se encarga de medir la discrepancia o el error que se tiene entre las predicciones que realiza la red neuronal y los valores reales esperado los cuales se conocen como ground truth[23]. En el proceso del entrenamiento el algoritmo de optimización tiene como propósito que se ajusten los pesos internos que tiene la red para que el resultado de dicha función de perdida sea lo mas cercana posible a cero [23]:

$$\text{Pérdida} = 1 - \frac{(2 \cdot m_r \cdot m_p) + coef}{(m_r^2 + m_p^2) + coef} \tag{19}$$

Para ciertos problemas en donde se busca predecir valores que son continuos como en el caso de la reconstrucción de alturas topográficas, la función mas usada es el Error Cuadrático Medio (ECM o MSE), cuya formula general como se muestra en la ecuación (20) ayuda a estimar el promedio de los errores al cuadrado[23]:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(f(x_i) - g(x_i))^2 \tag{20}$$

5.5. Desenvolvimiento de Fase y Modelos Híbridos (Hformer)

Cuando se obtiene la fase envuelta, el desenvolvimiento de fase (phase unwrapping) es una de las etapas mas críticas en la reconstrucción, ya que un error provocado por el ruido residual puede provocar que se genere imprecisión global haciendo que se arruine la reconstrucción [9]. Las arquitecturas que son solamente convolucionales suelen carecer de un campo el cual es el receptivo global masivo, esto tiene como consecuencia que esta arquitectura sea ineficiente para poder modelar ciertas dependencias a larga distancia en topografías complejas [9][10].

Se han creado modelos híbridos como lo es la arquitectura Hformer (Hybrid CNN-Transformer) para poder solucionar estas carencias [9][47]. Dicho modelo integra una red base la cual cuenta con una alta resolución (HRNet) la cual contiene módulos que son construidos sobre el bloque CAT (Cross-Attention Transformer) asi como una Red de Pirámide de Características (FPN) con el fin de fusionar los mapas a múltiples escalas [9]. Los mecanismos de auto-atención (self-attention) son de suma importancia para que la red logre comprender la continuidad visual global que tienen las franjas logrando que puedan predecir el orden exacto saltándose las discontinuidades abruptas [10]. Esto nos da como resultado que las superficies sean visualmente suaves y libres de errores algorítmicos [9].

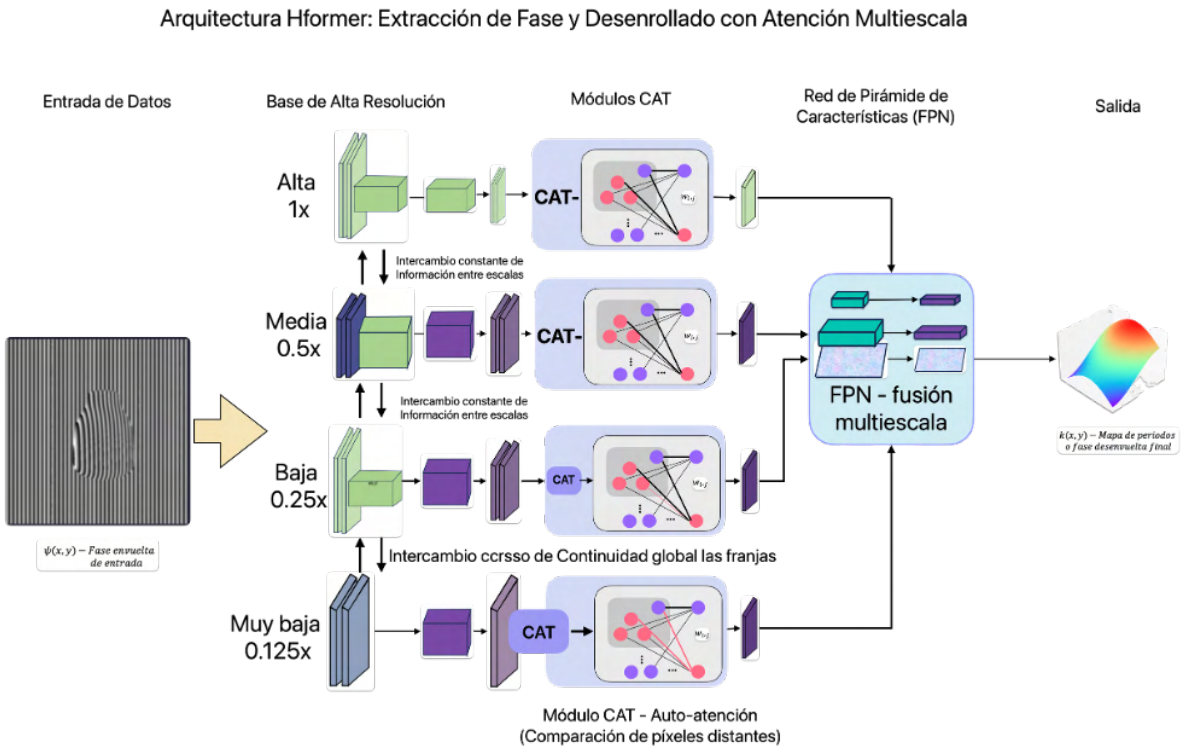


Figura 12: Arquitectura híbrida Hformer para el desenvolvimiento de fase. Fuente propia

En el diagrama de la figura 12 muestra la integración de una de una base la cual cuenta con alta resolución (HRNet) con bloques de atención cruzada (CAT). Se destaca el flujo multiescala en paralelo lo cual permite mantener la fidelidad de las franjas mientras que el mecanismo de auto-atención ayuda a modelar las dependencias globales ayudando a que se haga un desenvolvimiento de fase robusto ante discontinuidades y oclusiones complejas. En la figura 12 se muestra la arquitectura de red Hformer, en la parte de la Red de Alta Resolución a diferencia de la Red U-Net de la figura 11 la información toma rutas paralelas, se puede notar como esta fluye en niveles de resolución alta, media y baja de manera simultanea, esto nos permite que la red no pierda los detalles finos que tienen las franjas.

También se muestra el Mecanismo de Atención el cual es el Bloque CAT (Cross-Attention Trasnformer), este bloque es de suma importancia ya que la red lo utiliza para que esta pueda reconocer las información importante que tiene la imagen con el fin de resolver el desenvolvimiento de la fase, enfocandose en la parte donde hay sombras o ruido.

Dicha arquitectura también cuenta con un FPN la cual se refiere a la Red de Pirámide de Características, en esta parte se puede observar como todas las rutas se fusionan, este paso es de suma importancia ya que se obtiene un mapa de fase suave y continuo, dicha red combina la parte del contexto global con el detalle local que son los bordes.

El diseño Hformer es una arquitectura híbrida que tiene como fin superar las limi-

taciones de las CNN tradicionales como la U-Net. Dicha red cuenta con tres bloques importantes:

- **Backbone (Columna Vertebral CNN):** Usa una HRNet para poder procesar la fase envuelta de la entrada y extracción de los mapas de características a múltiples escalas, la cual aprovecha la capacidad que tienen las CNN para poder detectar los detalles locales finos [47].
- **Encoder-Decoder con Transformers:** Con ayuda del uso de los mecanismos de atención toma caracteristicas que tienen las CNN lo cual permite relacionar las zonas que estan muy lejanas en la imagen [47]
- **FPN Mejorada (Feature Pyramid Network):** Esta parte es la encargada de poder fusionarlas información que tienen las distintas etapas con el fin de emitir la predicción del orden de la franjas en la misma resolución que tiene la imagen original [47].

5.5.1. Matemática del Hformer: Bloque CAT

El núcleo de la innovación de Hformer esta en la parte del Bloque CAT (Cross-Attention Transformer), el cual se utiliza en la parte del codificador y el decodificador[47]. Dicho bloque ayuda a fraccionar la imagen en parches y aplicando una serie de cálculos que se alternan entre analizar lo que sucede dentro de un parche y lo que sucede entre los parches distantes [47].

El procesamiento de las características de entrada x_p hacia la salida y_p en un bloque CAT se modela de manera algebraica mediante una secuencia la cual incluye Normalización de Capa (LN), Perceptrones Multicapa (MLP), Auto-atención Intra-Parche ($IPSA$) y Auto-atención Cruzada entre Parches ($CPSA$) [47]:

$$\hat{I}_p = x_p + IPSA(LN(x_p)) \tag{21}$$

$$\hat{M}_p = \hat{I}_p + MLP(LN(\hat{I}_p)) \tag{22}$$

$$C_p = \hat{M}_p + CPSA(LN(\hat{M}_p)) \tag{23}$$

$$\tilde{M}_p = C_p + MLP(LN(C_p)) \tag{24}$$

$$\tilde{I}_p = \tilde{M}_p + IPSA(LN(\tilde{I}_p)) \tag{25}$$

$$y_p = \tilde{I}_p + MLP(LN(\tilde{I}_p)) \tag{26}$$

Tanto el modulo IPSA como el CPSA basan su funcionamiento en el cálculo de manera matemática de la Matriz de auto-atención [47]. Dicha operación permite que

la red pueda calcular la importancia del pixel A para poder comprender el píxel Z, sin importar que tan alejados estén el uno del otro.

De manera matemática, los tensores de la imagen se convierten en 3 matrices: Q (Query o Consulta), K (Key) y V (Valor) [47]. La relación de la atención se calcula como se muestra en la ecuación (27).

$$Attention(Q, K, V) = SoftMax(\frac{QK^T}{\sqrt{d}} + B)V \tag{27}$$

Donde:

- QK^T: representa en producto punto que se tiene entre la matriz de consultas y la de claves transpuestas. Esto ayuda al calculo de el puntaje de similitud que se tiene entre todos los elementos que tiene la imagen.
- $\sqrt{d}$: Es el factor de escalado el cual esta basado en la dimensión d que tienen las matrices, este es usado para poder mantener la estabilidad numérica que tienen los gradientes [47].
- B: equivale a la desviación relativa de posición (relative Position deviation), la cual le indica a la red en donde se ubica cada parche.
- $SoftMax$: es la función exponencial normalizada que ayuda a convertir los puntajes en una distribución de probabilidades que hay entre (0,1) [47].
- V: representa a la matriz de valores que se multiplican por dichas probabilidades con el fin de ponderar la información.

Cuando el mecanismo de atención y el bloque CAT han cumplido su propósito es cual es extraer su contexto global, la red no puede arrojar los resultados de manera inmediata esto se debe al uso de los Transformers ya que estos consumen una memoria computacional enorme. Para que el modelo pueda entrenar de extremo a extremo, el Hformer incorpora la FPN mejorada con el fin de concatenar los mapas de características resultantes de las cuatro etapas del decodificador esto lo hace mediante el uso de submuestreo (upsampling) [47].

Para la salida final a nivel pixel, la última capa hace uso de la función de activación $SoftMax$ ya que aborda el cálculo de $k(x, y)$ como un problema de multiclasificación [47]. Para poder estimar un número continuo, la red ayuda a predecir de manera estadística a que clase o número entero absoluto $(0, 1, 2, 3...)$ pertenece la franja en esa coordenada especifica [47].

5.6. Mapeo Directo Extremo a Extremo: Arquitectura UHR-Net

En investigaciones recientes han optado por eludir por completo las etapas tradicionales de extracción, filtrado y desenvolvimiento de fase mediante un mapeo directo

de franja a profundidad (Fringe-to-Depth) [1]. En lugar de extraer la fase envuelta mediante la función arcotangente y despues aplicar los algoritmos de desenvolvimiento los cuales tienen como limitante a que son muy vulnerables a propagar errores debido a las oclusiones o el ruido, dicho modelo aprende la función de la regresión matemática compleja que ayuda a transformar de manera directa la imagen de intensidad del patrón de franjas deformado $I(x, y)$ en un mapa de profundidad o altura $h(x, y)$ [39][47].

Una de las arquitecturas que representa este enfoque que es mapeo directo e extremo a extremo es UHRNet (U-shaped High-Resolution Network) [39] como se muestra en la figura 13.Dicha arquitectura cuenta con la siguiente estructura, la cual se explicara a continuación:

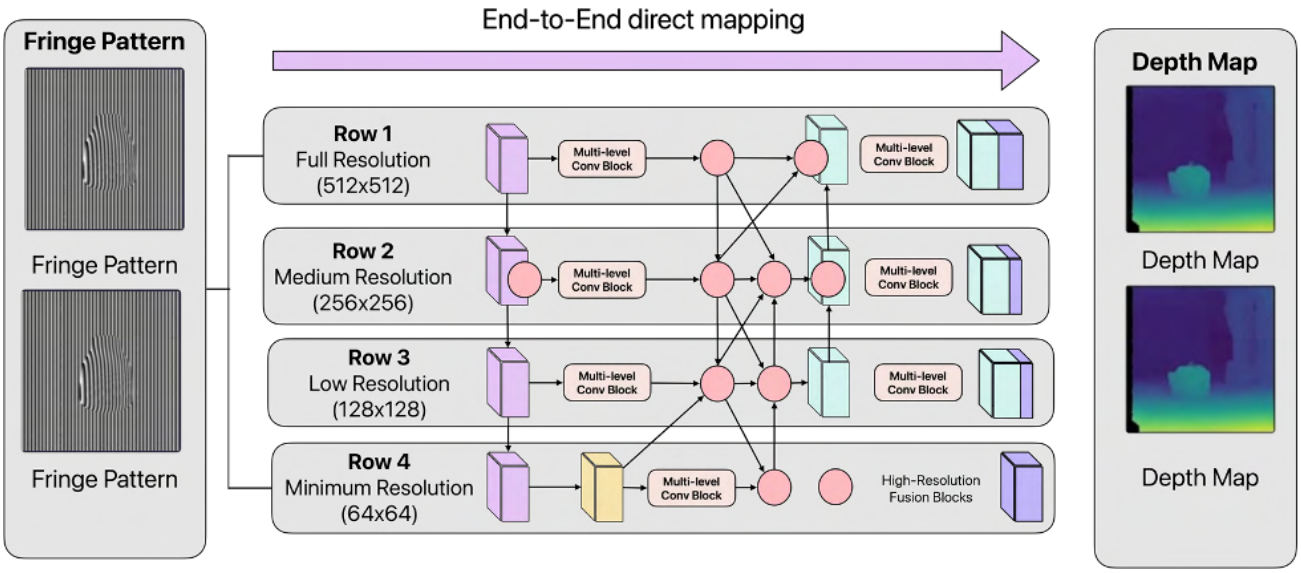


Figura 13: Arquitectura UHRNet para el mapeo directo de franja a profundidad. Fuente propia

La red UHRNet usa como base (backbone) una estructura clásica de codificador-decodificador de la red U-Net [39]. Durante el proceso de la codificación, la imagen pasa por cuatro niveles en donde el número de canales del mapa de características aumenta de manera progresiva (64C, 128C, 256C, 512C hasta 1024C), mientras sucede esto la resolución espacial se va reduciendo a $\frac{1}{2}$, $\frac{1}{4}$, $\frac{1}{8}$ y $\frac{1}{16}$ de la imagen original [39]. En la parte de la capa final, una convolución de $1X1$ hace la tarea de mapear los canales restantes hacia un único canal el cual representa el mapa topográfico 3D [39]. Para poder evitar la perdida de detalles finos loc cuales ocurren en la U-Net durante la etapa del submuestreo, la arquitectura de UHRNet introduce dos bloques matemáticos que son fundamentales:

A. Bloque de Convolución Multinivel (MLB - Multi-Level convolution Block)

Este bloque ayuda a reemplazar las convoluciones estándar con el fin de poder expandir el campo receptivo sin perder la resolución espacial. De manera matematica, el tensor de la entrada se divide en cinco ramas paralelas [39]:

- La rama izquierda es la que actúa como conexión de salto o mejor conocida como skip connection, su principal funcionamiento es asegurar que la propagación de la información superficial y los gradientes sea fluida [39].
- La cuatro ramas derechas como se muestran en la figura 13 aplican convoluciones las cuales son dilatadas con distintas tasas las típicas son (0,1,2 y 5). Esta dilatación ayuda a insertar espacios entre los valores del Kernel de convolución permitiendo que la red pueda extraer de manera simultanea las características locales las cuales son los detalles de la superficie también las globales que es la profundidad general. Finalmente se tiene como resultado que todas las ramas se concatenan [39]

B. Bloque de Fusión de Alta Resolución (FB - High-resolution Fusion Block)

Esta parte es fundamental debido a que sustituye las conexiones de salto simple que hay entre el codificador y el decodificador en los niveles 1, 2 y 3, para este caso el 4 se mantiene la conexión normal, esto se debe a su baja resolución [39]. Esto se debe a que los mapas de las características provienen de distintos módulos de codificación y tienen distintos tamaños, este bloque ayuda a ajustar matemáticamente sus dimensiones para que estas se puedan fusionar [38]. Para poder lograr este propósito es necesario aplicar operaciones de submuestreo (Downsampling) para poder reducir matrices grandes asi como operaciones de sobremuestreo con ayuda de convoluciones transpuestas (ConvTranspose2D) con el fin de ampliar las matrices pequeñas [6][38]. Al sumar de manera analítica las matrices igualadas, dicha red logra una alta integración que se tiene entre la información global profunda así como los detalles superficiales de alta resolución, sin perder los bordes finos del objeto 3D [38].

5.6.1. Motor Matemático: La función de pérdida compuesta

El principal funcionamiento de la UHRNet es que no optimiza sus pesos (W) basándose unicamente en el error estándar numérico general, ayudándola a tener un buen rendimiento. El modelo redefine el problema de optimización introduciendo la función de pérdida compuesta, la cual tiene como tarea atacar de manera simultanea las discrepancias numéricas y las deformaciones estructurales [38]. Esta se define de la siguiente manera como se muestra en la ecuación (28) [38].

$$Fussion_Loss = chunnked_L2_Loss + \alpha \cdot SSIMLoss \tag{28}$$

Donde α se refiere a un parámetro de ajuste el cual se enfoca en equilibrar ambas magnitudes, el cual es establecido empíricamente en $\alpha = 1000$ para poder igualar la

escala que se tiene de ambas métricas [38]. A continuación se desglosan cada uno de sus componentes:

A. Pérdida $L2$ fragmentada (chunked L2 Loss): se basa en la pérdida cuadrática media tradicional la cual tiende a ignorar los errores severos concentrados en pequeñas áreas complejas del objeto. Para poder solucionarlo, la red UHRNet divide el mapa de altura predicho y el valor real el cual es conocido como Ground Truth (GT) en una cuadricula de 16 parches o fragmentos [38]. Como primer punto, se calcula el error $L2$ independiente para cada parche i, como se muestra en la ecuación (29)[38].

$$L2_loss_i = \sum_{j=1}^{n} (y_i - \hat{y_j})^2 \tag{29}$$

Donde y_i representa a la altura real y $\hat{y_j}$ se basa a la altura predicha de los n pixeles del parche [38]. Como siguiente paso, dicho algoritmo implementa un mecanismo de reordenamiento dinámico: este se basa en que después de cada iteración (epoch), ordenas los 16 parches dependiendo de la magnitud que tenga su error asignandole un peso matemático dinámico $weight_i$, en cual se muestra en la ecuación (30) y esta definido por:

$$weight_i = i \times 0{,}2 - 0{,}1 \tag{30}$$

Los valores que se toman para la componente i=1,2,3...16 [39]. Finalmente se calcula el error global fragmentado como una suma ponderada como se muestra en la ecuación (31).

$$chunked_L2_Loss = \sum_{i=1}^{16} weight_i \times \frac{L2_loss_i}{16} \tag{31}$$

Dicho mecanismo matemático hace que la red pueda prestar atención de manera estadística la cual es mucho mayor a la regiones topográficas que son mas difíciles de reconstruir, evitando el estancamiento en el entrenamiento [38].

B. Pérdida por Similitud Estructural (SSIMLoss)

La métrica $L2$ tiene como desventaja que es muy sensible a las anomalías, teniendo como fallas en la captura de la percepción visual del objeto [38]. Por ciertas desventajas, se incorpora el índice de Similitud Estructural ($SSIM$), la cual se basa en evaluar la concordancia entre la altura real (x) y la altura predicha (y) tomando en consideración tres factores estadísticos: la luminancia/media (μ), la varianza (σ^2) y la covarianza cruzada (σ_{xy}) [39]. Esta métrica se formula de manera matemática como se muestra en la ecuación (32).

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{32}$$

Donde c_1 y c_2 son constantes de estabilización matemática que tiene el propósito de evitar las divisiones por cero [38]. Como el valor del $SSIM$ puede variar entre 0 y 1,

significando que cuando es 1 equivale a que tiene una similitud perfecta. En este caso el término minimizar dentro de dicha función se define como en la ecuación (33) [38].

$$SSIMLoss = 1 - SSIM(x, y) \tag{33}$$

Cuando se integra el modelo estructural SSIMLoss con el cálculo focalizado del chunked L2 Loss, el entrenamiento de la red UHRNet converge en una solución matemática la cual es mas robusta frente a los atípicos [37].

5.7. Arquitecturas complementarias en el procesamiento de franjas y mapas de profundidad

Ademas de los modelos que son basados en topología U-Net y Transformers, recientemente en la literatura se han explorado nuevas arquitecturas profundas las cuales son orientadas a resolver problemas específicos dentro del proceso de la reconstrucción 3D, como lo es el desvanecimiento del gradiente, la generación de los mapas de fase de mayor resolución y e aislamiento del ruido periódico complejo. A continuación se explicaran dichas arquitecturas así como sus fundamentos matemáticos.

5.7.1. Redes Residuales Profundas (ResNet) y Bloques Residuales

Cuando se entrenan Redes Neuronales Profundas (DCNN) para el análisis de franjas uno de sus mayores retos son los problemas en la degradación el cual también es conocido como desvanecimiento del gradiente, a medida de que se aumenta la profundidad de la red las cuales son la cantidad de capas para poder captar las características mas complejas, la tasa de error tiende a aumentar esto se debe a que los gradientes matemáticos se vuelven demasiado pequeños que no funcionan para actualizar los pesos de manera correcta [1].

Para poder solucionar este problema, se tomaron en cuenta las Deep Residual Nets (ResNet) dicha arquitectura tiene como innovación la incorporación de conexiones de acceso directo (shortcuts o skip connections residuales) [1]. Estas redes se han integrado como bloques extractores en los algoritmos de recuperación de fase a partir de un solo disparo [10].

La red residual obliga a que las capas puedan aprender una función de mapeo residual $\mathcal{F}(x)$, en vez de forzar a que un conjunto de dichas capas convolucionales intenten aprender la función directa $\mathcal{H}(x)$ [1][10]. La salida que tiene un bloque residual de define de manera matemática como en la ecuación (34).

$$y = \mathcal{F}(x, \{W_i\}) + x \tag{34}$$

Donde x representa la entrada al bloque que en este caso es el mapa de características del patrón de franjas, $\mathcal{F}(x, \{W_i\})$ es la transformación aprendida por las capas convolucionales intermedias con pesos W_i y el termino $+x$ es la conexión de saltos los cuales suman la entrada original a la salida de la función [1][10]. Cuando se aplica la

función de activación ReLU, la salid que se obtiene al final es $ReLU(y)$. Dicha formulación matemática nos garantiza que la red ya ha aprendido una representación óptima, la cual tiene como función hacer que los pesos de $\mathcal{F}(x)$ tienda a cero, dejando que pase la identidad x permitiendo que se pueda hacer entrenamiento de fases que son extremadamente profundas sin perdida de información, ayudando que la predicción exacta de los numeradores y denominadores de la función arcotangente [1][10].

