

Implementación de algoritmos de aprendizaje profundo para la superresolución de imágenes
satelitales

Juan Sebastian Jaramillo Diaz, y Jonathan Mauricio Rojas Rodriguez

Trabajo de Grado para optar por el título de ingeniero electrónico

Director

PhD. Hans Yecid Garcia Arenas

Codirectores

Msc. Leandro Sebasthyan Rojas Rodríguez

Msc(c). Sebastian Ardila Leal

Universidad Industrial de Santander

Facultad de Ingenierías Fisicomecánicas

Escuela de Ingenierías Eléctrica, Electrónica y de Telecomunicaciones

Bucaramanga

2026

Dedicatoria

Juan Sebastian Jaramillo Diaz

Dedico este triunfo con el corazón lleno de amor y una gratitud infinita a mis padres, Maria Cristina Diaz Cristancho y Fernando Antonio Jaramillo Muñoz, quienes han sido el principio, el sostén y el motor de mi existencia; gracias por cada sacrificio silencioso, por las largas jornadas de desvelo que asumieron como propias y por ese acompañamiento incondicional que se convirtió en mi refugio más seguro en los momentos de mayor incertidumbre académica y personal. Este título universitario no es solo mío, sino el fruto de su entrega diaria, de su fe ciega en mis capacidades y de haberme enseñado con su ejemplo que el trabajo honesto y la perseverancia son las llaves para alcanzar cualquier meta. Con esa misma profundidad y con una nostalgia que se transforma en orgullo, dedico este logro a la memoria de mi abuela, Julia Cristancho Moreno, quien, a pesar de que ya no nos acompaña físicamente en este plano terrenal, ha sido un pilar invisible y fundamental a lo largo de toda mi carrera; su recuerdo, su fortaleza inquebrantable y sus bendiciones eternas han sido la luz que guio mis pasos en la distancia, impulsándome a no rendirme y a honrar hoy su maravilloso legado. Asimismo, comparto la inmensa alegría de este peldaño alcanzado con mi abuelo, Juan de la Cruz Diaz Pedraza, agradeciéndole desde el fondo de mi alma su presencia constante, sus sabios consejos y por ser esa raíz firme y sabia de nuestra familia que hoy celebra con orgullo el cumplimiento de este gran sueño compartido.

Jonatan Mauricio Rojas Rodriguez

A Dios, por guiar mis pasos hasta este momento, por darme la salud, la sabiduría en los días de mayor incertidumbre y la resiliencia para culminar con éxito esta exigente etapa, y a mis

padres, Liliana Rodríguez Castro y Mauricio Rojas Rincón.

Dedico este trabajo de grado, y todo el inmenso esfuerzo que sus páginas representan, a ustedes, los pilares inquebrantables de mi vida y mi mayor ejemplo a seguir. A lo largo de todos estos años de formación, han sido mi mayor refugio, mi lugar seguro y mi fuente inagotable de motivación cuando el camino hacia este título se tornaba cuesta arriba.

Gracias por cada sacrificio silencioso que hicieron para que a mí no me faltara nada, por postergar sus propios anhelos para poder impulsar los míos con una generosidad que no conoce límites. Gracias por brindarme no solo las herramientas materiales indispensables para avanzar, sino también, y más importante aún, la fortaleza mental, la ética de trabajo y los valores para no rendirme ante las adversidades. Cada noche de traspasar frente a la pantalla resolviendo problemas, cada momento de frustración ante un código que no funcionaba o un examen complejo en la universidad, y cada pequeño avance logrado en esta investigación, llevan impregnados su amor incondicional y la confianza ciega que siempre depositaron en mí. Ustedes creyeron en mi capacidad y en mi talento mucho antes de que yo mismo lo hiciera, y esa fe intacta fue el verdadero motor que me impulsó a superar mis propios límites y a cruzar esta anhelada línea de meta.