A conitinuación se muestra la arquitectura de dicha red conocida con ResNet-34 como se muestra en la figura 14, así como la parte de su bloque residual básico (fig 15)

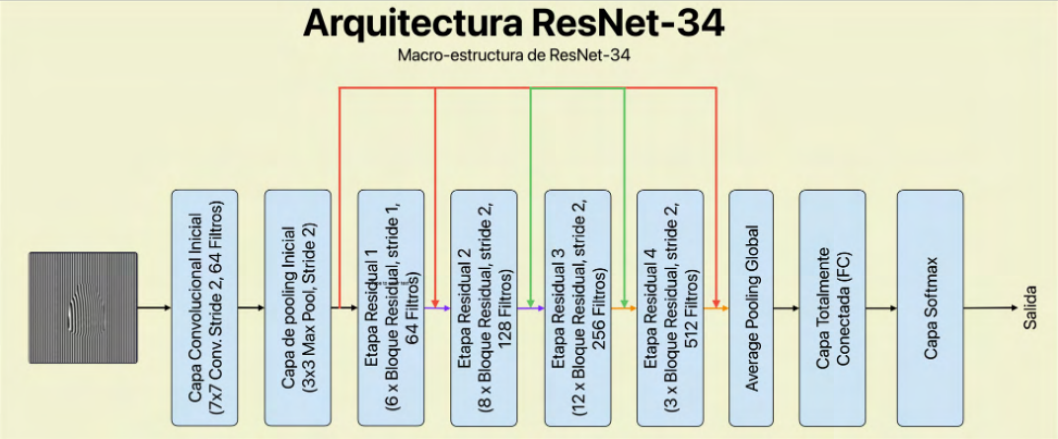


Figura 14: Arquitectura ResNet-34. Fuente propia

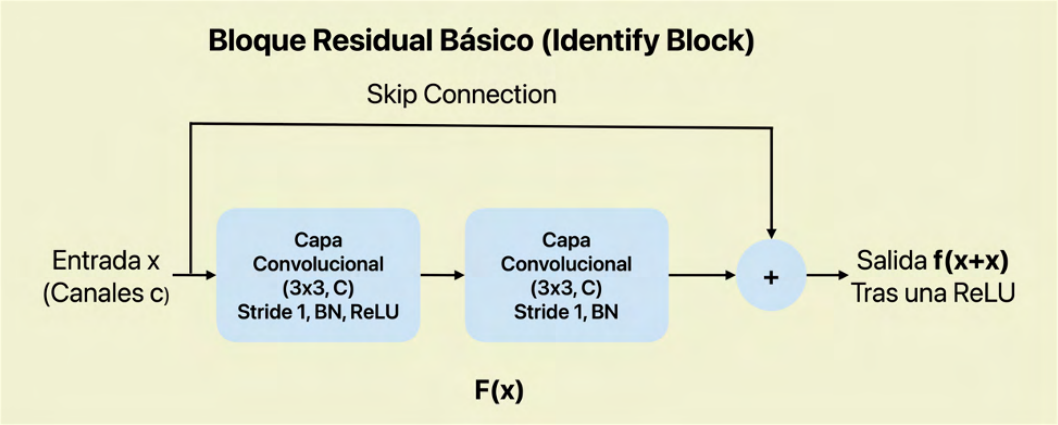


Figura 15: Bloque Residual Básico (Identify Block). Fuente propia

5.7.2. Redes Neuronales Tripletas (Triplet Networks)

Este tipo de redes son una variante de las redes que conocidas como siamesas las cuales son importantes para aprender las relaciones de similitud y diferencias topográficas,

son principalmente utilizadas para poder comparar las características que son extraídas de mapas de profundidad o reconstrucción 3D [4][8].

Dichas redes están compuestas por tres CNNs las cuales son idénticas su funcionamiento es en paralelo y comparten los mismos pesos. El sistema recibe tres imágenes en la entrada de manera simultanea: una imagen Ancla (I_a), una imagen positiva (I_p) la cual tiene la misma topografía que la imagen Ancla, también cuenta con una negativa (I_n) que representa un objeto con fallos en la fase [8][4]. La red se enfoca en comprimir estas imágenes en vectores d_a, d_p y d_n.

Esta red su aprendizaje se rige por la función de pérdida de triplete, esta obliga a que el algoritmo de optimización acerque a los descriptores similares en el espacio euclidiano y a los distintos los aleje de este mismo [4]. Se expresa de manera matemática como la ecuación (35)

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N}[D_{i,p}^{a} - D_{i,n}^{a} + m]_{+} \tag{35}$$

donde $D_{i,p}^{a} = ||d_a - d_p||_2^2$ representa a la distancia euclidiana que hay entre el ancla y el positivo. Por otro lado $D_{i,n}^{a} = ||d_a - d_n||_2^2$ simboliza a la distancia euclidiana entre el ancla y el negativo, m hace referencia a un margen de tolerancia el cual esta predefinido este tiene como propósito que la red pueda separar las muestras a la distancia que tiene predefinida [4][8] y por ultimo $[...]_+$ este equivale a la función de activación $ReLU(x) = max(0, x)$, nos asegura que la perdida sea igual a cero si la condición de separación ya se cumple [8].

A continuación, se presenta un diagrama de dicha red en la figura 16, en el cual se puede mostrar su funcionamiento

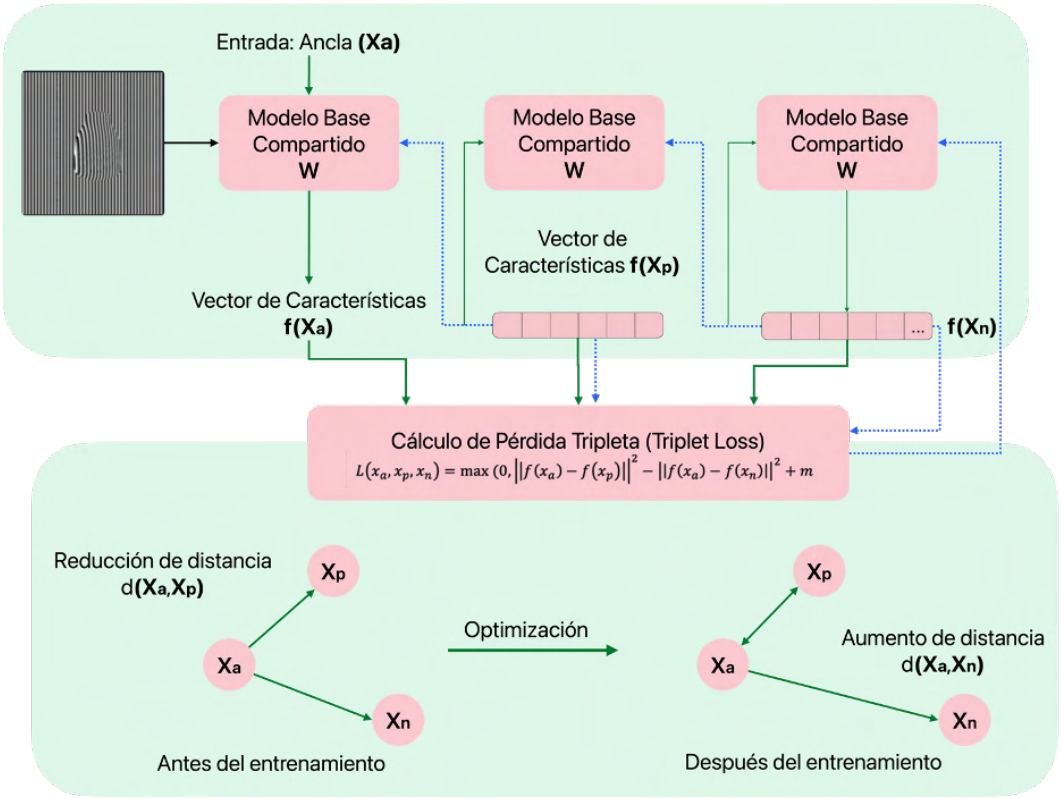


Figura 16: Arquitectura de las Redes Neuronales Tripletas. Fuente propia

5.8. Redes Generativas Antagónicas (GANs y Cycle-GAN)

En la parte del proceso de recuperación de fase e imágenes holográficas, se han implementados las Redes Generativas Antagónicas mejor conocidas como GAN, este tipo de redes tienen como propósito generar mapas de fase con una buena resolución a partir de patrones de intensidad que tiene baja resolución o también para realizar imágenes de fase cuantitativa sin ruido [37].

Su funcionamiento matemático de una Red GAN se basa en la suma cero (min-max) entre dos redes neuronales de las cuales cuentan de un Generados (G) y de un Discriminador (D). El Generador en la perfilometría intenta predecir un mapa de fase de alta calidad a partir de un patrón de franjas ruidoso, por otro lado, el Discriminador ayuda a evaluar si el mapa de fase es real, esto quiere decir que es proveniente del Ground Truth o falso si este es generado por la red. A traves de dichas arquitecturas como lo son la Cycle-GAN, dicha red puede aprender el mapeo topográfico sin la necesidad de pares de imágenes estrcitamente alineadas [37].

La red Generadora (G) lo que hace es minimizar la función de valor, por otro lado el Discriminador (D) maximiza esta función. La función de perdida antagonica convencional se puede definir de la siguiente manera (36):

$$\min_{G} \max_{D} V(D,G) = \mathbb{E}_x[logD(x)] + \mathbb{E}_z[log(1 - D(G(z)))] \tag{36}$$

Donde x nos representa el mapa de fase real y z hace referencia al patrón de entrada con franjas ruidosas. Para aplicaciones ópticas, dicha pérdida antagónica no se usa sola, sino que esta se suma a la métrica L_1 o L_2 que son las normas de error con el fin de asegurar la fase generada coincida con la topografía que tiene el objeto físico [37].

A continuación se muestra el diagrama de una Arquitectura Generativa Antagonica (GAN) (17), en la cual se puede observar que cuenta con 2 elementos fundamentales, el Generador el cual captura la la distribución de probabilidad de los datos reales con el fin de poder sintetizar nuevas muestras aceptables, se recibe como entrada un vector de ruido aleatorio z el cual pertenece a un espacio latente el cual no es estructurado así produciendo muestras sintéticas $G(z)$.

Por otro lado, el Discriminador se comporta como un clasificador binario el cual esta diseñado con el fin de poder evaluar la autenticidad que tienen las muestras, acá se reciben los datos reales del conjunto de entrenamiento x como datos sintéticos $G(z)$ y se devuelve una probabilidad escalar el cual representa si la muestra que se analizo es autentica significando que es verdadero o fue generada por el modelo en este caso falso [37].

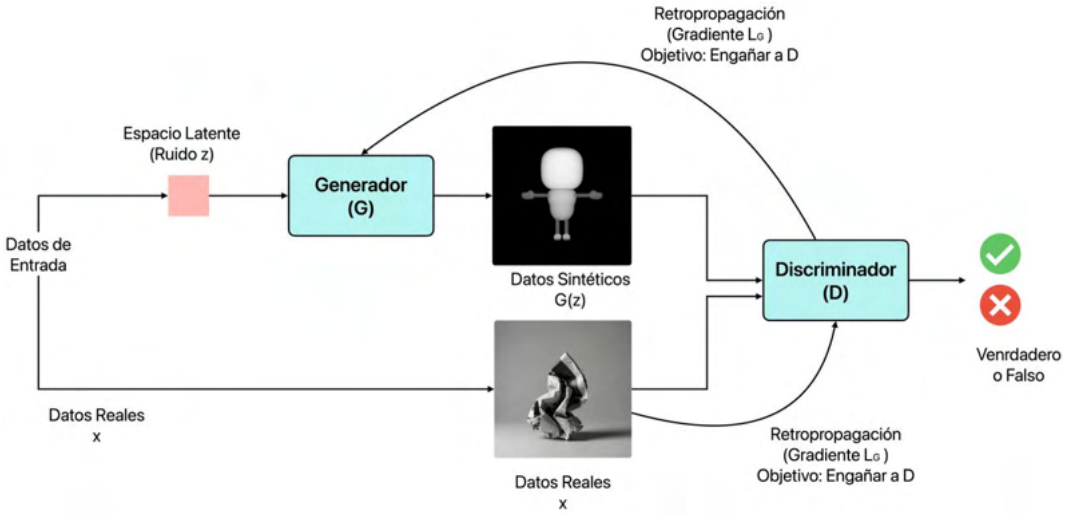


Figura 17: Arquitectura GAN Convencional. Fuente propia

Durante la etapa del entrenamiento, el Discriminador intenta maximizar la probabilidad de clasificar de manera correcta las muestras tanto reales como falsas, mientras que el Generador ajusta sus pesos con ayuda de la retropropagación con el fin de poder minimizar esta probabilidad, engañando de manera eficaz a D. La convergencia del sistema se alcanza cuando se logra un equilibrio de Nash, esto significa que la distribución de los datos generados imita con una alta fidelidad a la distribución de los datos reales [37].

También existe el mapeo de dominios no supervisado mediante Cycle-GAN este es usado en aplicaciones en las cuales se requiere la traducción de la imagen de un dominio fuente (A) a un dominio que va de destino (B) sin que se cuente con datos emparejados paso a paso uno de los ejemplos para este caso puede ser la transformación de imágenes de patrones de franjas que están distorsionadas a mapas de fase continuos o mapear las texturas a profundidad,esta arquitectura conocida como Cycle-GAN (Cycle-Consistent Adversarial Network) ofrece una solución sin supervisar lo que es la pareada directa [37][5].

En la figura 18 se muestra la arquitectura de la Red cycle-GAN para el mapeo de dominios:

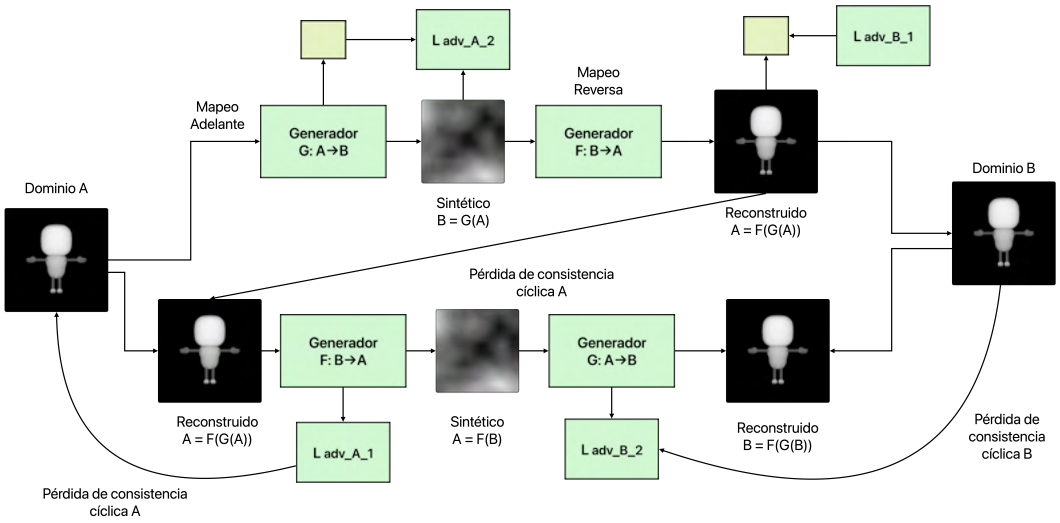


Figura 18: Arquitectura de Red Cycle-GAN para Mapeo de dominios. Fuente propia

A diferencia de una arquitectura GAN convencional, esta arquitectura mostrada en la figura 18 introduce dos pares Generadores y Discriminadores los cuales operan de manera acoplada:

a. **Mapeo Adelante** ($A \rightarrow B$): representa a un generador $G : A \rightarrow B$ que transforma las muestras del dominio A al dominio B. En este caso, su autenticidad es evaluada por un Discriminador D_B.

b. **Mapeo Reverso** ($B \rightarrow A$)**:** este es un segundo generador $F : B \rightarrow A$ el cual realiza la transformación opuesta, esta calidad es validada por el discriminador D_A.

Uno de los principales desafíos en el mapeo entre los dominios no supervisados es la posibilidad de que el generador sufra un fenómeno el cual es conocido como mode collapse esto puede tener como consecuencia que se distorsione el contenido estructural de la escena al enfocarse unicamente a engañar al discriminador. [37]. Para poder resolver esta ambigüedad matemática y garantizar que una imagen mapeada mantenga sus

características geométricas subyacentes, se introduce la pérdida de consistencia cíclica [38].

La pérdida de consistencia cíclica establece que si la muestra que esta en el dominio A se traduce al dominio B haciéndolo a través de G, posteriormente se vuelve a traducir al dominio A mediante F, el resultado ya reconstruido deber ser idéntico a la imagen original numéricamente $F(G(A)) \approx A$) y para el dominio B se exige que $G(F(B)) \approx B$ [37].

La función de pérdida total de Cycle-GAN esta definida mediante la suma ponderada de las pérdidas antagónicas tradicionales y la pérdida de consistencia de ciclo, como se muestra en la ecuación (37) [37].

$$\mathcal{L}_{CycleGAN}(G, F, D_A, D_B) = \mathcal{L}_{adv}(G, D_B, A, B) + \mathcal{L}_{adv}(F, D_A, B, A) + \lambda \mathcal{L}_{cyc}(G, F) \quad (37)$$

Donde la pérdida de consistencia de ciclo se representa de manera manera matemática de la siguiente manera como se muestra en la ecuación (38), a través de la norma L_1 [37]:

$$\mathcal{L}_{cyc}(G, F) = \mathbb{E}_{a \sim pdata(a)}[\|F(G(a)) - a\|_1] + \mathbb{E}_{b \sim pdata(b)}[\|G(F(b)) - b\|_1] \quad (38)$$

Donde el hiperparámetro λ ayuda a controla la relevancia relativa de la preservación del contenido tomando en cuenta la efectividad de la transformación antagónica entre los dominios.

VI. Material y Métodos o Metodología

En el presente capitulo se detalla y describen todas las etapas que se llevaron a cabo para la metodología empleada de dicha investigación, así como las herramientas y las fases implementadas para el desarrollo del proyecto, desde la creación de los datos sintéticos, hasta la implementación, entrenamiento y evaluación de la reconstrucción tridimensional mediante perfilometría de fourier y redes neuronales convolucionales. Como primer punto, se describe el proceso de la generación de la base de datos sintéticas y reales, las cuales son de suma importancia para realizar un análisis comparativo entre el método tradicional y el modelo basado en el uso de redes neuronales convolucionales.

Para el desarrollo de la base de datos sintética se implementó un flujo de procesamiento tanto en blender como en python con el fin de poder producir proyecciones de franjas, máscaras binarias, mapas de altura así como las fases envueltas con ayuda de los métodos de tres pasos y la Transformada de Fourier (FT). Dicho proceso permitió que se pudiera tener un control de las características como lo son los parámetros geométricos, las condiciones de la iluminación, ruido aditivo así como las variaciones en la orientación del proyector, esto con el fin de poder garantizar que los objetos sean lo suficientemente diversos para usarlo en el entrenamiento. De manera complementaria, se incorporaron objetos reales, en los cuales se utilizo una configuración interferométrica equivalente a la que se empleo en las generación de los datos sinteticos, de los cuales se obtuvieron distintos mapas low-wrap y multi-wrap en los cuales se incorporaron distintas perturbaciones como lo son el ruido, regiones que están parcialmente ocluidas y variaciones espaciales de la iluminación proyectada.

Por otro lado, se implementaron distintos algoritmos de desenvolvimiento de fase como lo fueron Itoh, Ghiglia-Romero, Zhao y Andreas en los objetos reales con el fin de poder hacer la comparación entre el métodos tradicional frente al método con redes neuronales convolucionales.

Como siguiente paso, se implementaron ciertas métricas cuantitativas como lo son el Error Cuadrático Medio en Imagenes (IMMSE), la relación de señal-ruido de pico (PSNR) y el Índice de Similaridad Estructural (SSIM), con el propósito de poder evaluar la calidad del objeto cuando se le aplica el método con el uso de la inteligencia artificial y comparar con los objetos reales aplicando el método tradicional.

Como parte de la metodología también se contempla el diseño, la implementación y el entrenamiento de tres arquitecturas de CNN las cuales son orientadas al procesamiento como la reconstrucción de la información de la fase: una de las arquitecturas usadas es la UHRNet la cual es caracterizada por procesar flujos paralelos de alta y baja resolución, la segunda es la SFNet esta basada en la estructura codificador piramidal la cual es complementada con el módulo de agregación de características conocido como (FAM) y por ultimo la Hformer esta es una arquitectura híbrida las cuales combina capas convolucionales con ciertos mecanismos de aatención en ventanada (Window Attention). Para la parte del entrenamiento de dichos modelos se optimizaron los parámetros con ayuda de los algoritmos Adam y AdamW a las tasas de aprendizaje que son ajustadas para cada red, se empleo funciones de activación no lineales como la ReLU y GELU, con

un tamaño de lote unificado ejecutado sobre aceleración por GPU usando una tarjeta gráfica de NVIDIA A100, cuyos hiperparametros como detalles estructurales.

Como punto final se describe como fue el proceso para la realización de este proyecto: estandarización y normalización de los datos, control de versiones y comparaciones entre ambos métodos.

Con el proposito de poder ilustrar la arquitectura integral de dicha investigación, en la figura 19 se muestra el flujo de la Perfilometría por Transformada de Fourier mediante el método tradicional haciendo comparación con el método basado en Aprendizaje Profundo. En este caso la vía analítica depende de las etpas consecutivas de filtrado en frecuencia y desdoblamiento espacial de fase la cual resulta vulnerable ante las discontinuidades abruptas, la red neuronal incorpora distintas topologías especializadas (UHRNet, SFNet y Hformer) las cuales son capaces de inferir la discontinuidad topografica tridimensional directamente desde el patrón de franjas o la fase envuelta, mejorando la fidelidad geométrica frente a los modelos de referencia

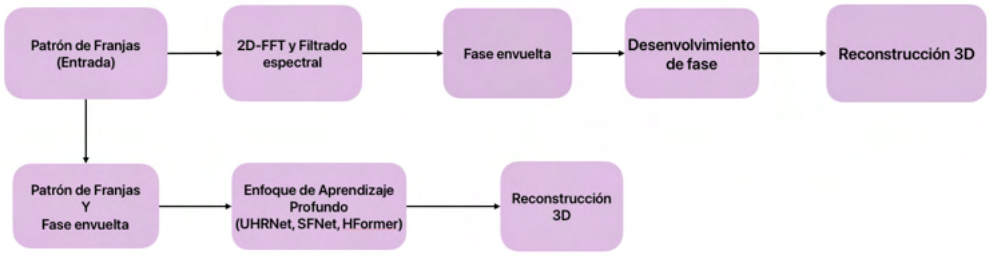


Figura 19: Diagrama de la metodología usada. Fuente propia

6.1. Base de datos

En esta bloque se describe como se realizo la base datos. Para el entrenamiento y la evaluación de las arquitecturas es necesario contar con un conjunto de datos los cuales combinen el Ground Truth con los patrones ópticos que representan las condiciones reales. Para poder hacer estos requerimientos se estructuro una base de datos sintética la cual fue creada en Blender con el propósito de poder controlar las propiedades geométricas de referencia, también se tomo en cuenta un conjunto de imágenes reales.

6.1.1. Base de datos sintética

La base de datos sintética fue creada en un software basado en la generación de modelos tridimensionales conocido como Blender versión 4.3.2 con ayuda de python para automatizarla y poder simular un sistema real de proyección de franjas sinusoidales. Para generar dichas imágenes se ocupo mallas poligonales de libre acceso las cuales fueron obtenidas del repositorio TurboSquid, para la selección de los objetos se tomo en cuenta las variaciones en relieve, curvaturas, variabilidad geométrica, asegurando que

exista la ausencia de cavidades internas, esto se debe a que se usa un material de tipo Spline para poder renderizar el Ground Truth (GT), también con las características mencionadas anteriormente nos permiten una interpolación continua del relieve superficial. El Ground Truth, es de suma importancia debido a que representa la verdad geométrica absoluta de la superficie tradicional, en la figura 20 se muestra como esta constituido en Ground Truth en la base de datos.

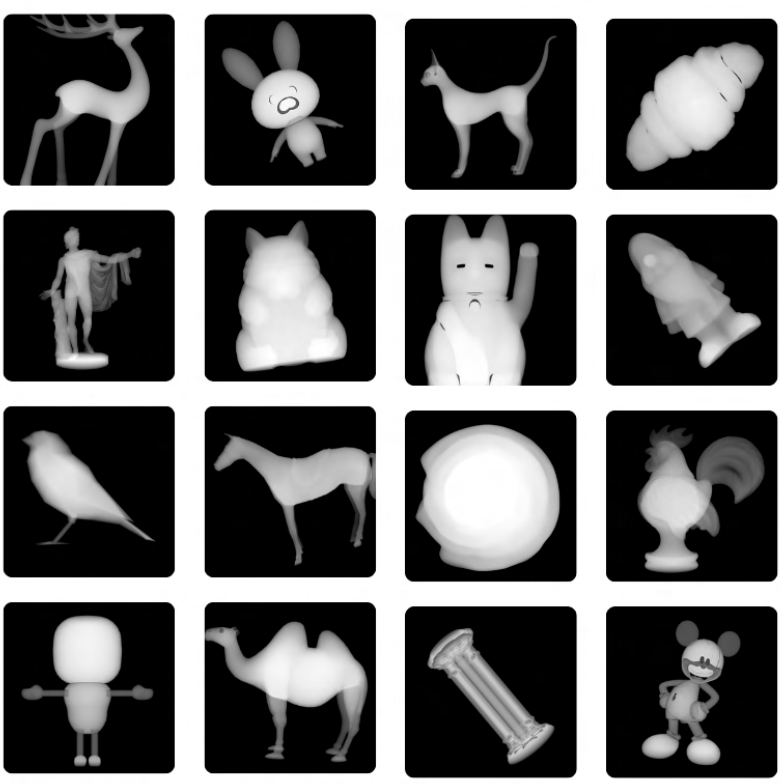


Figura 20: Conjunto de muestras del GT. Fuente propia

6.1.2. Composición y Generación de cada objeto

Para la creación de cada objeto que se integra en la base de datos sintética se implemento un proceso sistemático de generación y renderizado para poder crear un conjunto de objetos estandarizado los cuales fueron exportados en PNG sin compresión. Este proceso cuentas con las siguientes etapas:

- **Patrones de Franjas Moduladas con Desviación de Fase:** en esta fase se crearon tres proyecciones de franjas sinusoidales deformadas sobre la superficie del objeto estas son correspondientes a la desviación de la fase de 240°, 0° y 120°. Dicha modulación por pasos de fase es de suma importancia esto se debe a que complementa el análisis óptico enriqueciendo la información espacial que se necesita para obtener una extracción tanto precisa como robusta de la fase envuelta contrastándolo con los métodos de una sola captura.

- **Plano de referencia:** ayuda a capturar el patrón periódico que es proyectado sobre una superficie plana sin deformaciones. Esta tiene la capacidad de suministrar la frecuencia portadora espacial (f_0) la cual es requerida para poder sustraer la fase de fondo y así poder aislar solo la componente modulada.

- **Imagen Renderizado Original:** Esta es la representación visual del objeto con ayuda de la iluminación virtual sin el patrón de franjas, esto fue de gran utilidad para hacer un buen análisis morfológico directo y tareas de segmentación o extracción de las características de intensidad.

- **Ground Truth:** para este ultimo paso se generaron dos mapas de Ground Truth diferenciados a los cuales se le agregaron dos tipos de materiales procedimentales conocido como spline distintos sobre la geometría. Dichos materiales modelan y codifican de punto a punto la distribución de la altura y profundidad del objeto.

En la figura 21 se muestra como es la composición de cada uno de los objetos mostrando el plano de referencia estos se muestran apilados reflejando las proyecciones con desfase angular controlado (0°,120° y 240°), en el centro se muestra el objeto ya con las franjas proyectadas y por ultimo es el Ground Truth el cual nos va a ayudar en la parte del entrenamiento de la redes neuronales convolucionales, en este caso las zonas que se ven mas claras muestran que tienen una mayor proximidad al plano de proyección o cámara y las zonas que se ven con tonos oscuros representan que tienen menor relieve.



Figura 21: Composición del conjunto de imágenes por Objeto. Fuente propia

6.1.3. Especificaciones Técnicas y Adquisición de los datos

Todas las imágenes tienen una resolución de $512x512$ pixeles en una escala de grises de 16 bit, fueron exportadas en un formato matricial sin compresión. El uso de la profundidad de 16 bits fue de suma importancia para poder mantener la continuidad en las variaciones graduales de altura del Ground Truth y poder evitar que existan errores en la cuantización.

Por otro lado, no se aplico ningún tipo de normalización, ni se agrego ningún tipo de ruido sintético. Esto se hizo con el propósito de poder garantizar al momento de hacer el entrenamiento supervisado sea sobre la información geométrica pura sin ninguna alteración.

Cada objeto se aplicaron 9 rotaciones sucesivas las cuales fueron incrementando en 40° (0°, 40°,80°,...,320°), esto se hizo con el propósito de poder enriquecer la base de datos, permitiéndonos evaluar el comportamiento de los patrones de franja así como las sombras proyectadas en distintas perspectivas.

Otro de factores importantes es el costo computacional, el tiempo de renderizado de cada uno de los objetos fue en un intervalo de 5 a 40 minutos, esto varia debido la densidad poligonal de la malla, la curvatura superficial así como la complejidad morfológica global de la escena.

6.1.4. Mapas de fase envuelta

Tomando en cuenta las imágenes con franjas proyectadas (p_1, p_2, p_3) así como el plano de referencia homogéneo se obtuvo el mapa de fase envuelta mediante el método de perfilometria por transformada de fourier bidimensional (FTP2D).

Estos mapas de fase son de suma importancia en la Perfilometria por Transformada de Fourier debido a que es el núcleo analítico de este método ya que es el portador directo de la topografía 3D, cuando se proyecta un patrón de franjas sobre la superficie, las variaciones que se tienen en el relieve deforman de manera geométrica las lineas. Dicha distorsión física mencionada anteriormente se modula desde la fase $\phi(x, y)$ provocando que el mapa de fase resultante sea proporcional a la profundidad $z(x, y)$ del objeto con respecto al plano de referencia [20].

Otro de los factores es el aislamiento basado en la reflectancia e iluminación, en este caso la imagen que es capturada $I(x, y)$ esta afectada por ciertos componentes como lo es la reflectancia del material $b(x, y)$ y por la iluminación de fondo no homogénea $a(x, y)$ en el dominio de la frecuencia. Cuando se hace el proceso de trasladar la señal al dominio de fourier se aisla la componente portadora con ayuda de un filtrado selectivo provocando que el calculo de la fase desacople el nivel continuo (DC) así como los efectos de contraste local preservando la información del objeto sintético sin errores [20]. Este es un eslabón indispensable para realizar la reconstrucción continua esto es debido a forma matemática de la función arcotangente como se muestra en la ecuación (7) en este caso la fase extraída es discontinua esto se debe a que la fase envuelta esta en el intervalo $[-\pi, \pi]$, el mapa tiene como función ser la entrada en donde operan los algoritmos de desenvolvimiento de fase con el fin de remover los saltos de 2π y así se

pueda reconstruir una superficie tridimensional continua sin embargo si existe alguna distorsión o ruido el cual no fuera filtrado en el mapa afecta de manera directa la topografía final provocando que no se haga una buena reconstrucción [20].

Al hacer ese método se extrae la fase como se menciono anteriormente, dependiendo del tratamiento en la frecuencia se generaron dos variantes:

- **Fase envuelta low-wrap:** en este punto se tanto aisla como traslada el armonico fundamental al origen espectral antes de la IFFT, ayudando a generar mapas que tienen una baja densidad de saltos de 2π provocando que se recupere de manera directa la modulación topográfica [24].

- **Fase envuelta multi-wrap por FT:** en este caso es sin traslación portadora ya que se conserva el desfase de la frecuencia portadora espacial sin trasladarla al origen, esto produce que los mapas combinen las deformaciones del relieve con múltiples envolventes y saltos de fase de 2π a lo largo del dominio [24].

Se aplico un método de tres pasos (3-step) este es un algoritmo temporal de desfase tomando en cuenta la formulación de malacara, fue utilizado a partir de las proyecciones desfasadas a 0°,120° y 240°. Esto ayudo a producir los mapas de fase de tipo multi-wrap con una elevada cantidad de transiciones periódicas ≈ 25 saltos de 2π [20].

6.1.5. Base de datos con imágenes reales

Para poder comparar el desempeño de los métodos de reconstrucción entre los objetos sintéticos sin perturbaciones y los reales con perturbaciones, se utilizo una base de datos ya existente experimental la cual esta conformada con 26 objetos reales, los cuales se obtuvieron mediante una arreglo experimental de triangulación óptica la cual esta compuesta por un proyector digital el cual es el encargada de de emitir patrones de franjas con perfil sinusoidal también usaron una cámara digital orientada la cual tiene u angulo de inclinación con el fin de registrar las deformaciones morfológicas. En la figura 22 se muestran algunos de los objetos reales que componen esta base de datos.

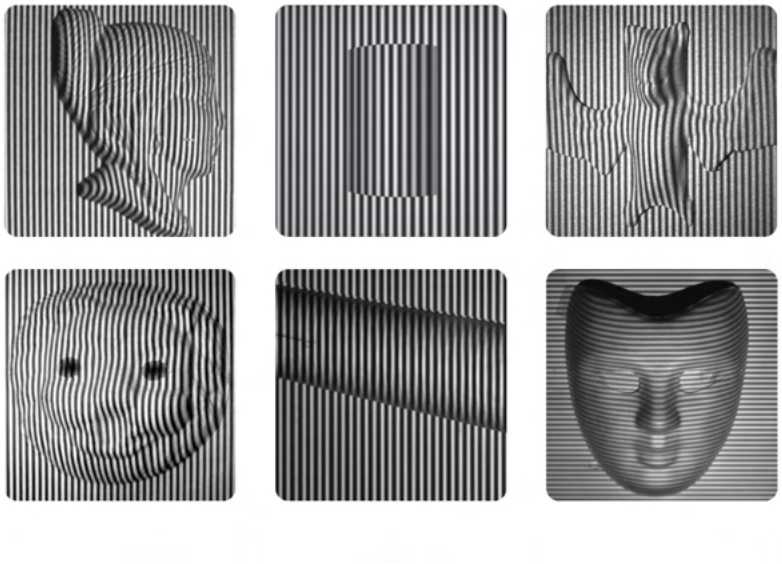


Figura 22: Ejemplo de datos reales. Fuente propia

6.2. Método Tradicional

Como siguiente parte en la metodología realizada es la implementación del método tradicional de Perfilometria por Transformada de Fourier Bidimensional (FTP2D) este se llevo a cabo en MATLAB, en donde se implementaron distintos algoritmos de desenvolvimiento de fase como lo son: Itoh, Herraez, Zhao, Andreas y Ghiglia con el fin de poder comparar los resultados de la reconstrucción a hacer uso de estos algoritmos. También se hizo uso de distintas mascaras espectrales las cuales fueron de las siguientes formas: la rectangular, elipse, circulo y diamante, cada una de ellas se usos con los distintos algoritmos de desenvolvimiento de fase respectivamente y así hacer una comparación entre ellos tanto de manera cualitativa y cuantitativa.

Se hizo una evaluación del impacto del diseño de los filtros espectrales y las robustez que tienen los algoritmos de desenvolvimiento de fase ante el ruido y las discontinuidades combinando ambos métodos.

El proceso de dicho método sigue un flujo secuencial que va mediante cinco fases principales como se muestra en la figura 23

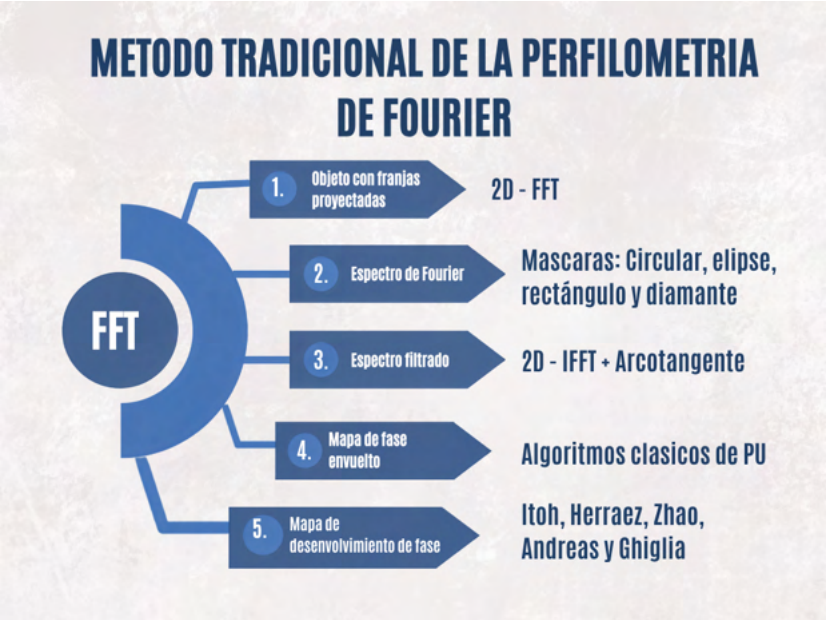


Figura 23: Método tradicional perfilometria de fourier. Fuente propia

6.2.1. Proyección de franjas

La técnica de la proyección de franjas es un método en la óptica sin contacto el cual permite que se recupere la topografía tridimensional de una superficie que tiene una resolución micrométrica y precisión subpixel. Este se basa en la triangulación óptica la cual es generada por un arreglo geométrico el cual esta compuesto por una unidad de proyección digital, una unidad de adquisición que en este caso es la cámara y un plano de referencial, como se muestra en la figura 24 [36][20].

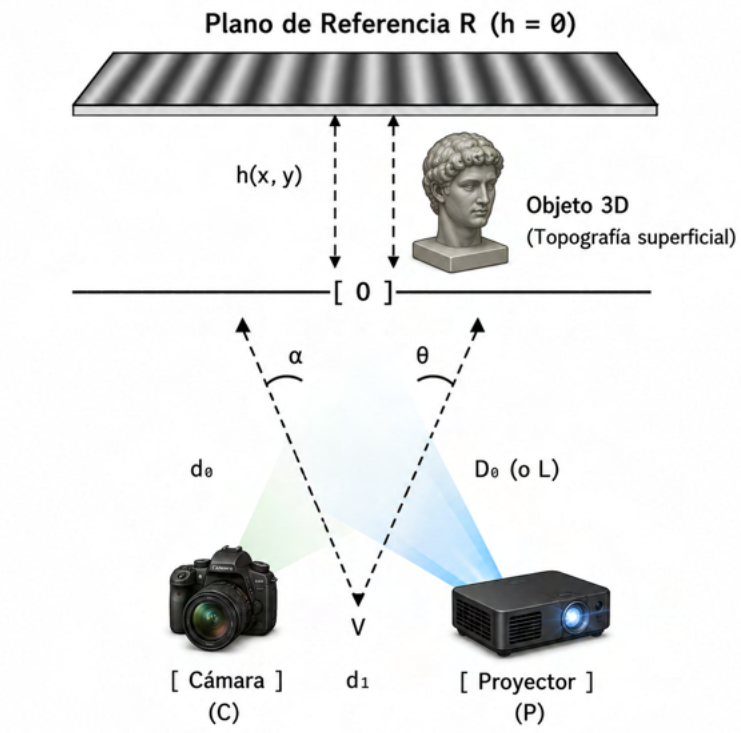


Figura 24: Proyección de franjas para objetos reales. Fuente propia

En este proceso la cámara y el proyector se encuentran situados a una distancia d_0 con respecto al plano de referencia horizontal R como se presenta en la figura 24, ambos están separados por una linea base de distancia conocida como d_1. Los ejes ópticos de ambos dispositivos convergen hacia la escena formando un ángulo de triangulación relativo θ o ángulo entre rayos α [20][36]

Cuando el patrón periódico se proyecta de manera directa sobre una superficie plana la cual esta situada $h(x, y) = 0$, la cámara se encarga de registrar un conjunto de franjas. Dicha captura suministra la frecuencia portadora (f_0) [36][24]. Por otro lado, cuando se interpone el objeto con variaciones de altura superficial $h(x, y)$ sobre lo que es el plano base sin embargo esto provoca que la trayectoria geométrica de los rayos se altera teniendo como consecuencia que los puntos de la malla física sufran un desplazamiento lateral aparente desde la perspectiva angular de la cámara, las lineas rectas originales se curvan y deforman debido a que se adaptan a la morfología local [36][20][24].

Como primer paso se toma en cuenta el objeto tanto sintético como el real con franjas, estas fueron importadas en matrices de $512x512$ en escala de grises con una profundidad de 16 bits, con esta proyección de franjas nos ayudo a ver como estaban los saltos de fase en los objetos reales. El patrón sinusoidal proyectado sobre el objeto experimento distorsiones espaciales las cuales fueron proporcionales a las variaciones de elevación y curvatura de la superficie. De manera analítica la intensidad luminosa registrada en el plano del sensor $g(x, y)$, se modela como se muestra en la ecuación (39)

[20][5]:

$$g(x,y) = a(x,y) + b(x,y)cos[2\pi f_0 x + \phi(x,y)] \tag{39}$$

La rejilla cosenoidal sufre una distorsión espacial se conoce como un desplazamiento angular dentro de la función de la intensidad, cuando se proyecta el patrón sobre el plano de referencia R, la fase que fue registrada por el sensor esta dada por $\phi_0(x,y) = 2\pi f_0 x$, al agregarle una elevación física $h(x,y)$ se produce un desfase diferencial neto, visto como la ecuación (40) [36][24]:

$$\Delta\phi(x,y) = \phi(x,y) - \phi_0(x,y) \tag{40}$$

Tomando en cuenta la semejanza de triángulos las cual esta definida por la posición relativa de la cámara como se muestra en la figura 24, el centro de emisión de proyector y el plano de referencia existe una relación matemática entre lo que es el cambio de fase y la cota vertical física $h(x,y)$ como se muestra en la ecuación (41) [20][36], también se puede mostrar como la ecuación (42) cuando se tienen configuraciones simplificadas de ejes cruzados con un angulo de inclinación (θ) el cual es constante entre los rayos de proyección y observación [36].

$$h(x,y) = \frac{d_0 \cdot \Delta\phi(x,y)}{2\pi f_0 d_1 + \Delta\phi(x,y)} \tag{41}$$

$$Z(x,y) = \frac{d_x}{2\pi tan\theta}\Delta\phi(x,y) \tag{42}$$

donde d_x representa el periodo espacial que tienen las franjas proyectadas sobre el plano de calibración. De dicha manera el cálculo preciso del mapa de fase $\phi(x,y)$ a lo largo de toda la matriz de la imagen constituye el requerimiento indispensable para poder reconstruir con fidelidad la elevación de la superficie [36][20]

6.2.2. Espectro de Fourier (2D-FFT)

La Trasnformada de Fourier permite que la información se pueda transformar al dominio espacial (x,y), en donde dichos datos se van organizando mediante coordenadas discretas de pixeles y niveles de intensidad hacia el dominio de la frecuencia (u,v), en este caso la señal se descompone en combinaciones ortogonales de función seno y coseno la cuales están parametrizadas por su amplitud y frecuencia espacial. Enfocándonos en el contexto de la perfilometria, dicha transformación ayuda a que se realice el desacople y poder aislar la fase modulada $\phi(x,y)$ de factores que puedan afectar la reconstrucción como la iluminación de fondo y las variaciones en la reflectancia.

Partiendo de que se tiene una imagen digital capturada de tamaño NxN pixeles, la intensidad se transforma al espacio de la frecuencia con ayuda de la transformada de Fourier discreta en dos dimensiones, como se muestra en la ecuación (43)[24]:

$$G(u,v) = \frac{1}{N}\sum_{x=0}^{N-1}[\sum_{y=0}^{N-1} g(x,y)e^{-j2\pi vy/N}]e^{-j2\pi ux/N} \tag{43}$$

Cuando ya se tiene el objeto con las franjas proyectadas se aplica el espectro de fourier en donde la fase angular $\phi(x,y)$ se desacopla de los errores de la iluminación continua $a(x,y)$ y reflectancia superficial $b(x,y)$, sobre la matriz de intensidad digital $g(x,y)$ después se aplicó la Transformada de Fourier Bidimensional (2D-FFT) como se muestra en la ecuación (44) en Matlab.

$$G(u,v) = \mathcal{F}\{g(x,y)\} = A(u,v) + C(u-f_0,v) + C^*(u+f_0,v) \tag{44}$$

Cuando se pasa al dominio de fourier, la información de la imagen se distribuye en lóbulos los cuales se dividen de la siguiente manera:

- **Componente de nivel continuo (DC)$A(u,v)$:** este es el lóbulo central en cual esta situado en el origen $(0,0)$ del espectro el cual se encarga de agrupar las variaciones de la iluminación ambiental así como los componentes de baja frecuencia espacial.

- **Lóbulos portadores fundamentales** $C(u-f_0,v)$ **y** $C^*(u+f_0,v)$**:** estas son las bandas laterales simétricas que están centradas en las frecuencias portadoras.

Para poder garantizar que la extracción de la fase $\phi(x,y)$ sea matemáticamente exacta y esta no represente distorsiones que sean irreversibles, es de suma importancia que el lóbulo $+f_0$ no se traslape con la componente continua central $A(u,v)$ [20].

De acuerdo con lo que formulo Takeda, dicha separación física en el espectro se cumple siempre y cuando las variaciones espaciales introducidas por el fondo $a(x,y)$, el contraste $b(x,y)$ y la fase modulada $\phi(x,y)$ sean significativamente mas lenta en comparación con la frecuencia de franjas f_0 como se presenta en la ecuación (45) [20][36][24].

$$f_{max} < \frac{1}{3}f_0 \tag{45}$$

donde f_{max} hace referencia al ancho máximo de banda que esta ocupado por el gradiente de las pendientes de la superficie. Si la frecuencia portadora conocida como f_0 es suficientemente alta, los tres lóbulos quedan aislados entre si [20][36].

6.2.3. Filtrado Espectral

Ya que se se saco el espectro de fourier como siguiente paso es aplicar un filtrado espectral, en este caso se aplicaron cuatro geometrías de mascara con el propósito de poder verificar cual de todas retiene la información y no deja pasar errores que afecten la reconstrucción.

Para poder retener esta información topográfica y poder descartar la componente de DC así como el armónico conjugado negativo $-f_0$, el algoritmo logro reconocer de

manera automática la coordenada del pico máximo del lóbulo portador fundamental (u_0, v_0). De manera inmediata se multiplicó punto a punto el espectro $G(u, v)$ por una función de transferencia pasa banda $H(u, v)$, representándose en su forma matemática como en la ecuación (46)[36][5][24].

$$G_f(u, v) = G(u, v) \cdot H(u - u_0, v - v_0) \tag{46}$$

Dentro de Matlab se programaron las cuatro geometrías para que estas puedan centrarse en la frecuencia exacta y no tomar otra frecuencia que no es también se evaluaron cada una de ellas.