Hoy miro hacia atrás, contemplo todo el camino recorrido y me doy cuenta de que este título de ingeniería no me pertenece solo a mí, sino a nosotros tres. Es el fruto directo de sus enseñanzas, de sus sabios consejos, de sus palabras de aliento constante y de su alegría sincera por cada una de mis metas alcanzadas. Todo lo que soy hoy, y todo lo que aspiro a ser mañana como profesional y como ser humano, se los debo enteramente a ustedes. Gracias por darme raíces fuertes para no olvidar nunca de dónde vengo, por enseñarme el valor de la constancia, y por darme unas alas enormes para volar tan alto como me lo proponga.

Este triunfo, más que académico, es un triunfo familiar. Es infinitamente nuestro.

Agradecimientos

Juan Sebastian Jaramillo Diaz

Mirar hacia atrás y ver el camino recorrido en estos años de formación me llena de una inmensa gratitud, comenzando por Dios, cuya guía espiritual y fortaleza fueron mi sostén en los momentos de mayor debilidad, otorgándome la perseverancia para culminar este proyecto de vida. Este peldaño alcanzado es el fruto del amor sembrado por mis padres y mi familia, quienes con su apoyo incondicional, su comprensión y sus sabios consejos se convirtieron en mi puerto seguro y en el motor que me impulsó a dar lo mejor de mí cada día, demostrándome que ningún objetivo es inalcanzable cuando se cuenta con un hogar que cree en ti. En el ámbito académico, expreso mi sincero agradecimiento al doctor Hans Yecid García Arenas, por su invaluable dirección, paciencia y por compartir su conocimiento y rigurosidad científica en el desarrollo de este trabajo. De igual manera, mi reconocimiento y gratitud son para mi compañero de grado, Jonathan Mauricio Rojas, por su compromiso, esfuerzo conjunto y por hacer posible la culminación exitosa de este proyecto mediante un trabajo en equipo excepcional. Asimismo, hago extensivo este agradecimiento a mis compañeros y amigos, quienes me brindaron su apoyo y alegría a lo largo de la carrera, compartiendo experiencias inolvidables no solo en las aulas de clase, sino en cada vivencia que enriqueció esta etapa universitaria. Con el mismo sentimiento, rindo un sentido homenaje a aquellos seres queridos que ya no se encuentran físicamente a mi lado, pero cuya huella imborrable permanece viva en mi corazón; su partida dejó un vacío inmenso, pero su recuerdo y las bendiciones que desde la eternidad me guían se transforman hoy en la fuerza que me inspira a honrar su memoria a través de este y de todos mis éxitos futuros.

Jonathan Mauricio Rojas Rodriguez

En primer lugar, quiero expresar mi más profundo agradecimiento a la Universidad Industrial de Santander (UIS) y, de manera muy especial, a la Escuela de Ingenierías Eléctrica, Electrónica y de Telecomunicaciones (E3T). Gracias por ser mi segundo hogar durante estos años y por brindarme una formación académica de excelencia, llena de rigor y exigencia, que me ha preparado para afrontar los retos de la vida profesional.

Extiendo mi gratitud a mi director de proyecto, el Doctor Hans Yecid Garcia Arenas, por su innegable paciencia, su guía experta y su constante disposición para orientar el rumbo de esta investigación. Sus exigencias, revisiones y sabios consejos fueron fundamentales para llevar este trabajo a buen término y elevar su nivel técnico. De igual forma, agradezco sinceramente a mis codirectores, Sebastian Ardila Leal y Leandro Sebasthyan Rojas Rodríguez, por su invaluable apoyo, sus aportes metodológicos y por creer en el potencial de este proyecto desde el primer día.

A mi compañero de tesis y amigo, Juan Sebastian Jaramillo Diaz. Gracias por el inmenso compromiso, por las incontables horas de trabajo frente a las pantallas buscando optimizar los modelos, por la paciencia en los momentos de mayor estrés y por hacer de este desafío un verdadero trabajo en equipo. La dedicación en el procesamiento de los datos y en la documentación técnica hicieron que este logro compartido brille con luz propia.