Se aplico una máscara circular la cual solo toma la frecuencia como se muestra en la figura 25 pero no de pico a pico esta definida por la condición de la ecuación (47). Esta máscara ejerció una atenuación simétrica en todas las direcciones del plano frecuencial en donde se suprimió de manera uniforme el ruido que no esta correlacionado sin embargo tiene tendencia a truncar las frecuencias laterales cuando la geometría presentó pendientes asimétricas marcadas.

$$\sqrt{(u - u_0)^2 + (v - v_0)^2} \leq r_c \tag{47}$$

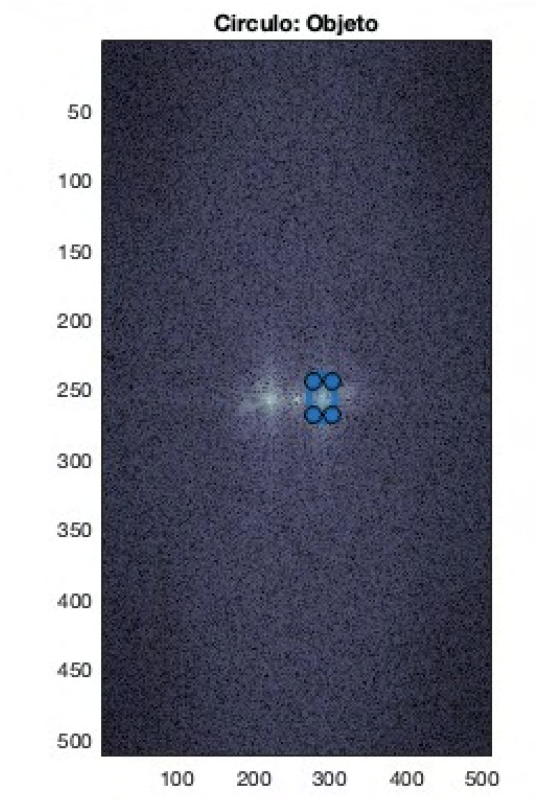


Figura 25: Máscara circular. Fuente propia

Por otro lado, se aplico una máscara elíptica esta si toma toda la frecuencia de pico a pico, esta relacionada con una función pasa banda que en su forma matemática es como el de la ecuación (48). En este caso se ajusta en función de proyección de las franjas con el fin de captar la dispersión alargada logrando que exista una delimitación que sea mas ajustada de la banda útil sin incorporar el ruido como se muestra en la figura 26.

$$\frac{(u-u_0)^2}{a^2} + \frac{(v-v_0)^2}{b^2} \leq 1 \tag{48}$$

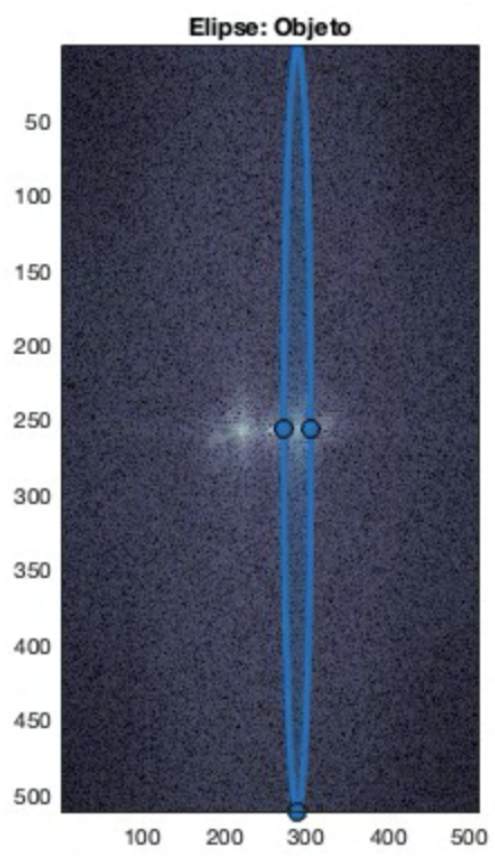


Figura 26: Máscara elipse. Fuente propia

Otra geometría que se tomo para la mascara es la rectangular como se muestra en la figura 27, esta al igual que la elipse toma en cuenta toda la frecuencia, esta delimitada por los intervalos $[u_o \pm \Delta v, v_0 \pm \Delta v]$. Se proporciona una banda pasante amplia a bajo costo computacional sin embargo esta presento un inconveniente que debido a su geometría y que toma la frecuencia de pico a pico dejo entrar errores en el momento de la reconstrucción.

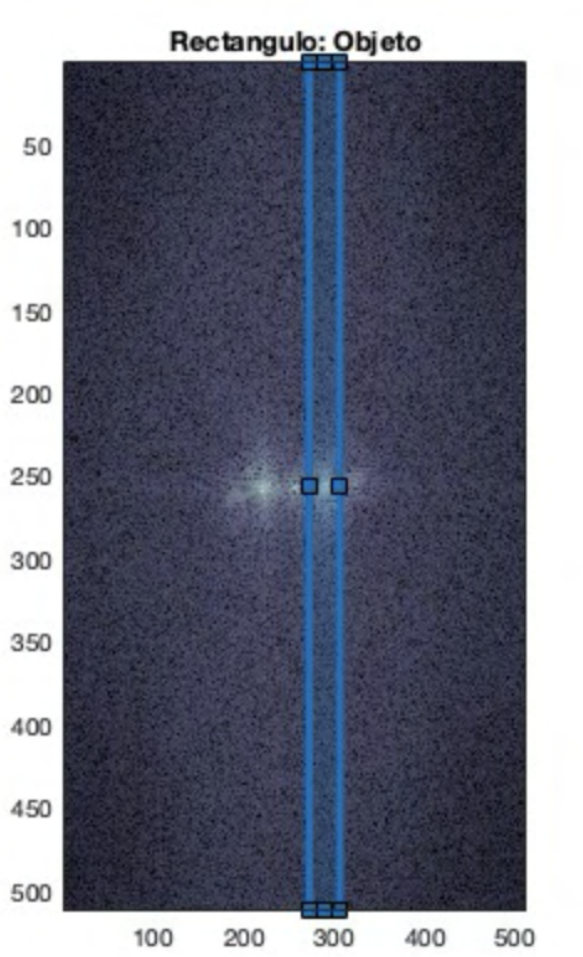


Figura 27: Máscara rectangular. Fuente propia

Por ultimo, se utilizo la mascara en forma de diamante como presenta la figura 28, esta es en función romboidal la cual esta delimitada por la norma $|u - u_0|/L_u + |v - v_0|/L_v \leq 1$. Esta forma también dio buenos resultados debido a su perfil diagonal proporcionando un decaimiento progresivo hacia lo que son las esquinas espectrales, ayudando a disminuir las distorsiones perimetrales proporcionando mapas demodulados estables.

Para la generación de las bases de datos de tipo low-wrap acá el lóbulo filtrado $G_f(u, v)$ se desplazo hacia el origen espectral $(0, 0)$, mientras que las multi-wrap conservo la ubicación desplazada en $+f_0$.

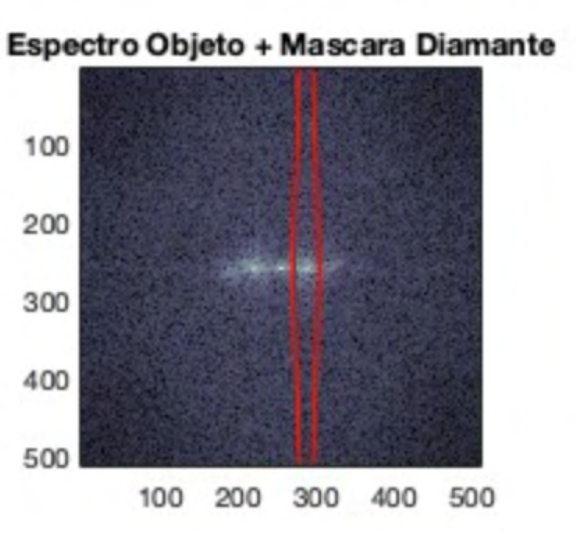


Figura 28: Máscara de diamante. Fuente propia

6.2.4. Transformada Inversa y Extracción de la fase envuelta

Ya que se filtro la frecuencia con las distintas geometrías de las mascaras como siguiente paso se aplico la transformada inversa de fourier para pasar del dominio de la frecuencia al dominio espacial. Lo que tiene como propositivo es sintetizar una señal analítica compleja a partir del lóbulo espectral el cual fue previamente aislado mediante un filtrado pasa banda logrando desacoplar el mapa de fase envuelta.

De manera analítica la transformada inversa de fourier bidimensional continua sobre el espectro filtrado $G_f(u, v)$ definiéndose mediante la integral doble como en la ecuación (49)[24].

$$gf(x, y) = \mathcal{F}^{-\infty}\{G_f(u, v)\} = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} G_f(u, v) exp[j2\pi(ux + vy)] dudv \qquad (49)$$

En el entorno en el que se implemento el método tradicional MATLAB, cuando se opera en matrices discreta que son de $N \times N$ pixeles, en este caso son de 512×512, se ejecuta el algoritmo de la transformada inversa de fourier bidimensional (2D-IFFT), cuya formulación discreta normalizada se expresa como la ecuación (50) [36][23].

$$\hat{g}(x, y) = \frac{1}{N^2} \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} G_f(u, v) exp\left[j2\pi\left(\frac{ux + vy}{N}\right)\right] \quad \text{para } x, y = 0, 1, \ldots, N-1 \quad (50)$$

Cuando se toma la IFFT de un espectro simétrico completo se obtiene como resultado espacial la imagen original de intensidades. Sin embargo, la perfilometria lo que hace es suprimir la componente continua así como el lóbulo conjugado simétrico,

reteniendo de manera asimétrica el armonio portador positivo, de manera analítica se presenta como la siguiente ecuación(51)[36][24].

$$G_f(u,v) = H(u,v) \cdot G(u,v) = C(u - f_0, v) \tag{51}$$

Se aplica la Transformada Rapida de Fourier Bidimensional sobre la componente unilateral aislada produciendo una señal bidimensional analítica compleja (52)[36][5].

$$\hat{g}(x,y) = \frac{1}{2} b(x,y) exp[j\phi(x,y)] \tag{52}$$

Esta señal bidimensional analítica compleja se descompone en el campo cartesiano tanto en la parte real como en la parte imaginaria en cada pixel, mediante la identidad de euler como se muestra las ecuación (53) de la cual su parte real se representa por la ecuación (54) y su parte imaginaria por la ecuación (55).

$$\hat{g}(x,y) = \frac{1}{2} b(x,y) cos[\phi(x,y)] + j\frac{1}{2} b(x,y) sin[\phi(x,y)] \tag{53}$$

$$\Re\{\hat{g}(x,y)\} = \frac{1}{2} b(x,y) \cos[\phi(x,y)] \tag{54}$$

$$\Im\{\hat{g}(x,y)\} = \frac{1}{2} b(x,y) \sin[\phi(x,y)] \tag{55}$$

La existencia de la parte de real y de la parte imaginaria ayuda que se pueda desacoplar de manera directa la modulación de la fase, para que esto suceda se aplica el operador arcotangente de los cuatro cuadrantes sobre el cociente entre ambas matrices espaciales, esto se representa de manera analítica como la siguiente ecuación (56).

$$\phi_w(x,y) = \arctan\left(\frac{\Im\{\hat{g}(x,y)\}}{\Re\{\hat{g}(x,y)\}}\right) = \arctan\left(\frac{\frac{1}{2} b(x,y) \sin[\phi(x,y)]}{\frac{1}{2} b(x,y) \cos[\phi(x,y)]}\right) \tag{56}$$

La función mostrada en la ecuación (56) tiene como propósito calcular el angulo a partir de razones trigonométricas, el resultado queda mapeado dentro de un rango modular restringido $\phi_w(x,y) \in (-\pi, \pi]$, esto se hace con el fin que cuando se origina el mapa de fase envuelto en el que las variaciones de relieve superen dicho rango angular mencionado anteriormente esto se manifiesta en forma de discontinuidades y saltos de 2π radianes [20].

6.2.5. Desenvolvimiento de Fase espacial (Phase Unwrapping)

Como ultimo paso de la reconstrucción mediante perfilometria de fourier es el desenvolvimiento de Fase Espacial o también conocido como Phase Unwrapping (PU), en esta parte del proceso se restaura la continuidad topográfica global y asi poder reconstruir la superficie continua $\Phi(x,y)$, para esto es necesario que determine el campo de saltos enteros $k(x,y) \in \mathbb{Z}$ con el fin de satisfacer la relación de compensación modular, en la ecuación (57) se muestra en su forma matemática [5]

$$\Phi(x, y) = \phi(x, y) + 2\pi k(x, y) \tag{57}$$

Para esta ultima etapa se implemento cinco algoritmos de desenvolvimiento de fase espacial al igual que se compararon los resultados en conjunto con las mascaras.

- **Itoh:** se implemento mediante la función estándar unwrap la cual fue aplicada en dos dimensiones, dicho método ayudo a que se evaluaran las diferencias finitas entre los pixeles adyacentes, en este caso si el gradiente local se pasa del umbral de π radianes, añade o sustraer múltiplos de 2π con el fin de restablecer la continuidad. Dicho algoritmo se destaca debido a su rapidez de computo y su bajo consumo de memoria pero tiene como limitación que es vulnerable al momento de la propagación de errores en especial cuando el objeto presenta sombras, discontinuidades o ruido

- **Herraez:** para este método se uso unwrap_phase, este se encarga de calcular un mapa de confiabilidad local es una conocida como vecindad de 3×3 pixeles. Basándose en estos valores los pixeles se ordenan por prioridad propagando el desdoblamiento a través de caminos que no son continuos los cuales van avanzando desde regiones que tienen mayor certidumbre hacia las zonas criticas que tienen un alto nivel de ruido con el fin de evitar que el error se propague.

- **Zhao:** este algoritmo se implemento con ayuda de la función Unwrap_TIE_DCT_Iter, aca se hace una combinación de la Ecuación de Transporte de Intensidad (TIE) con la Transformada Discreta de Coseno (DCT). En dicho esquema se integran las funciones de ponderación adaptativa y balanceo de residuos ópticos, esto resulta sumamente adecuado para la reconstrucción de geometrías que tienen presencia de ruido aleatorio así como discontinuidades suaves sin generar acumulaciones que sean severas en zonas que cuentan con alto gradiente.

- **Andreas:** para la ejecución de este método se utilizo un enfoque de optimización global por mínimos cuadrados con una regularización espacial de tipo Tikhonov. Su principal función minimizar la energía que penaliza las discrepancias entre los gradientes de la fase desenvuelta y los envueltos observados, este parámetro de regularización se ajustó de forma numérica con el fin de minimizar el error cuadrático medio ayudando a que la topografía sea visualmente homogénea y continua aunque tiene una cierta perdida de nitidez en los bordes finos.

- **Ghiglia & Romero / Ghiglia & Pritt:** su implementación fue mediante la función phase_unwrap, la cual es formulada a partir de la resolución numérica de la ecuación diferencial de Poisson ($\nabla^2\Phi = \rho$) en el sentido de mínimos cuadrados. Esto se calculó aplicando la Transformada Discreta del Coseno (DCT) bajo ciertas condiciones de frontera naturales de Neumann, asegurando que la fase desenvuelta sea suave, sin cortes artificiales y con una alta robustez frente a las perturbaciones locales.

6.3. Métricas

Para la evaluación del desempeño de la reconstrucción sera de manera cuantitativa con el propósito de garantizar que la confiabilidad y validez de los resultados de dicha investigación. Con el objetivo de poder hacer una comparación de manera objetiva la fidelidad de las reconstrucciones frente a su Ground Truth sintético, para poder lograr este objetivo se implementaron tres métricas las cuales son reconocidas en procesamiento y análisis de imágenes: IMMSE, PSNR y SSIM.

La métricas seleccionadas permitieron que se pudieran analizar de manera conjunta la precisión numérica puntual, la relación de potencia que tiene la señal frente al ruido residual así como la preservación de las características morfológicas y geométricas de la superficie evaluada frente a la referencia. Para esto en todos los experimentos que se hicieron se utilizo como referencia el Ground Truth (GT).

- **Error Cuadrático Medio en Imágenes (IMMSE):** esta métrica se encarga de cuantificar la magnitud de la discrepancia numérica promedio esta va de pixel a pixel entre la superficie de referencia $I(x, y)$ y la topografía reconstruida $K(x, y)$, en la ecuación (58) se muestra su formula en donde m y n hacen referencias a las dimensiones de la matriz analizada. Cuando se tienen valores mas bajos esto indica que se tiene un menor error y una mayor aproximación a la reconstrucción real.

$$IMMSE = \frac{1}{m \cdot n} \sum_{x=1}^{m} \sum_{y=1}^{n} [I(x, y) - K(x, y)]^2 \tag{58}$$

- **Relación Señal a Ruido Pico (PSNR):** Ayuda a evaluar la fidelidad de la superficie en decibeles (dB) relacionando la potencia máxima teórica admisible con el ruido cuadrático el cual es introducido durante la reconstrucción. Esta métrica se muestra en su forma matemática como en la ecuación (59). Cuando se tiene un valor mas elevado significa que se tiene una reconstrucción con menor ruido residual.

$$PSNR = 10 \cdot log_{10} \left(\frac{MAX^2}{IMMSE} \right) \tag{59}$$

- **Índice de Similitud Estructural (SSIM):** tiene como función modelar las distorsiones que son percibidas en la superficie comparando de manera simultanea la altura media, el contraste que es la desviación estándar del relieve y la correlación estructural de los bordes y curvaturas locales. En la ecuación (60) se presenta como se calcula esta métrica, cuando el resultado es cercano a 1 significa que se garantiza una alta conservación en la morfología y continuidad.

$$\text{SSIM}(I, K) = \frac{(2\mu_I \mu_K + C_1)(2\sigma_{IK} + C_2)}{(\mu_I^2 + \mu_K^2 + C_1)(\sigma_I^2 + \sigma_K^2 + C_2)} \tag{60}$$

6.4. Redes Neuronales convolucionales

El método tradicional de la perfilometria de fourier presentan limitaciones ante las discontinuidades geometricas abruptas, sombras, saltos de fase y ruido óptico. Para poder mitigar dichas limitaciones se integro el área de la inteligencia artificial en el área de aprendizaje profundo.

Para la realización de esta investigación se seleccionaron y modificaron tres arquitecturas: UHRNet con conexiones tipo U-Net, SFNet (fusión simétrica mediante mecanismos de atención cruzada) y Hformer (diseño híbrido basado en Vision Transformers con auto-atención).

Se seleccionó la arquitectura UHRNet (U-shaped High-Resolution Network) debido a que tiene la capacidad de mantener las rutas de procesamiento de alta resolución a lo largo de toda la red, esto es muy importante debido a que en la perfilometria, el submuestreo agresivo que tienen las redes clásicas provoca que se degraden los bordes finos de las franjas deprimiendo las discontinuidades morfológicas, es por eso que la elección de esta red solventa el intercambio de información de forma continua entre las representaciones finas y gruesas.

La segunda red que elegí para la realización de este proyecto fue la arquitectura SFNet (Symmetriic Fusion Network), se seleccionó esta red ayuda a aislar el desfase topográfico antes de proyectar la salida de fase absoluta, entonces una de las funciones de dicha arquitectura es abordar problemas en donde se requiera procesar de forma diferenciada la escena deformada y el plano de referencia esto lo hace mediante el uso de ramas siamesas simétricas y con fusión cruzada con atención.

Por ultimo se utilizo la Hformer (Hierarchical Transformer Network), esta fue incorporada para poder superar el campo receptivo local inherente a las capas convolucionales convencionales, existen ambigüedades en saltos de 2π en donde se requieren correlacionar zonas distantes del objetos, con ayuda de los módulos de atención espacial multiescala permite que se capture las dependencias globales de continuidad en la superficie.

6.4.1. Configuración de cada una de las redes usadas

A continuación, se describirá las configuración que tiene cada una de la redes utilizadas, en la figura 29 se muestra un cuadro con las especificaciones que cuenta cada una

Hiperparámetros	Arquitectura		
	UHRNet	SFNet	Hformer
Tipo de arquitectura	Ramas HR + LR paralelas	Encoder piramidal + FAM	CNN + Window Attention
Canales principales	hr_ch = 48, lr_ch=96	base=64	dims=(96,192,385)
Profundidad	n_stages = 4, n_blocks= 3	4 escalas, 2 conv/ etapa	depths= (2,2,6,2)
Kernel convolucional	3x3	3x3 y 7x7 (stem)	3x3 depth wise + 1x1 point wise
Tasa de aprendizaje	3E-04	1E-03	3E-04
Tipo de activación	ReLu	ReLu	GELU
Tamaño del lote	16 (A100)	16 (A100)	16(A100)
Optimizador	Adam	Adam	Adam W

Figura 29: Caracteristicas de cada una de las redes. Fuente propia

6.4.2. UHRNet (U-shaped High-Resolution Network)

La arquitectura UHRNet adopta una topología de codificador y decodificador modificada con el fin de poder conservar la resolución espacial. Esta cuenta con una entrada de un solo canal en donde va imagen de 512×512 la cual alimenta una rama de codificación estructurada en 4 niveles progresivos que van de 32, 64, 128 y 256 canales, cada uno de los niveles consta de operaciones convolucionales que van de 3×3 acopladas con normalización por lotes y activación ReLU como se muestra en la figura 30.

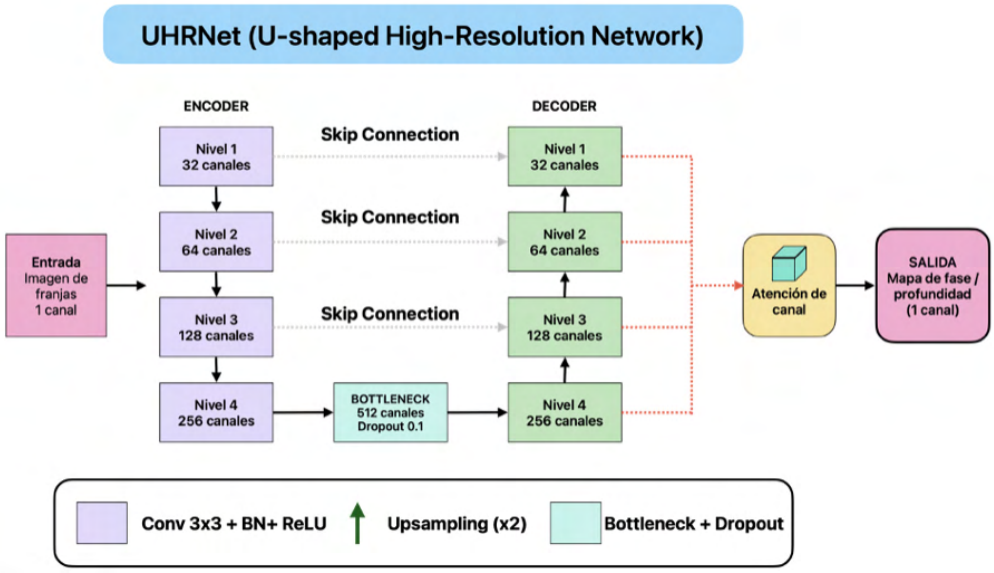


Figura 30: Arquitectura de la UHRNet modificada. Fuente propia

Como se puede observar en la figura 30 esta cuenta con un cuello de botella y con un decodificador, la base de la red el bottleneck es el encargado de procesar 512

canales en los cuales se incorpora una tasa de Dropout que es de 0.1 para poder mitigar el sobreajuste, por otro lado el funcionamiento del decodificador es recuperar la resolución espacial de manera gradual esto mediante capas de remuestreo ascendente (upsampling $\times 2$) estas van acopladas a los bloques convolucionales las cuales emparejan la profundidad de canales (256,128,64 y 32).

Las Skip Connections ayudan a evitar la perdida de alta frecuencia producida en las etapas de contracción, estas se integran de manera directa entre los niveles homólogos de codificación y decodificación. En la fase terminal, las características que tienen todos los niveles convergen en un modulo de atención de canal, el cual tiene la función de recalibrar la ponderación de los mapas de características antes de generar una salida continua en 1 solo canal.

6.4.3. SFNet (Symmetric fusion Network)

Dicha arquitectura esta diseñada en un enfoque siamés multiescala la cual esta diseñada para procesar de manera simultanea dos flujos de información complementarios, en la figura 31 se muestra como esta integrada esta arquitectura y a continuación se explicaran a detalle sus características.