Quiero extender un agradecimiento muy especial a todos mis compañeros y amigos con quienes compartí aulas, pasillos, laboratorios y largas jornadas de estudio a lo largo de toda la carrera. Gracias por la camaradería, por el apoyo mutuo en los momentos de mayor exigencia académica y por las anécdotas que hicieron de esta etapa universitaria una experiencia invaluable. El camino fue mucho más llevadero gracias a su compañía.

Asimismo, agradezco a todos los profesores de la facultad que, con su vocación y dedicación, aportaron a mi formación integral, cimentando las bases de mi conocimiento como ingeniero.

Finalmente, a todas las personas que, de manera directa o indirecta, brindaron un consejo, una palabra de aliento o su apoyo para la culminación de esta etapa. A todos ustedes, mil gracias.

Tabla de contenido

Introducción	18
1 Objetivos	21
1.1 Objetivo general	21
1.2 Objetivos específicos	21
2 Marco teórico y estado del arte	22
2.1 Motivación operativa y el problema de la resolución	22
2.2 Fundamentación física: Respuesta al impulso del sistema óptico	23
2.2.1 Caracterización de núcleos de degradación suministrados	23
2.3 Modelado matemático de la degradación y formación de imágenes	24
2.3.1 Caracterización estadística del ruido estocástico (η)	25
2.3.2 La superresolución como problema inverso	26
2.4 Estado del Arte y evolución de la Superresolución (SR)	26
2.4.1 La era de los Transformers en visión: SwinIR y modelos de atención global	27
2.5 Fundamentos de optimización y funciones de pérdida en superresolución	27
2.5.1 Funciones de pérdida híbridas y preservación estructural	28
2.5.2 Dinámica del aprendizaje y regularización estocástica	28
2.6 Métricas de evaluación de calidad	29
2.6.1 PSNR (Peak Signal-to-Noise Ratio)	29
2.6.2 SSIM (Structural Similarity Index)	30

3	Diseño del protocolo de preprocesamiento de datos	31
3.1	Acondicionamiento del dataset base	32
3.2	Degradación multiescala	34
3.3	Validación y resultados del protocolo de degradación	35
3.3.1	Inspección visual multiescala	36
3.4	Análisis del Impacto Estocástico (SNR)	37
4	Selección y Arquitectura de los Modelos de Superresolución	38
4.1	Arquitectura Base: Enhanced Deep Residual Networks (EDSR)	38
4.2	Arquitectura Avanzada: Swin Image Restoration (SwinIR)	39
4.3	Justificación del Análisis Comparativo	41
5	Entrenamiento de los Modelos	43
5.1	Evaluación Preliminar de las Funciones de Degradación	43
5.2	Configuración del Entorno y Hardware	44
5.3	Diseño de las Funciones de Pérdida Híbridas	45
5.3.1	Función de pérdida para la arquitectura convolucional (EDSR)	46
5.3.2	Función de pérdida para el modelo de atención (SwinIR)	46
5.4	Dinámicas de Entrenamiento	47
5.5	Monitoreo de Convergencia y Prevención de Sobreajuste	47
5.6	Protocolo de Evaluación Cuantitativa	48
5.7	Consideraciones de Complejidad Computacional y Memoria	48
6	Evaluación del Desempeño y Validación Experimental	50
6.1	Validación del Protocolo de Preparación de Datos	50
6.2	Análisis Comparativo EDSR vs. SwinIR	51
6.2.1	Convergencia y Funciones de Pérdida	51
6.2.2	Evaluación Cuantitativa	51
6.2.3	Inspección Visual y Discusión	55

SUPERRESOLUCIÓN DE IMÁGENES SATELITALES	8
7 Conclusiones	56
8 Recomendaciones	59
8.1 Ampliación y validación del modelo de degradación	59
8.2 Optimización arquitectónica y eficiencia computacional	59
8.3 Evolución de las métricas de evaluación	60
8.4 Proyección hacia el procesamiento en órbita	60
Referencias Bibliográficas	61
Apéndices	63

Lista de tablas

1	Relación entre factores de escala y aplicaciones en sensores satelitales.	23
2	Distribución original de imágenes satelitales en el DeepGlobe Dataset.	32
3	Especificaciones del entorno de desarrollo y hardware.	45
4	Matriz de experimentación y variables de degradación orbital.	49
5	Resultados cuantitativos comparativos para EDSR y SwinIR bajo diferentes degradaciones (PSF Gaussiana y Direccional).	53
6	Comparativa de hiperparámetros de entrenamiento: EDSR vs. SwinIR.	72

Lista de figuras

1	Impacto visual de los núcleos de degradación suministrados	24
2	Arquitectura del pipeline para la generación de datos sintéticos y degradación multiescala.	31
3	Muestra representativa de la diversidad topográfica y urbana presente en el conjunto de datos DeepGlobe.	33
4	Pipeline de preparación y partición del dataset.	34
5	Efectos visuales del pipeline de degradación multiescala sobre parches del DeepGlobe.	36
6	Arquitectura del modelo EDSR evidenciando la estructura de bloques residuales sin capas de normalización.	39
7	Arquitectura del modelo SwinIR destacando el bloque RSTB y el mecanismo de atención por ventanas desplazadas.	41
8	Evaluación preliminar de convergencia del modelo EDSR frente a cuatro funciones de degradación distintas, utilizada para la selección metodológica de los casos de prueba.	44
9	Tablero comparativo de degradación multiescala y variabilidad radiométrica.	51
10	Evaluación del rendimiento sobre el conjunto de test bajo condiciones de degradación.	54
11	Arquitectura de la red neuronal SRCNN	68
12	Arquitectura propuesta del modelo RCAN.	69

13	Dinámica de optimización de la arquitectura convolucional (EDSR). Se contrasta la evolución de la función de pérdida frente al desenfoque gaussiano (a) y direccional (b)	74
14	Evolución de la pérdida del SwinIR para desenfoque gaussiano (a) y direccional (b).	75
15	Comparativa a diferentes condiciones para escalamiento x2	77
16	Comparativa a diferentes condiciones para escalamiento x4	79
17	Comparativa a diferentes condiciones para escalamiento x16	81

Lista de Apéndices

Apéndice A	Repositorio del Proyecto	64
Apéndice B	Demostración Matemática de la Degradación Óptica	65
Apéndice C	Evolución Histórica de la Superresolución	67
Apéndice D	Dinámicas y Configuración de Entrenamiento	70
Apéndice E	Análisis de Convergencia y Funciones de Pérdida	73
Apéndice F	Inspección Visual Multiescala	76

Glosario

CubeSat / Satélite Pequeño: plataforma satelital miniaturizada y estandarizada. Presenta costos reducidos en comparación con misiones satelitales de tamaño y complejidad mucho mayores. Su diseño permite despliegues en constelaciones, aunque impone restricciones severas en el tamaño de la óptica y, por consiguiente, en la resolución espacial.

DeepGlobe (Dataset): conjunto público de datos de imágenes satelitales ópticas RGB. Con una resolución nativa comercial de 50 [cm/píxel], actúa como la referencia para el aprendizaje de la red neuronal y la validación de los algoritmos de restauración espacial.

Factor de Escala: magnitud de submuestreo espacial que define la proporción entre la imagen de baja resolución y el objetivo de alta resolución. En este estudio, se emplean factores extremos (ej. x16) para simular matemáticamente las limitaciones ópticas de un satélite compacto.

Función de Dispersión de Punto (PSF): modelo matemático tridimensional que describe la respuesta y difracción de un sistema óptico al captar una fuente de luz puntual. En esta investigación, se emplea como método de degradación para agrupar matemáticamente las aberraciones y efectos de lente antes de la digitalización electrónica.