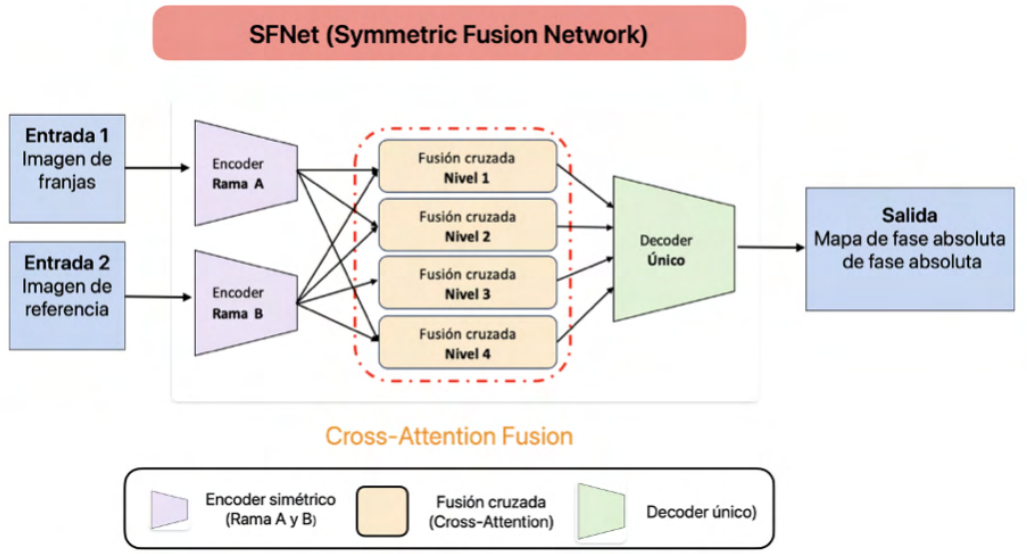


Figura 31: Arquitectura SFNet. Fuente propia

Dispone de dos ramas de codificación idénticas y paralelas a este tipo de codificación se le conoce como Dual Simétrica, en este caso la rama A tiene como función procesar la imagen del objeto con franja proyectadas y la rama B es la que recibe la captura del plano de referencia homogéneo. EL preprocesamiento de ambas señales se efectúa mediante una capa stem la cual tiene un gran campo receptivo de 7×7, seguida de convoluciones de 3×3 con una base inicial de 64 canales los cuales están distribuidos a lo largo de 4 escalas espaciales las cuales son 2 bloques convolucionales por etapa.

Cuenta con una fusión cruzada (Cross-Attention Fusion) en donde ingresan las salidas de cada nivel de las ramas A y B, en lugar de realizar una concatenación simple como se muestra en la figura 31. Dicho mecanismo ayuda a calcular los mapas de covarianza cruzada entre la referencia y la deformación, permitiendo que la red se desacople de manera activa la portadora espacial base y así pueda asignar mayor peso a las variaciones de fase inducidas por la topografía. Por otro lado, la representaciones que fueron enriquecidas por la atención cruzada convergen hacia un único decodificador piramidal mejor conocido como Feature Alignment Module (FAM) el cual puede reconstruir de manera progresiva la geometría del objeto produciendo como salida un mapa continuo de fase absoluta desenvuelta.

6.4.4. Hformer (Hierarchical Vision Tranformer)

La Hformer representa un paradigma hibrido el cual acopla la capacidad de extracción local de las redes usando el modelado de relaciones contextuales que son de largo alcance de los Tranformers. En la figura 32 se presenta un diagrama en donde muestra como es la arquitectura de esta red.

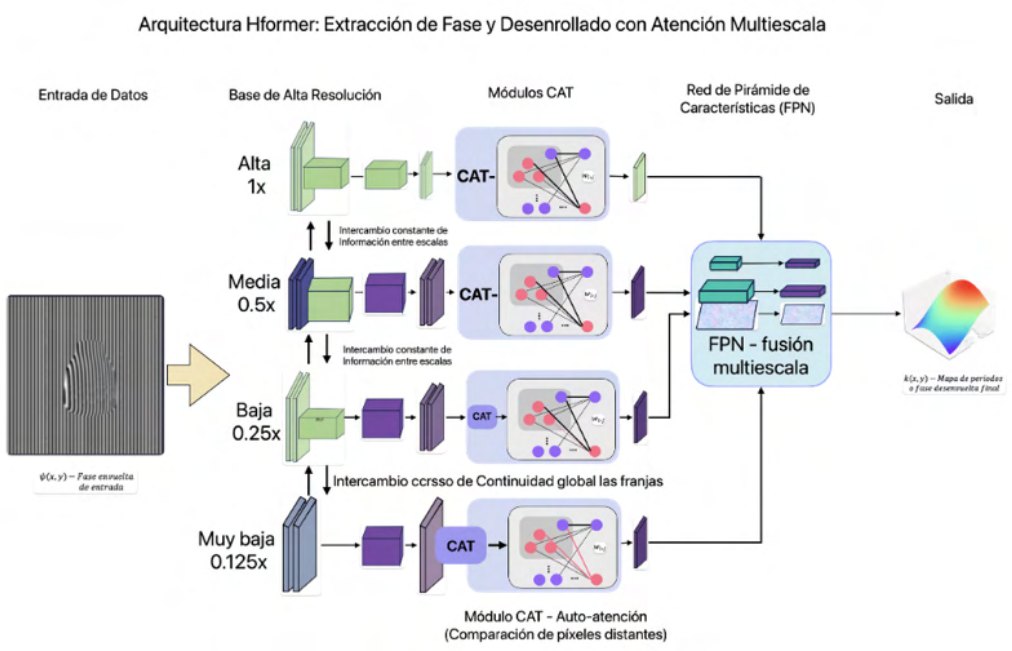


Figura 32: Arquitectura Hformer. Fuente propia

Tiene una estructura jerárquica en donde la red recibe el objeto como entrada y lo subdivide en cuatro jerarquías piramidales de resolución que va de manera descendente: Alta (1×), media (0,5×), baja (0,25×) y muy baja (0,125×) con dimensiones de canal proyectadas a 96, 192 y 385 dimensiones espaciales. A lo largo de estar ramas se mantiene un intercambio que va de manera constante de información bidireccional entre las

escalas con el fin de asegurar la coherencia contextual. En la parte del procesamiento se utiliza convoluciones que son separables en profundidad (3×3 Depthwise Convolutions) estas van seguidas de proyecciones puntuales (1×1 Pointwise Convolutions) los cuales ayudan a optimizar el uso de los parametros. En cada nivel se incorporan los módulos CAT (Cross-scale Auto-attention) estos aplican una atención dentro de las ventanas locales desplazadas o también conocidas como window attention permitiendo que los pixeles que estan geometricamente distantes dentro del patrón de franjas puedan comparar su estado de la fase para poder determinar la existencia de saltos de $2k\pi$. Para esta arquitectura se uso una activación GELU (Gaussian Linear Unit) ya que esta aporta transiciones mas suaves y diferenciables en el cálculo del gradiente en compraci´ón con la activación ReLU. Los mapas que son procesados por los módulos CAT se integran mediante una Red de Pirámide de Características (FPN) que tiene como función sintetizar la información multirresolución para poder producir el mapa de períodos compensados $k(x, y)$ y la fase continua final desenvuelta.

6.5. Entorno experimental y protocolo de entrenamiento

Se implementaron en un entorno computacional de alto rendimiento el cual fue estructurado sobre el framework PyTorch. En la etapa de optimización y retropropagación de gradientes se requirió una aceleración por hardware la cual esta dedicada para manejar una alta dimensional de las matrices de entrada así como de los volúmenes de características multiescala.

El entrenamiento de los modelos se realizo en Colab en un nodo acelerado por hardware con una GPU NVIDIA A100, para cada ciclo del entrenamiento proporcionando el ancho de banda así como la capacidad de memoria de video (VRAM) la cual es de 42 GB, los tensores de entrada y las etiquetas de referencia que es el Ground Truth mediante las llamadas directas al dispositivo a través de CUDA. Se hizo así con el fin de evitar los cuellos de botella que suceden por saturación en la memoria durante las operaciones intensivas de atención en ventanas de Hformer así como en las ramas paralelas de alta resolución UHRNet. Con base a que se uso una GPU A100 se fijo un Batch Size de 16 para las tres redes, dicho valor garantizo una ocupación balanceada de los Tensor Cores ayudando a que se mantenga una estabilidad numérica en el calculo de los gradientes y así evitar la sobrecarga en la memoria dinámica del entorno de Colab.

La base de datos sintética cuenta con 1563 objetos de los cuales 1250 se ocuparon para el conjunto de entrenamiento lo que equivale al 80 % que esta destinado al ajuste iterativo de los pesos así como para los sesgos de cada red mediante la retropropagación de errores. Por otro lado, se ocuparon 156 objetos para el conjunto de validación siendo el 10 % de la base de datos completa este es evaluado de manera periódica al final de cada época con el fin de poder monitorizar la generalización sobre las superficies no vistas durante el ajuste, supervisar la evolución de las pérdidas y también registrar los puntos de control que cuentan con menor error. Por ultimo se usaron 157 muestras para el conjunto de prueba equivalente al 10 % en donde se aísla completamente el proceso de optimización el cual se reserva para el paso final al momento que se hace la evaluación

de las métricas de la reconstrucción 3D.

Los tensores se cargaron mediante módulos DataLoader optimizados, para poder garantizar que el flujo de datos no hiciera mas lento el ciclo de cómputo de Colab, a esto también se le aplico una reorganización conocida como shuffle al inicio de cada época. Esta mezcla rompió correlaciones secuenciales entre las vistas consecuencias de un mismo objeto forzando que las redes puedan aprender representaciones geométricas que son invariantes a la orientación espacial.

6.5.1. Optimización Paramétrica

Para hacer el ajuste de los parámetros entrenables para cada arquitectura esto se realizo mediante algoritmos de optimización los cuales fueron basados en estimación adaptativa de momentos. Se utilizo el optimizador Adam para poder calcular tasas de aprendizaje independientes para cada uno de los parámetros combinando momentos que son de primero y segundo orden.

Para la arquitectura UHRNet se configuró con una tasa de aprendizaje inicial (η) de 3×10^{-4} al elegir esta tasa tuvo como ventaja que favoreció la convergencia estable a lo largo de sus cuatro etapas paralelas, evitando que existan oscilaciones en los mapas que tienen alta resolución. Por otra parte para la red SFNet se calibró con una tasa de aprendizaje de 1×10^{-3} facilitando en la alineación de características así como la sincronización entre el codificados siames dual y los módulos de atención cruzada.

Para la red Hformer se ocupo un optimizador AdamW con una tasa de aprendizaje de 3×10^{-4}, se uso este tipo de optimizador debido a que este desacopla el decaimiento de los pesos del calculo del gradiente directo esto ayuda a impedir la sobreponderación paramétrica que existen en las proyecciones lineales de los módulos de auto-atención CAT (Croos-scale Auto-Attention).

Se realizo un ciclo de retropropagación en donde cada itereación se ejecuto de forma secuencial desde el reinicio de gradientes el cual se hizo mediante la función optimizer.zero_grad(), por segundo se aplico un pase hacia adelante mejor conocido como el forward pass, por otra parte se realizo la evaluación de la función de costo frente al Ground Truth, otro aspecto es la derivación hacia atrás (loss.backward()) y por ultimo se realizo la actualización de los pesos sinápticos (optimizer.step())

6.5.2. Supervisión y función de perdida

La supervisión de las arquitectura juega un papel importante para realizar esto se llevo a cabo utilizando como referencia el mapa continuo de elevación en donde se normalizo en el rango que va de $[0, 1]$. Se hizo esto debido a que la reconstrucción de la superficies tridimensionales la cuales requieren preservar de manera simultanea las curvaturas suaves así como los cambios abruptos de cota como lo son los bordes y las transiciones de contorno, para esto se implemento una función de pérdida multiobjetivo balanceada $\mathcal{L}_{total}$, esta función se representa por la siguiente ecuación (61), donde la pérdida $\mathcal{L}_1$ representa al Error Absoluto Medio (MAE).

$$\mathcal{L}_{total} = \alpha\mathcal{L}_1 + \beta\mathcal{L}_{MSE} \tag{61}$$

La pérdida $\mathcal{L}_1$ se calcula como se muestra en la ecuación (62), el cual penaliza las discrepancias morfológicas de manera lineal sin sobredimensionar los valores atípicos, lo que ayudo a mantener la nitidez en tanto aristas como regiones que cuentan con pendientes pronunciadas.

$$\mathcal{L}_1 = \frac{1}{M}\sum_{i=1}^{M}\left|\Phi_{pred}^{(i)} - \Phi_{GT}^{(i)}\right| \tag{62}$$

Para la perdida del Error Cuadrático Medio (MSE) se aplica una penalización cuadrática sobre las diferencias globales pronunciadas, acelerando el descenso del gradiente en las etapas tempranas de entrenamiento garantizando la suavidad de las zonas continuas de la geometría. En la ecuación (63) se presenta en su forma matemática.

$$\mathcal{L}_{MSE} = \frac{1}{M}\sum_{i=1}^{M}\left(\Phi_{pred}^{(i)} - \Phi_{GT}^{(i)}\right)^2 \tag{63}$$

Estos coeficientes de ponderación se establecieron de forma equitativa ($\alpha = 0{,}5$) y ($\beta = 0{,}5$) esto nos permitió que el entrenamiento en la GPU NVIDIA A100 dentro de Google Colab, esto se hizo con el fin de que pudiera alcanzar un balance óptimo entre la precisión métrica puntual, estabilidad estructural y convergencia global en la reconstrucción tridimensional.

VII. Resultados y discusión

En este capitulo se presentan los resultados obtenidos durante la investigación en donde se hace un análisis integral y la discusión de estos resultados, a lo largo de esta investigación se realizaron distintos procesos desde la creación de la base de datos hasta la aplicación de la redes neuronales para la reducción del error.

El resultado que se obtuvo de la reconstrucción 3D a partir de patrones de franja continuas sobre un conjunto de datos sintéticos, al obtener estos resultados se evaluó el desempeño del método tradicional en la perfilometria de fourier por transformada de fourier bidimensional (FTP2D) combinándola con cinco algoritmos de desenvolvimiento de fase, posteriormente se examino el proceso de convergencia, optimización y capacidad inferencial de tres tipos de redes neuronales convolucionales de aprendizaje profundo supervisado directo (fringe-to-depth): se uso una red con ramas multirresolución en paralelo UHRNet, como segunda red fue la SFNet la cual trabaja con un codificador piramidal simétrico y por ultimo fue la red Hformer es un modelo híbrido el cual esta basado en transformadores de ventana, ya que se entrenaron estas redes por ultimo paso se hizo una comparación cuantitativa con ayuda de las métricas usadas así como en la fidelidad morfológica superficial en representaciones 2D y las malla volumétrica 3D tomando como referencia para dicha comparación el Ground Truth, también se hizo una comparación de los resultados con el uso de la redes neuronales convolucionales y el método tradicional.

7.1. Base de datos sintética

La base de datos esta conformada por 1632 objetos de los cuales cada uno cuenta con la imagen del objeto con el patrón de franjas proyectado el cual codifica las deformaciones las cuales van asociadas a la topografía también cada uno tiene su Ground Trurh (GT-SP) como se muestra en la figura 33.

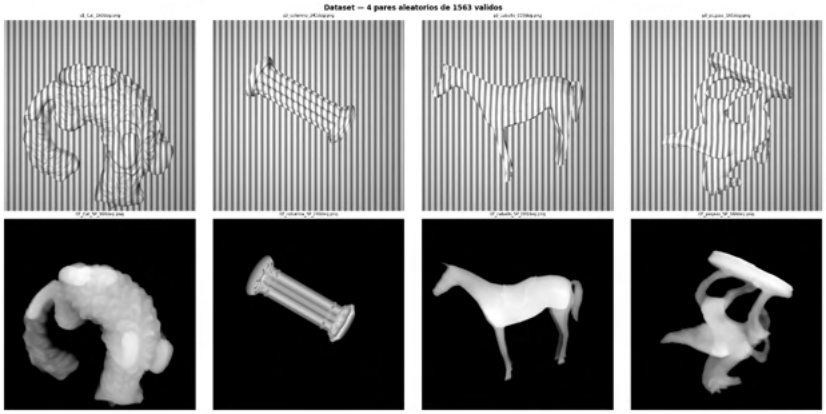


Figura 33: Base de datos sintética en el entrenamiento. Fuente propia

Se cargaron muestras en lotes de 16 pares generando una señal de gradiente balanceada, estas muestras fueron de suma importancia para la parte de las curvas de entrenamiento ya que se obtuvo una reducción continua el error sin divergencia numérica, sin embargo se indujeron pequeñas oscilaciones que tienen una alta frecuencia las cuales fueron acotadas en las fases de validación, estas se notaron entre las épocas 10 y 80 en las arquitecturas SFNet y Hformer. Para poder mitigar el sobreajuste se aplicó una reorganización aleatoria de los 1632 en donde en cada época se aseguró que ninguna red pudiera memorizar la correlación geométrica obteniendo como resultado que la perdida de entrenamiento y la de validación pudieron converger hacia un valor asintótico por debajo de 0.015 sin mostrar un ensanchamiento en la brecha los cuales nos llevaron a confirmar que las representaciones aprendidas corresponden a invariantes óptico-geométricas de la portadora.

Por otro lado, al hacer un contraste directo de la predicción de la red con el mapa GT en cada uno de los pases los modelos fueron aprendiendo de forma progresiva a suprimir el fondo sin las franjas así como a concentrar la densidad de computo que va exclusivo en la silueta volumétrica del objeto. Esto provoco que que en la métrica SSIM de validación evolucionara de manera rápida alcanzando una zona de saturación superior $SSIM > 0{,}98$ antes de poder completar el 50 % de las épocas programadas en las arquitecturas UHRNet y Hformer el cual supera la limitación que tienen los algoritmos de desenvolvimiento de fase los cuales colapsan ante los bordes que no tienen continuidad física.

7.2. Evaluación del Desempeño del Método Tradicional

El procedimiento de la reconstrucción se ejecuto sobre los objetos sintéticos en donde se aplico una descomposicion en el dominio de la frecuencia espacial haciendo uso de la Transformada de fourier bidimensional, esta parte del proceso ayudo a descomponer el patrón de la entrada en su componente continua (DC) la cual esta posicionada en el centro del espectro y los dos lóbulos armónico laterales correspondientes a las portadoras fundamentales $\pm f_0$, esto se hizo con el propósito de poder aislar la componente útil descartan el ruido de fondo para esto se empleo una máscara espectral que tiene forma de diamante, esta mascara se centro verticalmente en el lóbulo $+f_0$. Después se aplico la Transformada Inversa de Fourier (2D-IFFT) con esto se extrajo el mapa analítico complejo con esto también se calculo la distribución de la fase envuelta modulada $\phi(x, y)$ la cual fue acotada en el intervalo de $(-\pi, \pi]$ con ayuda de la función arcotangente de cuatro cuadrantes. En la figura 34 se muestra el método tradicional en donde se usaron distintos objeto sintéticos, en esta misma se muestra el objeto con franjas junto con su plano de referencia de igual manera se observa aplicación de la mascara de diamante la cual abarca toda la frecuencia de pico a pico sin dejar entrar perturbaciones.

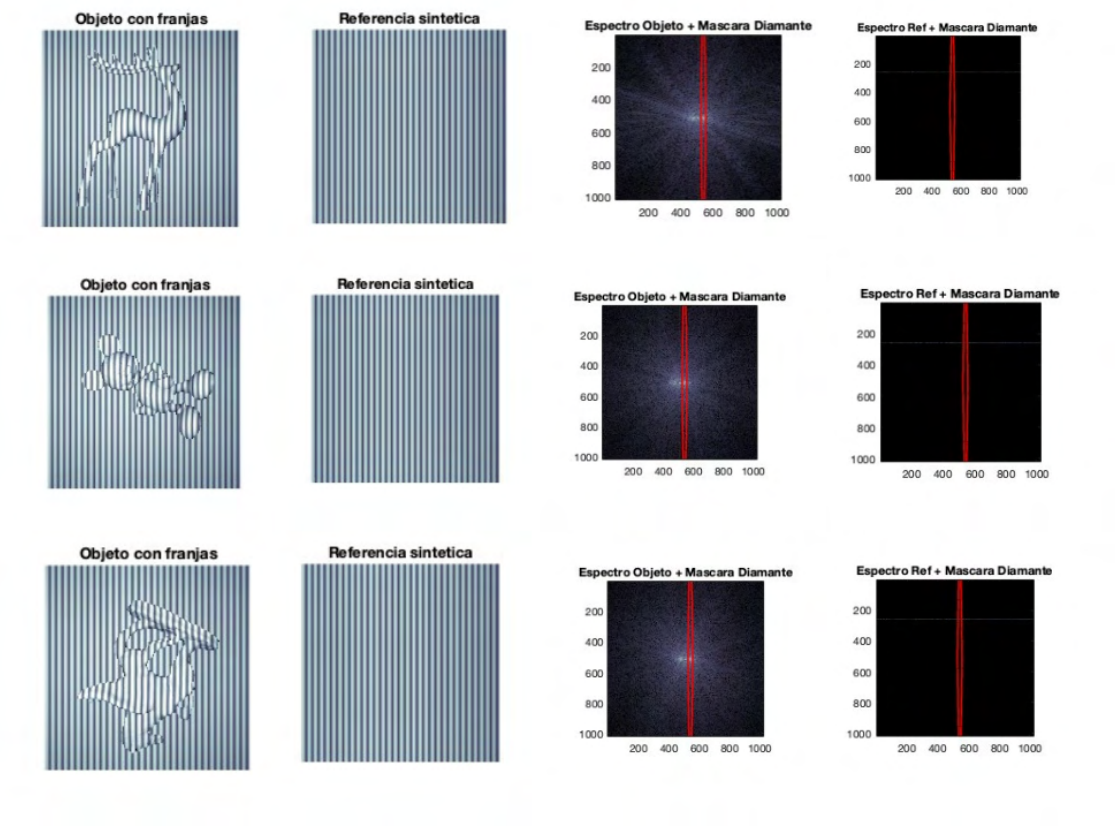


Figura 34: Objeto con franjas y mascara de diamante. Fuente propia

Para poder restablecer la continuidad topográfica de la fase y así poder remover los saltos artificiales que existen de 2π se implementaron y evaluaron cinco algoritmos de desenvolvimiento de fase Itoh, Herraez, Zhao, Andreas y Ghiglia, para poder tener un análisis cuantitativo se obtuvieron los resultados los cuales se muestran en la tabla (1), estas métricas se calcularon aplicando un protocolo de estandarización de normalización que va en el rango de $[0, 1]$ los cuales son delimitados por la máscara binaria del objeto también se optimizo el factor de compensación de la nivelación global Δ

Con resultados mostrados anteriormente se puede observar que los algoritmos de Itoh, Herraez, Zhao y Andreas alcanzaron los mismos valores de error y correlación, obtuvieron un $IMMSE = 0{,}215$, $PSNR \approx 6{,}66dB$ y $SSIM \approx 0{,}0367$ esto nos demuestra que los algoritmos secuenciales de integración fallan de manera sustancial antes las discontinuidades abruptas las cuales separan la silueta del objeto con respecto al plano de fondo homogéneo. Para el caso de Itoh hubo una acumulación errónea e 2π a lo largo de las lineas de rastreo lo que tuvo como consecuencia que se crearan fracturas las cuales se desfasaron verticalmente.

Por otro lado, el algoritmo de Herraez funciona por el critrerio de seguimiento guiado

Máscara Espectral	Algoritmo PU	IMMSE ↓	PSNR (dB) ↑
Itoh	0.21578	6.6599	
Herraez	0.21578	6.6599	
Zhao	0.21578	6.6599	
Andreas	0.21578	6.6599	
Ghiglia	0.16812	7.7438	

Cuadro 1: Comparación de algoritmos de Phase Unwrapping (PU) con Máscara Diamante. Fuente propia

por calidad sin embargo este no logra que se pudiera resolver el salto perimetral lo cual creo un plano inclinado o rampa de fase masiva la cual eleva la superficie de manera artificial por encima de los 60 a 100 radianes provocando que se suprimiera cualquier relieve interno útil como se muestra en la segunda gráfica de la figura 35. La propagación de residuos en los limites del objeto limita de manera parcial la base horizontal lo cual indujo a deformaciones severas que tienen forma de espigas verticales en los que son las fronteras perimetrales.