Ground Truth: en el aprendizaje supervisado, corresponde a la imagen original intacta y de alta fidelidad que se establece como la etiqueta o el objetivo geométrico perfecto que la red neuronal debe esforzarse por imitar durante su convergencia.

LEO (Low Earth Orbit): órbita Terrestre Baja. Región del espacio situada a una altitud de entre 160 y 2.000 kilómetros sobre la superficie terrestre. Es la órbita predilecta para las cons-

telaciones de satélites de observación del paradigma NewSpace, ya que permite una alta frecuencia de revisita. Sin embargo, exige altas velocidades espaciales (aproximadamente 7.5 [km/s]), lo que induce fenómenos de desenfoque por movimiento (Motion Blur) en los sensores integrados.

LIIF (Local Implicit Image Function): decodificador de representación continua de imágenes empleado en conjunto con modelos avanzados para compensar el método de degradación y reconstruir detalles en factores de escala arbitrarios.

Motion Blur (Desenfoque por movimiento): degradación óptica de carácter direccional ocasionada por la alta velocidad orbital transversal de la plataforma satelital en relación con la rotación terrestre, generada durante el tiempo de integración de luz del sensor.

Mecanismo de Atención:]capacidad de arquitecturas como los Vision Transformers para modelar dependencias a largo plazo y correlacionar información espacial en regiones distantes de una imagen, superando las limitaciones del campo receptivo local típico de las redes neuronales convolucionales.

NewSpace: paradigma contemporáneo de la industria aeroespacial caracterizado por el despliegue de constelaciones de satélites de menor tamaño, priorizando la ubicuidad y la frecuencia de revisita sobre la resolución óptica individual.

PSNR (Peak Signal-to-Noise Ratio): métrica cuantitativa expresada en decibelios (dB) que evalúa matemáticamente el error absoluto a nivel de píxel entre la imagen generada por la red neuronal y la imagen original.

Ruido blanco: componente de incertidumbre estadística aditiva (denotado como η) en la captura, proveniente del ruido de lectura, el ruido de disparo y la agitación térmica de los circuitos electrónicos en órbita.

SSIM (Structural Similarity Index Measure): métrica de calidad de imagen que modela la percepción visual humana. A diferencia del PSNR, evalúa específicamente la preservación de la

luminancia, el contraste y la continuidad topológica estructural en la imagen reconstruida.

Superresolución (SR): proceso computacional que infiere, sintetiza y reconstruye información de alta frecuencia espacial a partir de una o múltiples capturas degradadas de baja resolución, logrando compensar de manera algorítmica el impacto del submuestreo óptico.

SwinIR (Swin Image Restoration): arquitectura avanzada de aprendizaje profundo basada en Vision Transformers. Utiliza un mecanismo de partición de ventanas desplazadas (Shifted Windows) para capturar el contexto global de la imagen manteniendo una complejidad computacional eficiente, la cual escala de manera lineal (y no cuadrática) respecto al tamaño de la imagen de entrada.

SWaP (Size, Weight, and Power): acrónimo empleado en la ingeniería aeroespacial que hace referencia al Tamaño, Peso y Potencia. Representa las restricciones críticas de diseño en misiones espaciales. En el contexto de los satélites pequeños, las severas limitaciones de SWaP obligan a utilizar sistemas ópticos de menor apertura y electrónica de menor consumo, lo que compromete directamente la resolución espacial nativa y la relación señal-ruido de las imágenes capturadas.

Resumen

Título: Implementación de algoritmos de aprendizaje profundo para la superresolución de imágenes satelitales*

Autores: Juan Sebastian Jaramillo Diaz, Jonathan Mauricio Rojas Rodriguez **

Palabras Clave: Superresolución, Aprendizaje profundo, Imágenes satelitales, Vision Transformers, NewSpace, Redes Neuronales Convolucionales

Descripción:

El presente proyecto aborda las limitaciones de resolución espacial inherentes a las plataformas satelitales compactas, cuyas restricciones de diseño en tamaño, peso y potencia comprometen la calidad de los datos para análisis detallados en gestión territorial y monitoreo ambiental. Si bien existen misiones de alta resolución y satélites con capacidades multispectrales avanzadas, el despliegue masivo de constelaciones de bajo costo suele entregar imágenes con una resolución espacial insuficiente para tareas de ingeniería de precisión. Ante esta problemática, se propone el desarrollo y análisis de una solución basada exclusivamente en software, empleando algoritmos de aprendizaje profundo para la superresolución de imágenes ópticas. Esta técnica permite reconstruir texturas, bordes y patrones espaciales complejos, mitigando las restricciones físicas de los sistemas ópticos de pequeña apertura y superando las capacidades de los métodos tradicionales de procesamiento digital.

La investigación se estructura en cuatro fases metodológicas: el diseño de un protocolo de preprocesamiento de datos para la simulación de degradaciones orbitales reales, la adaptación de arquitecturas del estado del arte, el entrenamiento supervisado utilizando el lenguaje de programación Python y la evaluación cuantitativa de los resultados. Para garantizar la validez del modelo, se emplean métricas estandarizadas de calidad de imagen, específicamente PSNR y SSIM, comparando el desempeño de los algoritmos propuestos frente a métodos convencionales de interpolación bicúbica. Con este desarrollo, se busca fortalecer la autonomía técnica en el procesamiento de información geoespacial, permitiendo optimizar el aprovechamiento de datos de libre acceso y mejorar la precisión en la detección de infraestructura y coberturas terrestres a nivel global.

*Trabajo de Grado

**Facultad de Ingenierías Fisicomecánicas. Escuela de Ingenierías Eléctrica, Electrónica y de Telecomunicaciones. Director: Hans Yecid Garcia Arenas. Doctor en ingeniería. Codirector: Leandro Sebasthyan Rojas Rodríguez, Sebastian Ardila Leal. Magíster en ingeniería electrónica

Abstract

Title: Implementation of deep learning algorithms for satellite image super-resolution.***

Authors: Juan Sebastian Jaramillo Diaz, Jonathan Mauricio Rojas Rodriguez ****

Key Words: Super-resolution, Deep Learning, Satellite Imagery, Vision Transformers, NewSpace, Convolutional Neural Networks.

Description:

This project addresses the spatial resolution limitations inherent in compact satellite platforms, whose Size, Weight, and Power design constraints compromise data quality for detailed analysis in land management and environmental monitoring. While high-resolution missions and satellites with advanced multispectral capabilities exist, the massive deployment of low-cost constellations often delivers images with insufficient spatial resolution for precision engineering tasks. Given this problem, a software-based solution is proposed, employing deep learning algorithms for optical image super-resolution. This technique allows for the reconstruction of textures, edges, and complex spatial patterns, mitigating the physical constraints of small-aperture optical systems and surpassing the capabilities of traditional digital processing methods.

The research is structured into four methodological phases: the design of a data preprocessing protocol for the simulation of real orbital degradations, the adaptation of state-of-the-art architectures, supervised training using the Python programming language, and the quantitative evaluation of the results. To ensure model validity, standardized image quality metrics are employed, specifically PSNR and SSIM, comparing the performance of the proposed algorithms against conventional bicubic interpolation methods. This development aims to strengthen technical autonomy in geospatial information processing, enabling the optimization of open-access data utilization and improving precision in infrastructure detection and land cover monitoring on a global scale.