El algoritmo que tuvo mejores resultados fue el de por mínimos cuadrados el cual se conoce como Ghiglia, su funcionamiento principal el formular el problema de desenvolvimiento como la ecuación de poisson discreta la cual esta sujeta a ciertas condiciones de frontera de Neumann la cual se resuelve mediante la trasformada discreta del coseno (DCT), para esto Ghiglia evita que exista una propagación direccional de errores de integración haciendo que este algoritmo se consolide con un mejor rendimiento cuantitativo con respecto a los otros algoritmos teniendo como resultados en las métricas los siguientes: $IMMSE = 0{,}16812, PSNR = 7{,}7438dB, SSIM = 0{,}0388858$. Sin embargo a pesar de que obtuvo buenos resultados en comparación con los demás este produjo un suavizado excesivo en las pendientes que están pronunciadas impidiendo que se tenga una buena reconstrucción aplanando las cotas de profundidad máxima ante las discontinuidades aisladas.

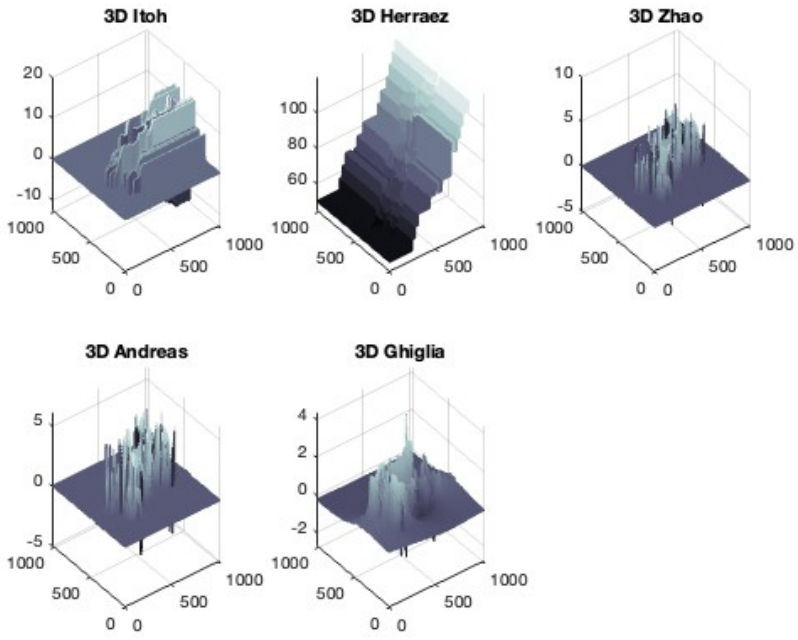


Figura 35: Reconstrucción 3D del objetos sintético con distinto algoritmos de desenvolvimiento de fase. Fuente propia

También se realizo el mismo procedimiento con objetos reales los cuales nos dieron mejores resultados sin embargo en comparación con el experimento con objeto sintéticos Ghiglia nos dio una mejor reconstrucción sin dejar huecos como se muestra en la figura 36, el algoritmo que no dio buenos resultados fue el de Itoh dejando entrar varios errores en la reconstrucción.

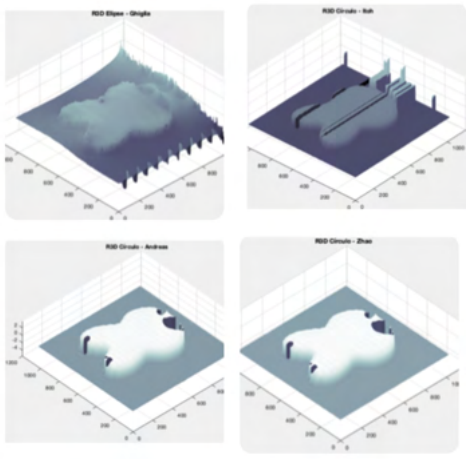


Figura 36: Reconstrucción 3D de objetos reales con distinto algoritmos de desenvolvimiento de fase. Fuente propia

7.3. Entrenamiento, Dinámica de Pérdida y Convergencia de las Redes Neuronales

El proceso tradicional como lo mencione anteriormente cuenta con limitaciones en el filtrado de las frecuencias así como el desenvolvimiento secuencial, para poder terminar con estas limitaciones se implementaron métodos de aprendizaje automático lo que son las Redes Neuronales Convolucionales mediante un paradigma de fringe-to-depth en donde se entrenaron tres redes profundas: UHRNet, SFNet y Hformer. El entrenamiento se llevo a cabo de manera supervisada en Google Colab utilizando un clúster acelerado por una tarjeta gráfica GPU NVIDIA A100 con un tamaño de lote de 16 pares con una base de datos sintética con 1563 pares de los cuales son el objeto con franjas y su Ground Truth correspondientemente.

En la siguiente tabla (2) se muestran los resultados generales cuantitativos que se obtuvieron en el entrenamiento, haciéndose una comparación entre los resultados de cada métrica de cada red. A continuación se presentaras los resultados que tuvo cada red.

Métricas	Arquitectura		
	UHRNet	SFNet	Former
IMMSE	0.000224	0.001102	0.000609
PSNR (dB)	37.37 dB	30.62 dB	33.05 dB
SSIM	0.9984	0.9917	0.9954

Cuadro 2: Comparación de métricas según la arquitectura. Fuente propia

7.3.1. Resultados UHRNet

La arquitectura UHRNet (U-shaped High-Resolution Network) se usaron 200 epocas en el entrenamiento mediante un optimizador Adam ($LR = 3 \times 10^{-4}$) con una activación ReLU, su diseño es muy importante para la preservación continua de las representaciones que tienen alta resolución en paralelo a lo largo de las cuatro etapa las cuales estan interconectadas mediante los bloques de intercambio de multirresolución. A continuación en la figura 37 se muestra la evolución del desempeño del modelo a lo largo de las 200 épocas del entrenamiento, en donde se evalua de manera simultanea la función de pérdida así como las tres métricas.

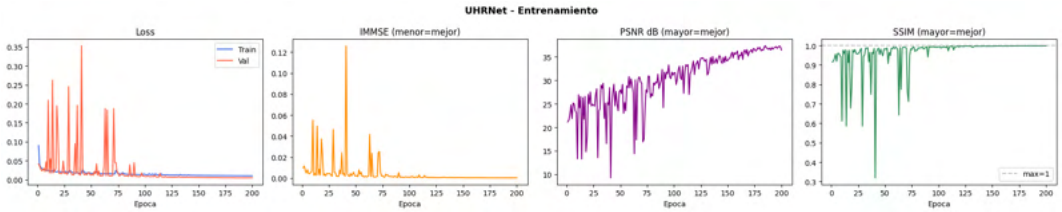


Figura 37: Curva de entrenamiento de la red UHRNet. Fuente propia

En la primera gráfica la cual es la figura 38 se muestra la función de perdida la parte del entrenamiento la cual es la linea azul demuestra un decaimiento continuo desde la época 1, la cual se fue estabilizando de forma asintótica por debajo de 0.01. Por otro lado, la perdida va descendiendo con una tendencia global sin embargo presenta fluctuaciones y picos pronunciados entre las épocas 10 y 75 el cual alcanzo un máximo cercano a 0.35 en la época 40 acoplándose suavemente a la curva de entrenamiento a partir de la época 80.

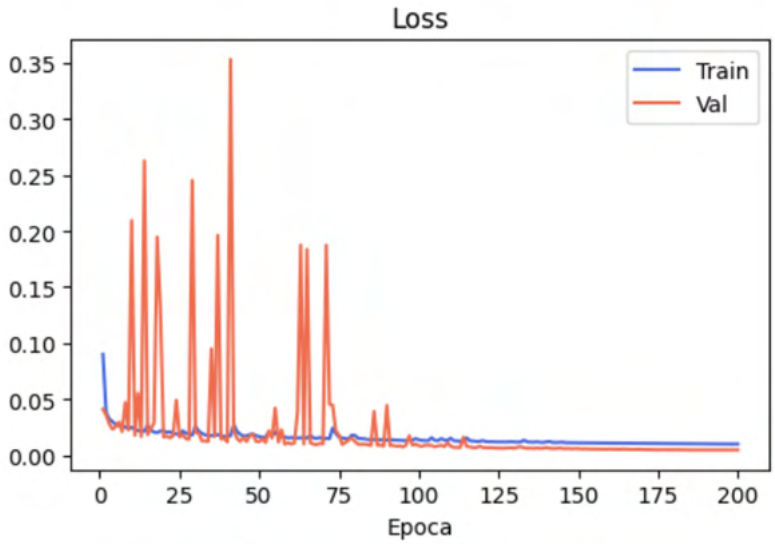


Figura 38: Función de perdida de la red UHRNet. Fuente propia

De igual manera se pueden observar oscilaciones intermedias en la validación, estas son características del optimizador cuando se procesan geometrias las cuales tienen discontinuidades abruptas y oclusiones complejas este optimizador experimenta ajustes dinámicos de gradiente esto sucede al procesar lotes que tienen alta variación morfológica. La convergencia que tienen ambas hacia un mínimo común estable ayuda a descartar la presencia de sobreajuste (overfitting) garantizando que el modelo generaliza de manera adecuada ante los datos no vistos.

También se presenta el comportamiento de la curva de las métricas en la figura 39, en la primera gráfica se puede observar la curva del error cuadrático medio en imágenes

(IMMSE) la cual presenta picos durante las primeras 75 épocas con una dispersión máxima de ≈ 0,125 en la época 40 entrando en una fase de decaimiento a partir de la época 85. Dicha métrica lo que hace es penalizar de forma cuadrática las diferencias de elevación pixel a pixel entre lo que es la inferencia y el mapa de referencia (GT), la eliminación total de las oscilaciones residuales en el ultimo 60 % del ciclo ayuda a confirmar que la red logró minimizar los errores de fase de manera uniforme en toda la escena.

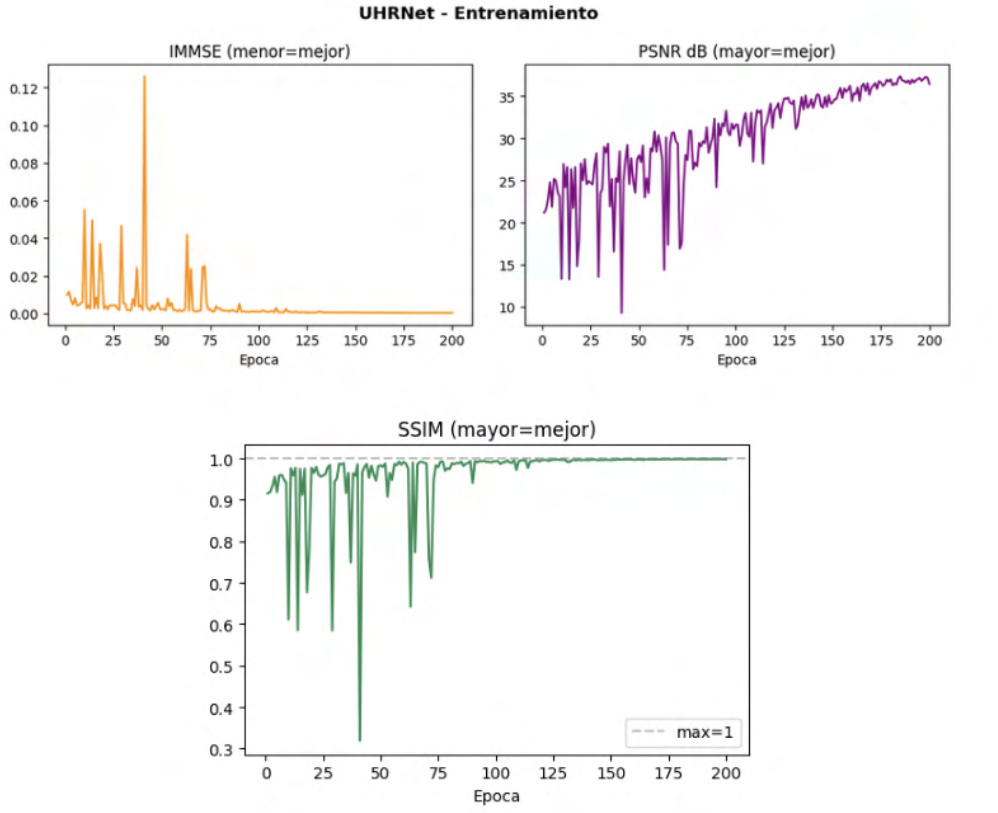


Figura 39: Métricas de validación durante el entrenamiento. Fuente propia

En la figura anterior en la tercera gráfica representa la relación señal a ruido pico (PSNR) entre mayores dB de es mejor, en este caso presenta una trayectoria predominante la cual va ascendiendo a lo largo de todo el entrenamiento, inicialmente presenta caídas transitorias iniciales las cuales coinciden temporalmente con las perturbaciones que tiene la función de costo, las métricas van ascendiendo de manera sostenida desde las cotas de 15-20 dB hasta que supera los 37dB al irse aproximando a la época 20. En el contexto de la metrología óptica y visión por computadora los valores de PSNR que son mayores a 30-35 dB representan un reconstrucción que tiene alta fidelidad con la mínima presencia de ruido. En este caso el incremento neto de 20 dB a lo largo del proceso valida que el mecanismo de intercambio de características preserva la amplitud

verdadera de la topografía.

Por ultimo en la gráfica de abajo se presenta el índice de similitud estructural (SSIM) en la época 40 se pueden observar caídas bruscas de $\sim$ 0,32, esta métrica experimento una recuperación rápida. A partir de la época 80 entro en una zona de saturación superior por encima de 0.98, la cual alcanzo una cota prácticamente unitaria de$>$ 0,998 hacia el final del ciclo, a diferencia del error cuadrático, el SSIM esta evaluando la preservación estructural considerando contraste y coherencia espacial en los bordes. Cuando se tuvo una estabilización cercana a a 1.0 ayudo a comprobar que la UHRNet puede preservar la integridad de los contornos finos y concavidades complejas ayudando a superar el suavizado excesivo el cual es característico en el método tradicional.

7.3.1.1. Interpretación de la métricas

A continuación se presentan los resultados de las métricas en la tabla (3), las cuales se explicaran en este bloque:

Métrica	**Valor**
IMMSE	0.000224
PSNR (dB)	37.375070
SSIM	0.998373

Cuadro 3: Resultados del conjunto de prueba (Test) para UHRNet.

En la métrica del IMMSE dio un valor de 0.000224 situándose en el orden de 10^{-4} esto confirma que la discrepancia cuadrática que va de pixel a pixel entre la superficie predicha y el Ground Truth es prácticamente despreciable, esto significa que la red aprendió a desacoplar la componente de la portadora sin propagar los desajustes de cota en todo el objeto.

La relación señal a ruido pico (PSNR) fue de 37.375070 dB, en la tarea de perfilometria cuando se tiene un valor arriba del 35dB significa que la reconstrucción tiene una alta fidelidad representando la supresión casi total de ruido el cual tiene alta frecuencia en la topografia estimada.

Por ultimo, tuvo 0.998373 en la métrica del índice de similitud estructural, el limite teórico de esta métrica es de 1, demostrando que el modelo preserva prácticamente el 100 % de la integridad estructural de objetos esto quiere decir que no deforma contornos, respeta pendientes naturales y conserva las relaciones espaciales entre las curvaturas continuas y zonas de fondo plano.

7.3.1.2. Análisis de las muestras de prueba

En el entrenamiento de estas redes fue con muestras aleatorias en este caso se presentaran tres, en la figura 40 se muestran los resultados en donde se observa en la fila

superior el procesamiento bidimensional y cuantitativo del error, mientras que en la fila inferior se evalúa la reconstrucción tridimensional.

La fila superior esta constituida por:

- Objeto con franjas
- Ground Truth
- La predicción de la UHRNet
- Error absoluto

Por otro lado en la fila inferior muestra las topografías en 3D:

- Superficie 3D GT vs UHRNet
- Superposición: el azul hace referencia al GT y el rojo a la predicción

En la muestra uno se presenta mickey el cual cuenta con curvaturas especificas y oclusiones de unión, con esta muestra se busco que la red evalué la capacidad para resolver problemas con cuerpos esféricos y cóncavos continuos, en la mallas tridimensionales la red pudo alcanzar las cotas máximas de elevación $z \approx 0{,}8$ en la cabeza y las extremidades sin sufrir ningún aplanamiento que normalmente aplica el método tradicional, en el mapa de error absoluto se puede observar que el fondo esta completamente negro lo que equivale a que no introdujo errores sin embargo se concentro en los pequeños desajustes residuales los cuales son los tonos amarillos y naranjas estos solo están en los tipos valles por ejemplo el cuello, orejas y rostro en donde la sombra proyectada de las franjas genera una ambigüedad local en lo que es la frecuencia espacial. Los valores que dio en la metricas locales fueron: IMMSE=0.000345, PSNR=34.6266 dB, SSIM=0.9956.

En la segunda muestra la cual es un dinosaurio este objeto presenta un cuerpo alargado con estructuras finas, los resultados de las métricas fueron las siguientes IMMSE=$8{,}30 \times 10^{-5}$, PSNR=40.8157 dB y un SSIM = 0.9976. Esta es una muestra la cual tiene un desempeño mas sobresaliente ya que cuenta con un PSNR mayor a 40 dB, su estructura delgada en la parte de la cola y las patas representan un desafió para el método tradicional, sin embargo la UHRNet preservo el espesor de la cola y en la parte de la malla roja se acopla de manera idéntica sobre la azul sin causar desfases verticales.

En la ultima muestra que es el pegaso es complejo debido a que tiene discontinuidades abruptas, paredes verticales y huecos, como por ejemplo en las alas y el pedestal contiene cavidades u orificios vacíos, en el método clásico cuando existe este tipo de limites colapsa generando errores de 2π, sin embargo la red delimita con nitidez los huecos que tiene el pedestal asignándole como cota cero sin desbordar el relieve falso manteniendo su fidelidad estructural de 0.9975. La métricas para esta muestra fueron: IMMSE = 0.000234, PSNR = 36.3145 dB y SSIM=0.9975

Estos resultados cuantitativos ayudaron a corroborar como esta arquitectura es robusta y funciona para resolver este problema, el uso de las ramas de alta resolución

en paralelo a lo largo las cuatro etapas de la red previene la perdida de las frecuencias espaciales altas traduciéndose como una delimitación fidedigna de las aristas verticales, estructuras final sin requerir etapas intermedias del filtrado.

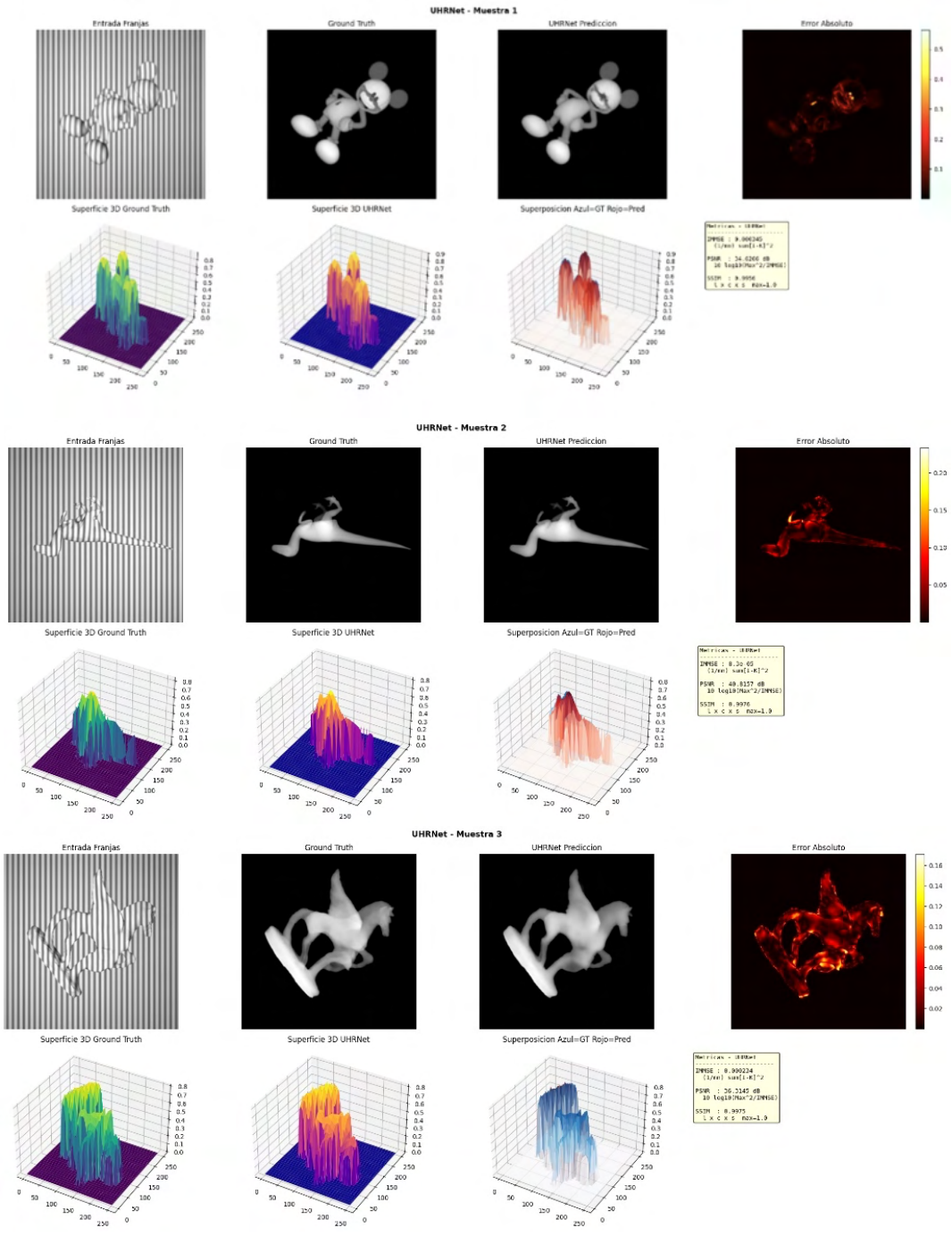


Figura 40: Muestras de prueba. Fuente propia

7.3.2. Resultados SFNet

A continuación se mostraran los resultados de la arquitectura SFNet la cual se entreno con 150 épocas con el optimizados Adam ($LR = 1 \times 10^{-3}$) y con una función de activación ReLU, esta arquitectura dispone de un codificador piramidal simétrico a cuatro escalas con un modulo de agregación de características (FAM).

En la figura 41 se presenta la evolución del desempeño del modelo en las 150 épocas de entrenamiento, se muestra la función de perdida y las métricas utilizadas.

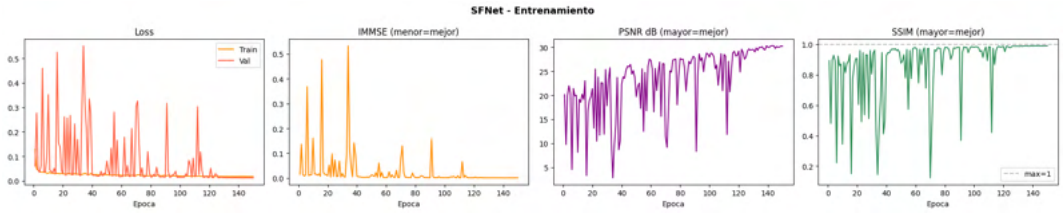


Figura 41: Entrenamiento de la red SFNet. Fuente propia

En la figura 42 se muestra la función de perdida en el conjunto de entrenamiento esta presenta un descenso continuo desde el principio pero se logro estabilizar por debajo de 0.02, por otro lado en la parte de validación se pueden observar fluctuaciones severas con picos que son notablemente durantes las primeras 115 épocas con apmplitudes que sobrepasan 0.5 en las épocas 15, 35 y 70. Dicho comportamiento demuestra una dificultad inicial del módulo de agregación de características piramidales (FAM) para poder generalizar sobre las geometrías con discontinuidades abruptas, pero a partir de la época 120 las oscilaciones desaparecen logrando que se acoplen las dos curvas de manera estable cercano a 0.015 confirmando la ausencia de sobreajuste.

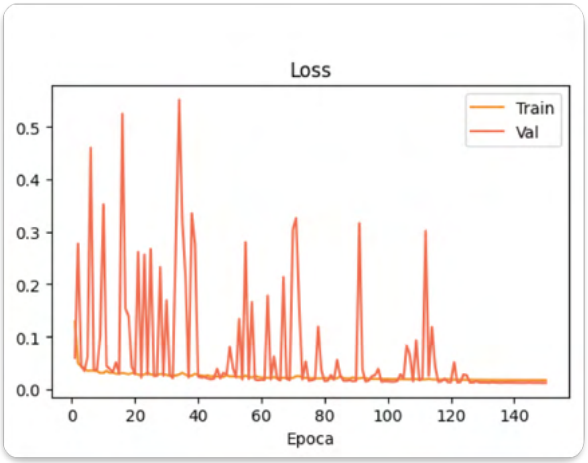


Figura 42: Función de perdida SFNet. Fuente propia

Para la figura 43 se presentan las curvas de entrenamiento de la parte de las métricas. Para el IMMSE se puede observar que se reproduce fielmente a las perturbaciones de la

perdida de validación de las cuales registro un dispersión máxima de 0.52 en la época 35, conforme iba avanzando se ajustaron los filtros de fusión simétrica, la métrica va descendiendo de manera drástica hasta que alcanza una meseta monótona llegando a valores inferiores de 0.002 en las últimas 30 épocas.

Por otro lado en la gráfica de la derecha se observa que el PSNR tiene una una tendencia global ascendente, al principio tiene caídas transitorias las cuales fueron causadas pos las oscilaciones de los pesos, la métrica se escala progresivamente desde las cotas que son inferiores a 10dB hasta superar los 30dB al aproximarse a la época 150 esto logra validar que la red logro reconstruir el objeto así como reducir el ruido.

Por ultimo, el SSIM experimento caídas transitorias en la primera mitad del entrenamiento llegando a valores mínimos de 0.15 en la época 70, este indice entra a una zona de saturación superior a partir de la época 115 lográndose estabilizar sobre 0.98 y aproximándose a 0.992 al finalizar el ciclo.

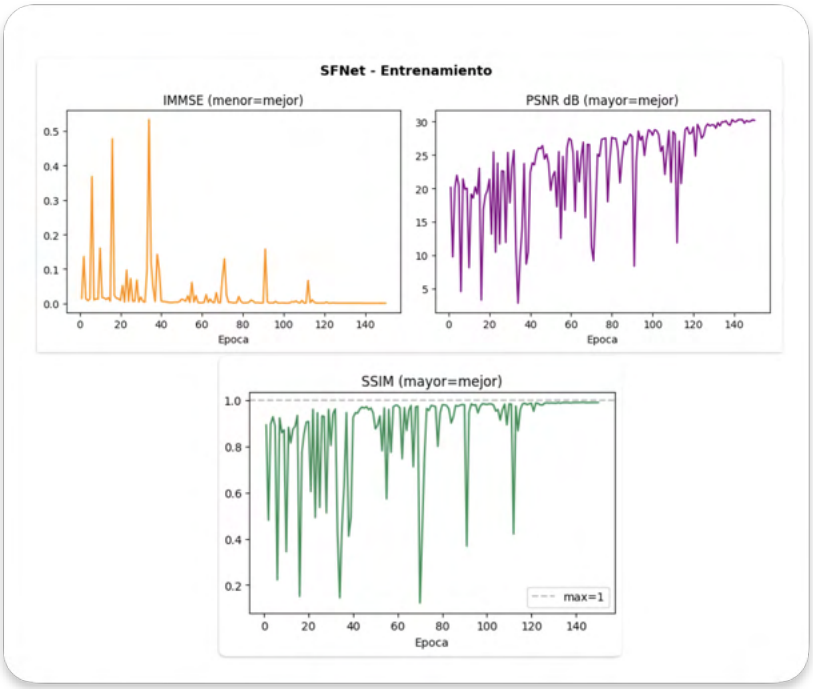


Figura 43: Curvas de entrenamiento de las métricas SFNet. Fuente propia

7.3.2.1. Interpretación de la métricas

La evaluación de esta arquitectura puede confirmar que puede resolver de manera reconstrucción 3D obteniendo los siguientes resultados los cuales se muestran en la tabla (4).

Métrica	Valor
IMMSE	0.001102
PSNR (dB)	30.620004
SSIM	0.991758

Cuadro 4: Resultados del conjunto de prueba (Test) para SFNet.

- IMMSE: la red dio un valor de 0.001102 en donde la discrepancia normalizada se mantiene en el orden 10^{-3} en comparación con lo obtenido en la red UHRNet esta ha sido el superior situandose a mas de dos ordenes de magnitud por debajo del método tradicional usando el PU de Ghiglia confirmando que este tipo de arquitectura puede suprimir la acumulación de desfase de 2π.

- PSNR: Acá dio un valor de 30.620004 dB superando el umbral de 30 dB demostrando que se tiene una recuperación adecuada de la profundidad promedio

- SSIM: dio un valor de 0.991758 situándose por encima de 0.99 dando a entender que al red no deforma las siluetas ni destruye la estructura de los objetos

7.3.2.2. Análisis de las muestras de prueba

En la figura 44 se pueden observar lo resultados de los mismos tres objetos que se usaron con UHRNet donde en la fila superior se encuentran la imagen con franjas, el GT, la predicción de la SFNet y el mapa termico del error absoluto ($|y - \hat{y}|$) y en la fila inferior la representaciones tridimensionales. Las muestras llevan el mismo orden que la red UHRNet, en este caso apra la SFNet con el primer objeto la red logro obtener como resultado un IMMSE = 0.001584, un PSNR = 28.0019 dB y un SSIM = 0.9792. Para este caso la red logro recuperar la geometría del cuerpo así como de la extremidades, sin embargo, el mapa de error absoluto denota concentraciones intensas de desviación local que son en este caso los tonos blancos y amarillos los cuales alcanzan valores de 0.4-0.5, en la parte tridimensional tiene una ligera perdida de definición en las facciones y un leve aplanamiento en los pico de máxima elevación en contraste con el plano de referencia. Para la segunda muestra se obtuvieron como resultados: un IMMSE=0.000310, PSNR=35.0918 dB y SSIM=0.9892, esto nos da a conocer que la red presenta un comportamiento sobresaliente con un PSNR ¿35 dB, para la parte tridimensional muestra un acoplamiento casi perfecto, por otra parte, el mapa de error absoluto se mantiene en los niveles mínimos lográndose observar unicamente unas ligeras diferencias de gradiente en la parte del lomo dorsal y en la punta extrema de la cola.

Por ultimo, en la muestra 3 las capas convolucionales piramidales de las SFNet atenúa la resolución de alta frecuencia en los contornos, esto se ve reflejado en el mapa térmico con los errores concentrados en las aristas de las alas y el soporte, en la parte tridimensional se producen bordes con rugosidades sutiles y con menor nitidez en la delimitación de los huecos libres.

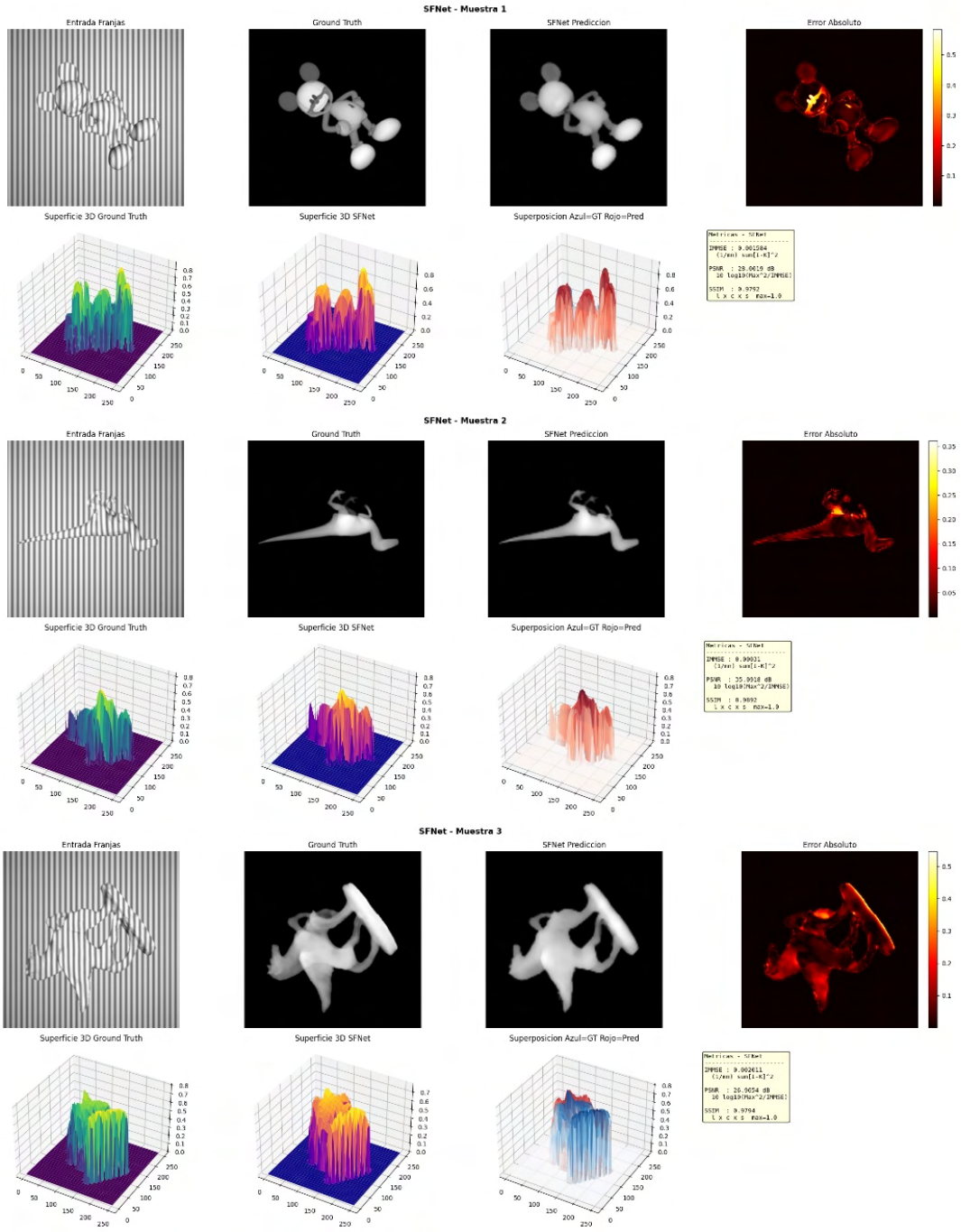


Figura 44: Muestras de prueba para la red SFNet. Fuente propia

7.3.3. Resultados de la Red Hformer

Esta es una arquitectura híbrida la cual combina convoluciones locales las cuales son separables en profundidad y bloques de autoatención en ventana cruzada (Cross-scale Auto-Attention, CAT), este tipo de red fue entrenada durante 150 épocas mediante

un optimizador AdamW ($LR = 3 \times 10^{-4}$) con decaimiento de pesos desacoplados y activaciones GELU. En la figura 45 se muestran las curvas de entrenamiento de dicha red.

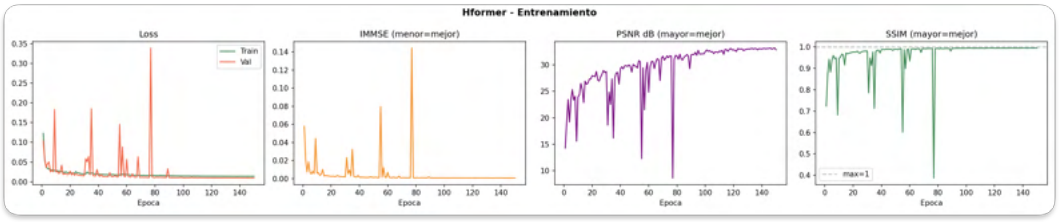


Figura 45: Entrenamiento de la red Hformer. Fuente propia

En la figura 46 se muestra la curva de entrenamiento de la validación de la perdida, se observa que va decayendo de manera continua desde la primera época, lográndose estabilizar por debajo de 0.012, por otro lado la parte de la perdida de validación sigue en la misma trayectoria descendente sin embargo esta registra picos transitorios pronunciados en las épocas 10, 35, 55 y un punto alto en la época 78 alcanzando hasta 0.34. Dicho fenómeno se atribuye al reajuste de los pesos en los bloques de autoatención en ventana cruzada (CAT) los cuales se basan en recalcular de manera dinámica las dependencias de largo alcance sobre los lotes con variaciones topográficas complejas, pasando la época 80 la curva de validación se estabiliza de manera uniforme acoplándose a la del entrenamiento causando la convergencia sin sobreajuste.

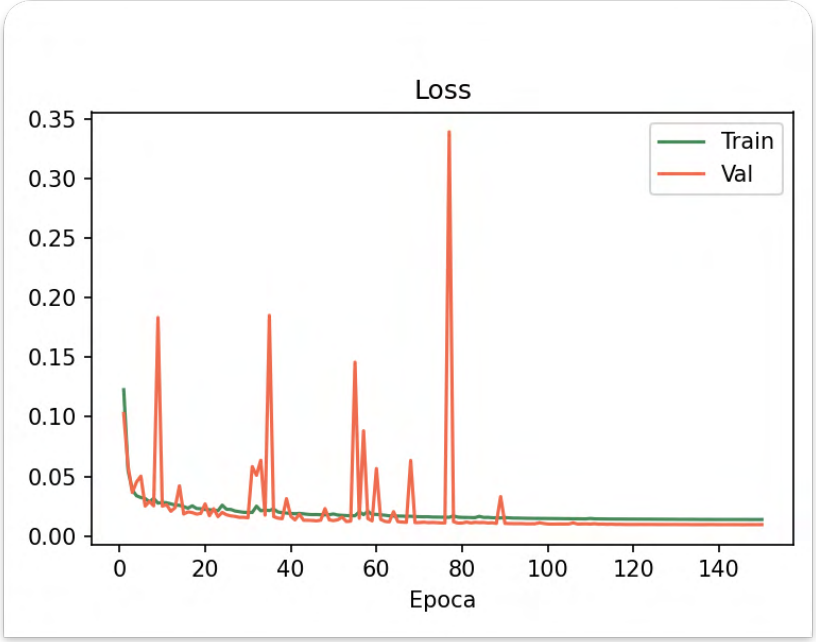


Figura 46: Validación de perdida Hformer. Fuente propia

Por otro lado en la figura 47 se muestran las curvas de entrenamiento pero de las métricas, en la gráfica del IMMSE tiene perturbaciones en la primera mitad del entrenamiento con un pico máximo de 0.14 en la época 78, la curva va descendiendo de una forma abrupta entrando en una fase de decaimiento asintótico plano a partir de la época 90, fijándose en valores cercanos a 0, logrando demostrar que la red aprendió a minimizar las diferencias cuadráticas pixel a pixel en toda la superficie.

El PSNR refleja una tendencia predominante ascendente sin embargo presenta caídas transitorias con los picos de perdida en la época 78 se puede observar de manera notoria esta caída logrando llegar a los 8.5 dB, el modelo se recupera de manera inmediata, a partir de la época 90 la métrica supera los 30 dB logrando superar los 33 dB hasta el final del ciclo.

Por ultimo, el SSIM tiene una caída llegando a 0.38 en la época 78, este índice se restituye con rapidez entrando en saturación por encima de 0.99 durante las ultimas 60 épocas, esta estabilización en valores cercanos a la unidad comprueba que la combinación de convoluciones locales y transformadores de atención preserva la correlación espacial, el contraste de curvaturas y los contornos geométricos de relieve.

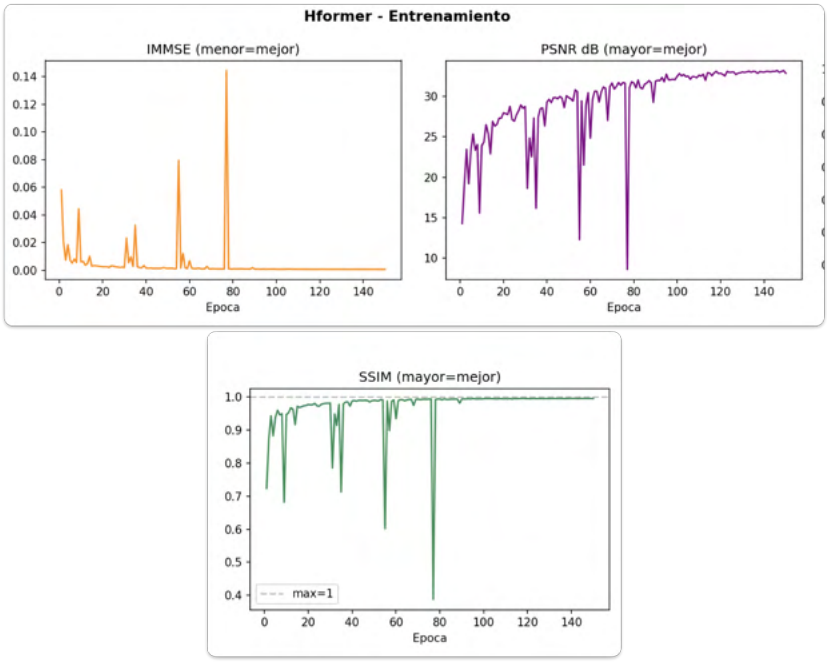


Figura 47: curvas de entrenamiento de las métricas Hformer. Fuente propia

7.3.3.1. Interpretación de la métricas

En la tabla (5) se observan los resultados para esta arquitectura.

Métrica	Valor
IMMSE	0.000609
PSNR (dB)	33.055269
SSIM	0.995486

Cuadro 5: Resultados del conjunto de prueba (Test) para Hformer.

Tomando en cuenta los resultados mostrados en la tabla se pueden interpretar de la siguiente manera:

- IMMSE = 0.000609: se mantiene en un orde 10^{-4} esto evidencia que el modelo desacopla de manera eficaz la deformación de las franjas sin introducir desfases continuos ni distorsiones acumulativos

- PSNR = 33.055269 dB: este supera el umbral de 30 dB obteniendo una recosntrucción de alta calidad en donde la señal reconstruida predomina sobre cualquier residuo numerico frente al GT

- SSIM = 0.995486: teniendo un valor mayor a 0.995 se puede validar una conservación estructurar el cual es prácticamente integral de la geometrías tridimensionales.

7.3.3.2. Análisis de las muestras de prueba

La figura 48 cuenta con las mismas características que la arquitectura anteriores, sin embargo en esta se presentan las pruebas para Hformer.

Para la primera muestra se obtienen los siguientes resultados: IMMSE = 0.001320, PSNR=28.7940 dB y SSIM = 0.9831, en este punto la Hformer recupera con fluidez la volumetría del torso, cabeza y extremidades alcanzando una cota maxima de elevación de $z \approx 0{,}8$ en la parte tridimensional, en el mapa de error absoluto se puede observar que el fondo permanece con un valor nulo registrándose las mayores discrepancias en las transiciones abruptas del rostro asi como la unión del cuello.

Para la muestra dos los resultados obtenidos fueron: IMMSE=0.000329, PSNR=35.3775 dB, SSIM=0.9910, con estos resultados se constituye una reconstrucción de alta precisión geométrica, logrando que se pueda acoplar directamente sobre GT en la superposición 3D.

Para la ultima muestra se obtuvieron las métrica locales de la siguiente manera: IMMSE=0.000676, PSNR=31.7014 dB y un SSIM=0.9930, en esta parte de evalúa el comportamiento ante las regiones que no son conexas y cavidades interiores. Los módulos CAT así como la red de pirámide de características (FPN) de Hformer permiten que se pueda comprender la continuidad global del objeto, logrando que se delimiten los huecos sin propagar errores al fondo, en la reconstrucción 3D se logra observar la nitidez de las paredes de las alas y el soporte manteniendo una similitud estructural de 0.9930

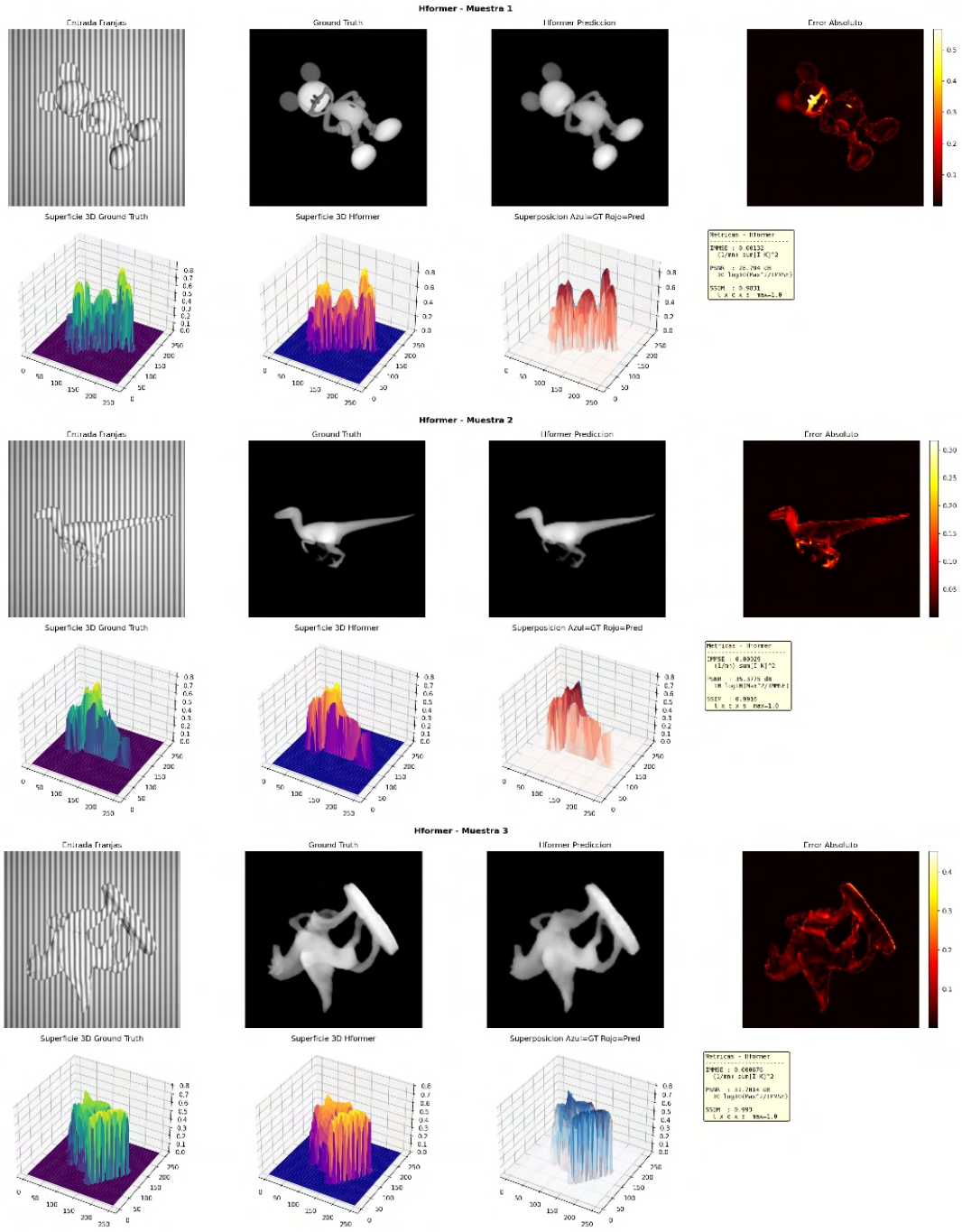


Figura 48: Muestras de prueba para la red Hformer. Fuente propia

7.4. Comparativo de las redes

A continuación se hará una comparativa entre las tres arquitecturas utilizadas (UHRNet, SFNet y Hformer) en donde se analizaran sus métricas cuantitativas, su

fidelidad mediante el error cuadrático medio la reconstrucción 3D frente al Ground Truth.

En la figura 49 se muestra una gráfica de barras en donde se comparan de manera estadística los resultados de las métricas para estas tres redes.

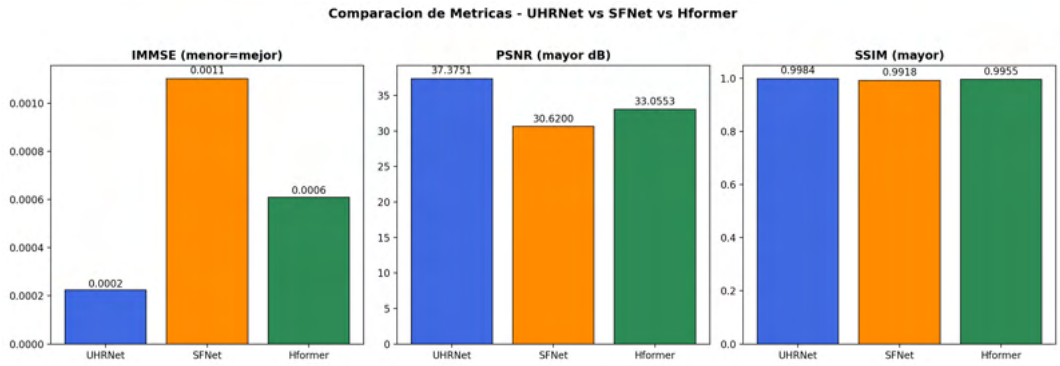


Figura 49: Comparativa entre las tres redes. Fuente propia

Se puede observar la diferencia entre estas tres redes para cada una de sus métricas, en el caso del IMMSE la arquitectura SFNet presenta el mayor error cuadrático medio con un valor de 0.0011 esto se debe a perdida de resolución espacial en la parte de codificador piramidal, sin embargo la Hformer reduce el error de SFNet en un 44.7 % con un valor de 0.0006 evidenciando que los módulos de atención pueden modelar con mayor precisión la fase que las convoluciones estándar y por ultimo la UHRNet logra tener una menor discrepancia cuadrática global lo ayudo a disminuir el error de Hformer en un 63.2 % y el de SFNet en un 80 % obteniendo un valor de 0.0002.

En el PSNR la que tuvo un mejor resultado fue la UHRNet de 37.3751 dB superando a la Hformer por mas de 4.32 dB y a la SFNet por 6.75 dB lo que hace que tenga una reconstrucción con mínimo ruido residual. Por otro lado en el SSIM la UHRNet preserva de manera mas consistente las transiciones de bordes con un valor de 0.9984 lo cual es mas cercano a 1. Los valores de las otras arquitecturas se pueden observar en la tabla (6).

Modelo Evaluado	**IMMSE Global ↓**	**PSNR Global (dB) ↑**	**SSIM Global ↑**
SFNet (150 épocas)	0.001102	30.620004	0.991758
Hformer (150 épocas)	0.000609	33.055269	0.995486
UHRNet (200 épocas)	0.000224	37.375070	0.998373

Cuadro 6: Comparación global de métricas por modelo evaluado.

En la figura 50 se muestra la comparativa visual en 2D, mostrando la entrada que es el objeto con franjas, el GT y la predicción de cada red, tomando en cuenta las tres muestras.

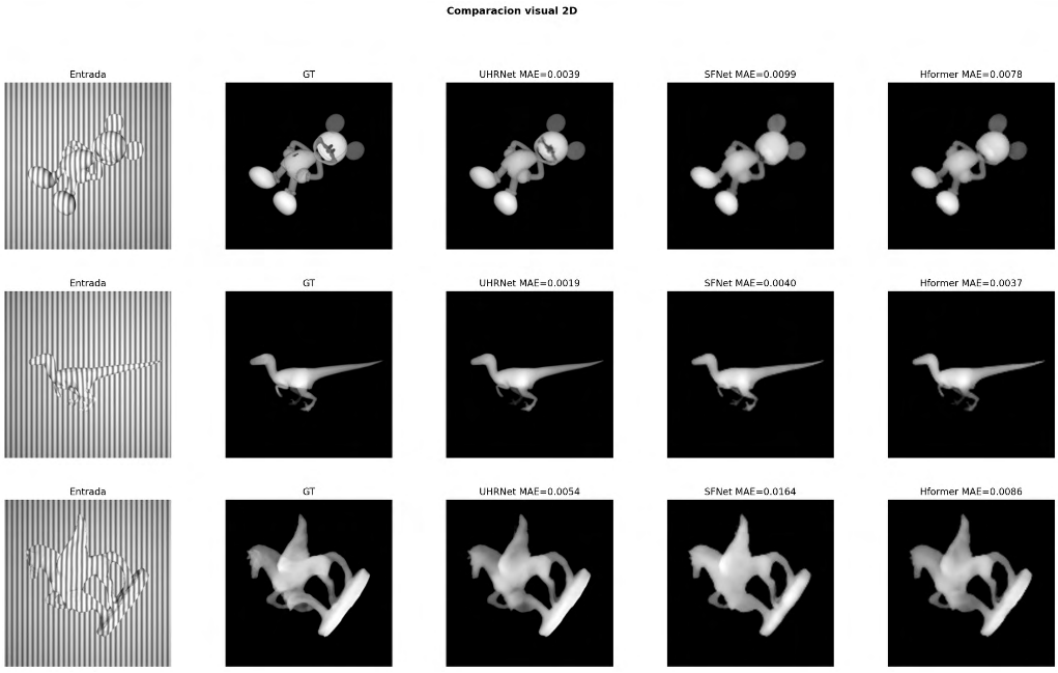


Figura 50: Comparativa entre las tres redes. Fuente propia

En la siguiente tabla (7) se presentan los resultados del error cuadrático medio de cada una de las arquitecturas. En donde se puede observar que para primera muestra la UHRNet tiene mejores resultados debido a que preserva con nitidez la separación facial asi como la orejas esféricas, la Hformer reproduce la geometría global con un error contenido mientras que la SFNet presenta un difuminado mayor en las cavidades del rostro lo que hace que se incremente su error medio de 10^{-2}. Para la muestra dos que es el dinosaurio al ser un objeto suave y continuo, las tres redes obtiene una predicción con una alta fidelidad sin embargo la UHRNet reduce a la mitad el error absoluto en la extremidades inferiores así como en la terminación final de la cola con respecto a Hformer y SFNet. Por ultimo, el pegaso representa un desafió debido a los orificios del pedestal en este caso la SFNet sufre una degradación notable lo que genero ensanchamiento en la fronteras, la Hformer mitiga este efecto con ayuda de la ventana de atención mientras que la UHRNet logra aislar los huecos manteniendo una delimitación limpia.

Modelo	**MAE (Mickey)**	**MAE (Dinosaurio)**	**MAE (Pegaso)**
UHRNet	**0.0039**	**0.0019**	**0.0054**
Hformer	0.0078	0.0037	0.0086
SFNet	0.0099	0.0040	0.0164

Cuadro 7: Comparación de MAE por modelo en los objetos evaluados (Mickey, Dinosaurio y Pegaso).

Para el análisis tridimensional esta vez solo se observara con la primera muestra como se observa en la figura 51 en donde se contrasta con respecto al GT y las tres mallas generadas.

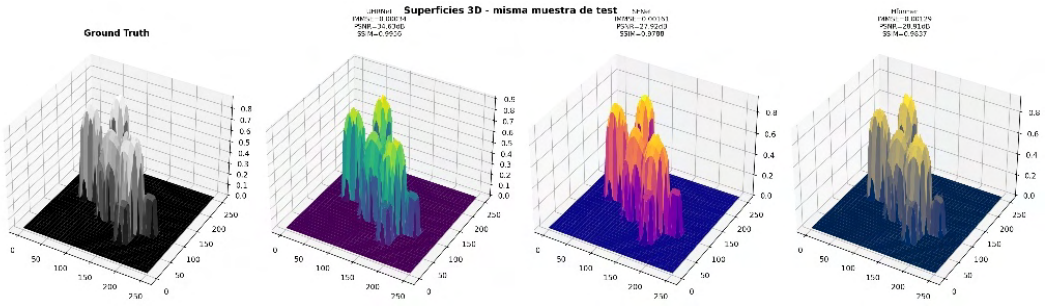


Figura 51: Comparativa entre las tres redes 3D. Fuente propia

Para la UHRNet se obtuvieron los resultados: IMMSE=0.00034, PSNR = 34.63 dB, SSIM=0.9956, en este caso el volumen de los tres picos coincide con GT sin truncar su altura máxima $Z \approx 0{,}8$, la pendientes que van de caída hacia lo que es el plano base se conservan verticales y definidas, sin desbordamiento de fase hacia el fondo. Por el contrario, la SFNet obtuvo un IMMSE=0.00158, PSNR=27.92 dB y SSIM=0.9700 aca se puede apreciar un aplanamiento en las esferas en donde la elevación maxima decae ligeramente por debajo de 0.8. Las transiciones entre la cabeza y los hombros representan rugosidades las cuales son visibles dispersando la continuidad superficial. Por ultimo, la Hformer obtuvo: IMMSE=0.00129, PSNR=28.91 dB y SSIM=00.9837, en este caso esta arquitectura recupera mejor la curvatura cóncava así como las alturas máximas en comparación con SFNet, logrando mantener una morfología mas suaves aunque tiene menor definición de aristas que UHRNet.

VIII. Conclusiones

En esta tesis se describió todo lo realizado en esta investigación para poder disminuir el error de la perfilometria por transformada de fourier con Redes Neuronales Convolucionales, demostrando como el aprendizaje automático ayuda a superar las limitaciones estructurales y matemáticas inherentes a las técnicas ópticas analíticas clásicas resolviendo las demodulaciones de la señal luminosa y la reconstrucción tridimensional de objetos complejos. A partir de del desarrollo del método tradicional se pudo comprobar que el pipeline clásico basado en filtrado espectral y desenvolvimiento de fase presenta una limitación ante geometrías aisladas y bordes abruptos. La aplicación de filtros en frecuencias ayudo a aislar la componente portadora atenuando severamente las componentes que tienen altas frecuencias espaciales, suavizando los detalles finos de la superficie. Por otro lado los algoritmos de desenvolvimiento de fase colapsan ante los saltos abruptos de altura y sombras esto es debido a que dependen de la continuidad matemática espacial o de la formulación de mínimos cuadrados, esto provoco que se introdujeran rampas de fase globales, filtrado excesivo provocando que no se lograra una buena reconstrucción.

La experimentación sistemática ayudo a evidenciar que el filtrado de la frecuencias introduce una alternativa crítica: las mascaras estrechas pueden atenuar el ruido pero suprimir de manera irreversible las altas frecuencias espaciales las cuales contienen los detalles morfológicos finos del objeto, sin embargo las ventanas amplias inducen solapamiento espectral con la componente continua y los armónicos superiores. Los algoritmos de desenvolvimiento de fase como Ghiglia, Itoh, Herraez, Zhao, entre otros al sustentarse en supuestos de continuidad suaves colapsan de manera irremediable en áreas de sobras o salto de fase pronunciados.

Para poder resolver la falta de datos reales volumétricos con una referencia exacta, se creo una base de datos sintética con patrones de franjas proyectadas y Grund Truth demostro ser un entorno de modelado indispensable y reproducible. Esto permitió que se desacoplaran las fuentes de error externas como las fluctuaciones de la luz ambiental, ruido de cuantización del sensor o distorsiones no lineales del proyector, esto se logro debido a que en blender se puede poner las características que se necesitan con la ventaja de no introducir ruido, también se suministro una referencia geométrica sin sesgos lo cual guio de forma optima la convergencia supervisadas de las arquitecturas.

Tomando en cuenta este contexto, la propuesta e implementación de modelos convolucionales profundos destacando en este caso la arquitectura UHRNet con ramas multirresolución en paralelo al dar buenos resultados logro evidenciar que el aprendizaje automático es capaz de poder emular e integrar la física de deformación de franjas sin requerir de una secuencia analítica convencional de transformada de fourier, diseño de los filtros, calculo de la fase envuelta ni desdoblamiento de fase, la red logro superar al ambigüedad periódica de 2π con ayuda de las dependencias contextuales y espaciales completas, conservando los detalles finos así como las discontinuidades morfológicas sin sufrir la degradación de los métodos clásicos.

La evaluación comparativa cuantitativa entre el método tradicional y en enfoque con

aprendizaje automático confirmo que se logro la reducción del error de reconstrucción. Los modelos con redes neuronales convolucionales lograron reducir el error cuadrático medio hasta tres veces su magnitud respecto al desempeño optimo obtenido por algoritmos clásicos, aumentando los niveles de correlación y alcanzando índices de similitud estructural cercanos a la unidad frente al Ground Truth. Sin embargo la perfilometria de fourier tradicional introduce un suavizado excesivo y el modelo convolucional puede restaurar con alta fidelidad las aristas vivas, las pendientes pronunciadas y la textura volumétrica del objeto.

Asimismo, tanto las pruebas experimentales como el protocolo de optimización de los hiperparametros demostraron una elevada capacidad de generalización de los hiperparametros demostraron tener una alta capacidad de generalización y robustez operativa. El empleo de las funciones de perdida compuestas con ayuda de esquemas de aprendizaje adaptativo, permitió que se pudiera calibrar las sensibilidad frente a las perturbaciones luminosas, zonas de baja reflectancia y fluctuaciones operativas de los objetos reales.

Este trabajo logro demostrar que el uso de la perfilometria de fourier tradicional tiene bastante limitaciones sin embargo este mismo se puede mejorar con la ayuda de modelos de inteligencia artificial superando una disyuntiva historia entre la velocidad de adquisición que es la captura a partir de un solo patrón y la precisión geometría en superficies que están morfológicamente complejas. También se puede validar que el uso de las redes neuronales convolucionales puede ser una herramienta precisa, eficiente y completamente automatizada para la digitalización y el análisis tridimensional de los objetos en ingeniería, ciencias de los materiales e inspección óptica avanzada.

Referencias bibliográficas

[1] S. Aciar, H. Chanampe, and M. de la Vega, "Deep learning para el reconocimiento de patrones," proyecto CICYT 056/17, s.f.

[2] L. A. Català, "Reconstrucción de objetos 3d mediante el uso de redes neuronales: del modelo real a su impresión física," Barcelona, España, 2024.

[3] M. Takeda and K. Mutoh, "Fourier transform profilometry," *Applied Optics*, vol. 22, no. 24, pp. 3977–3982, 1983.

[4] M. A. Pérez, "Redes neuronales con convoluciones en el dominio de fourier para la localización de robots móviles," Master's thesis, Universidad Miguel Hernández de Elche, Elche, España, 2024.

[5] M. A. C. Chávez, "Medición de perfiles utilizando técnicas de proyección de franjas," Master's thesis, Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE), Puebla, México, 2012.

[6] L. Chi, B. Jiang, and Y. Mu, "Fast fourier convolution," in *Advances in Neural Information Processing Systems*, vol. 33, 2020, pp. 4479–4488.

[7] A. D. et al., "An image is worth 16x16 words: Transformers for image recognition at scale," in *International Conference on Learning Representations (ICLR)*, 2021.

[8] D. D. González and N. J. C. Jaramillo, "Identificación de armas de fuego mediante técnicas de reconocimiento en el procesamiento digital de imágenes," Pereira, Colombia, 2019.

[9] O. A. E.-B. et al., "Quasi/periodic noise reduction in images using modified multiresolution-convolutional neural networks for 3d object reconstructions and comparison with other convolutional neural network models," *Computers*, vol. 13, no. 6, p. 145, 2024.

[10] S. F. et al., "Fringe pattern analysis using deep learning," *Advanced Photonics*, vol. 1, no. 2, p. 025001, 2019.

[11] ——, "Fringe pattern analysis using deep learning: Supplementary information," 2019.

[12] H. A. F. Hernández, "Inversión 3d de datos gravimétricos mediante redes neuronales u-net en la zona de yucatán, golfo de méxico," Master's thesis, UNAM, Ciudad de México, México, 2025.

[13] Y. Han and B.-W. Hong, "Deep learning based on fourier convolutional neural network incorporating random kernels," *Electronics*, vol. 10, no. 16, p. 2004, 2021.

[14] X. He and K. Qian, "A comparative study on temporal phase unwrapping methods in high-speed fringe projection profilometry," *Optics and Lasers in Engineering*, vol. 142, p. 106613, 2021.

[15] G. E. K. et al., "Physics-informed machine learning," *Nature Reviews Physics*, vol. 3, pp. 422–440, 2021.

[16] Q. Kemao, "Windowed fourier transform for fringe pattern analysis: principles, applications and implementations," *Optics and Lasers in Engineering*, vol. 45, no. 2, pp. 304–317, 2007.

[17] S. Kumar, "Phase retrieval with physics informed zero-shot network," *Optics Letters*, vol. 46, pp. 5942–5945, 2021.

[18] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[19] T. L. et al., "Feature pyramid networks for object detection," in *CVPR*, 2017, pp. 2117–2125.

[20] D. E. L. Portillo, "Corrección de oclusiones para la obtención de información 3d en objetos complejos," Master's thesis, Instituto Tecnológico de Chihuahua, Chihuahua, México, 2019.

[21] "Machine learning algorithms and performance metrics in engineering," *Eng*, vol. 4, 2023.

[22] V. M. M. Suárez, "Reconstrucción de formas 3d mediante algoritmo de desplazamiento de fase estéreo usando proyección de franjas y una red neuronal convolucional," Master's thesis, UAQ, Querétaro, México, 2023.

[23] I. M. M. Cuadrado, "Entrenamiento de redes convolucionales mediante la transformada de fourier," España, 2024.

[24] J. C. M. Morales, "Desarrollo e implementación de un algoritmo para la reconstrucción tridimensional de sólidos basado en técnicas perfilométricas," Master's thesis, UAQ, Querétaro, México, 2014.

[25] H. Nguyen, E. Novak, and Z. Wang, "Accurate 3d reconstruction via fringe-to-phase network," *Measurement*, vol. 190, p. 110663, 2022.

[26] H. Nguyen, Y. Wang, and Z. Wang, "Learning-based 3d imaging from single structured-light image," *Graphical Models*, vol. 126, p. 101171, 2023.

[27] O. Pereira and R. Kindelan, "Reconstrucción tridimensional de modelos anatómicos a partir de imágenes médicas digitales," in *Informática Habana 2009*, Cuba, 2009.

[28] H. P. et al., "Fcnn: Fourier convolutional neural networks," in *Springer*, 2017, pp. 786–798.

[29] Rama Estudiantil IEEE CUN, "Materiales de investigación: Tecnologías de impresión 3d y prototipado rápido," colombia, s.f.

[30] J. R. et al., "Reconstrucción tridimensional de superficies en el cuerpo humano," *Información Tecnológica*, vol. 24, no. 5, pp. 31–40, 2013.

[31] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *MICCAI*, 2015, pp. 234–241.

[32] A. A. S. Tufiño, "Desarrollo de un sistema de visión artificial basado en redes convolucionales," Ecuador, 2022.

[33] G. E. S. et al., "Phasenet," *IEEE Signal Processing Letters*, vol. 26, no. 1, pp. 54–58, 2019.

[34] S. V. der Jeught and J. Dirckx, "Deep neural networks for single shot structured light profilometry," *Optics Express*, vol. 27, no. 12, p. 17091, 2019.

[35] S. V. der Jeught et al., "Optimized loss function in deep learning profilometry," *Journal of Physics: Photonics*, vol. 3, no. 2, p. 024014, 2021.

[36] C. W. Gutierrez, "Tratamiento de imágenes con machine learning para reconstrucción tridimensional," Master's thesis, TESCI, México, 2022.

[37] K. W. et al., "Deep learning for phase processing," *Light: Science & Applications*, vol. 13, 2024.

[38] L. Wang, D. Lu, R. Qiu, and J. Tao, "3d reconstruction from structured-light profilometry," *EURASIP Journal*, vol. 2022, no. 14, 2022.

[39] Y. W. et al., "Uhrnet," *arXiv*, 2023.

[40] Z. W. et al., "Review of fringe projection profilometry," 2024.

[41] W. Y. et al., "Physics-informed deep learning for fringe pattern analysis," *Opto-Electronic Advances*, 2024.

[42] Z. Z. et al., "Fringe-based structured-light 3d reconstruction," *Sensors*, 2025.

[43] Y. Z. et al., "Cnn vs transformer vs mlp," *arXiv*, 2021.

[44] M. Y. G. Echevarría, F. E. A. Serrano, J. C. P. Ortega, O. A. E. Bernal, J. M. R. Arreguin, and S. A. Tovar, "Mejora en la perfilometría por transformada de fourier 2d utilizando filtrado con distintas formas de máscara," Querétaro, México, documento académico.

[45] K. L. V. Ruiz, O. A. E. Bernal, J. C. P. Ortega, J. M. R. Arreguín, A. S. Olmedo, and M. A. A. Fernández, "Estudio comparativo de algoritmos de desdoblamiento de fase espacial y aplicación en reconstrucción 3d," Querétaro, México, s.f.

[46] C. Zuo, J. Qian, S. Feng, W. Yin, Y. Li, and Q. Chen, "Deep learning in optical metrology: a review," *Light: Science & Applications*, vol. 11, no. 1, p. 39, 2022.

[47] X. Z. et al., "Hformer," *arXiv*, 2021.

[48] ——, "Phase retrieval from single frame projection fringe pattern," *Optics and Lasers in Engineering*, 2014.

[49] C. Z. et al., "Phase shifting algorithms review," *Optics and Lasers in Engineering*, 2018.

[50] ——, "Temporal phase unwrapping review," *Optics and Lasers in Engineering*, 2016.

[51] M. Takeda, H. Ina, and S. Kobayashi, "Fourier-transform method of fringe-pattern analysis for computer-based topography and interferometry," *Journal of the Optical Society of America*, vol. 72, no. 1, pp. 156–160, 1982.

[52] C. Quan, W. Chen, and C. J. Tay, "Phase-retrieval techniques in fringe-projection profilometry," *Optics and Lasers in Engineering*, vol. 48, no. 2, pp. 235–243, 2010.

[53] J. C. Pedraza-Ortega, E. Gorrostieta-Hurtado, M. A. Aceves-Fernández, A. Sotomayor-Olmedo, J. M. Ramos-Arreguín, S. Tovar-Arriaga, and J. E. Vargas-Soto, "A 3d sensor based on a profilometrical approach," *Sensors*, vol. 9, no. 12, pp. 10 326–10 340, 2009.

[54] J. C. M. Morales, C. A. Ramos-Arreguín, A. Sotomayor-Olmedo, E. Gorrostieta-Hurtado, J. M. Ramos-Arreguín, J. C. P. Ortega, S. T. Arriaga, and J. E. V. Soto, "A strategy for 3d object digitalization using pre-filtering and post-filtering stages," *Procedia Technology*, vol. 3, pp. 273–281, 2012.

[55] R. C. González and R. E. Woods, *Digital Image Processing*, 3rd ed. Upper Saddle River, NJ, USA: Prentice Hall, 2008.

[56] K. J. Gasvik, *Optical Metrology*, 3rd ed. Chichester, UK: John Wiley & Sons, 2002.

[57] J. W. Goodman, *Introduction to Fourier Optics*, 4th ed. New York, NY, USA: W. H. Freeman, 2017.

[58] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[59] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2nd ed. Upper Saddle River, NJ, USA: Prentice Hall, 1999.

[60] R. Flórez and J. Fernández, *Las redes neuronales artificiales: fundamentos teóricos y aplicaciones prácticas*. Oleiros, España: Netbiblo, 2008.

[61] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 25, 2012, pp. 1097–1105.

[62] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.

[63] L. Grafakos, *Classical and Modern Fourier Analysis*. New York, NY, USA: Springer, 2014.