***Degree Work

****Facultad de Ingenierías Fisicomecánicas. Escuela de Ingenierías Eléctrica, Electrónica y de Telecomunicaciones. Director: Hans Yecid Garcia Arenas. Doctor of Engineering. Codirector: Leandro Sebasthyan Rojas Rodríguez, Sebastian Ardila Leal. Master's degree in electronic engineering

Introducción

La teledetección contemporánea se encuentra en la era del NewSpace, un paradigma caracterizado por el despliegue de constelaciones de satélites en órbita terrestre baja (Poghosyan y Golkar, 2017). Esta estrategia de misiones distribuidas permite una alta frecuencia de revisita sobre un mismo punto geográfico. Sin embargo, estas plataformas están sujetas a estrictas restricciones de diseño conocidas en la ingeniería aeroespacial como SWaP (Poghosyan y Golkar, 2017). La limitación en el volumen físico restringe la apertura del sistema óptico, y las limitaciones de potencia afectan la capacidad de los sensores y los tiempos de integración. Como resultado, mientras que los grandes satélites de observación alcanzan resoluciones espaciales submétricas, las plataformas compactas suelen capturar imágenes con resoluciones del orden de 3 a 5 [metros/píxel] (Demir et al., 2018).

El desarrollo de la superresolución ha buscado mitigar esta limitación de muestreo espacial mediante el aprendizaje profundo (Elad y Feuer, 1997). Los antecedentes en este campo inician con modelos convolucionales como SRCNN (Lim et al., 2017), demostrando la viabilidad de las redes neuronales para mapear características de baja a alta resolución. Posteriormente, la introducción de redes residuales profundas, específicamente el modelo EDSR (Lim et al., 2017), estableció un estándar de rendimiento. En la actualidad, el estado del arte incorpora arquitecturas basadas en Vision Transformers, como SwinIR (Liang et al., 2021), las cuales emplean mecanismos de atención por ventanas desplazadas para modelar dependencias espaciales complejas de manera más robusta que las convoluciones locales tradicionales.

A pesar de estos avances algorítmicos, gran parte de los modelos de SR se validan bajo submuestreos ideales, ignorando las aberraciones intrínsecas del sistema de captura satelital. En un

escenario real, la señal se ve corrompida por factores como el desenfoque por movimiento derivado de la velocidad orbital, variaciones en el enfoque por dilatación térmica de las lentes, y ruido en la electrónica del sensor (Demir et al., 2018). A partir de esta limitación metodológica, surge la pregunta de investigación: ¿En qué medida la implementación de una arquitectura de atención global supera a las redes convolucionales en la restauración de características espaciales de alta frecuencia, al procesar imágenes satelitales sometidas a degradaciones sintéticas que modelan matemáticamente el entorno físico orbital?

El objeto de estudio de esta investigación es el análisis comparativo del desempeño de los modelos EDSR y SwinIR en tareas de superresolución, bajo factores de escala operativos de $x2$, $x4$ y $x16$ (Liang et al., 2021). Para ello, se utiliza el conjunto de datos DeepGlobe como verdad base (Demir et al., 2018). El propósito es someter estas imágenes a un preprocesamiento que inyecta degradaciones sintéticas mediante una función de dispersión de punto modelada y la adición de ruido con el fin de evaluar la capacidad de las redes para revertir computacionalmente dicha pérdida de información. La hipótesis plantea que el modelo SwinIR, por su capacidad de procesar el contexto global, logrará una recuperación estructural superior frente al EDSR en los escenarios de degradación más severos, validado mediante métricas objetivas de PSNR y SSIM.

La justificación de este proyecto radica en su fundamentación técnica y su aplicabilidad directa. Técnicamente, aporta un marco de evaluación bajo ruidos que emulan la física del sensor, superando la degradación bicúbica estándar (Keys, 1981). A nivel de aplicación, propone compensar limitaciones ópticas mediante software, facilitando el análisis de infraestructuras, vegetación y desastres al optimizar la relación costo-beneficio en la adquisición de datos espaciales.

El desarrollo de este trabajo sigue la lógica de nuestra propuesta técnica. Partimos de un marco teórico y estado del arte que justifica la necesidad de herramientas para el procesamiento de imágenes satelitales (Chen et al., 2021). Con esa base, describimos cómo construimos el entorno de simulación para degradar las imágenes y cómo adaptamos los modelos para procesarlas. Analizamos cómo aprenden estas redes y qué resultados ofrecen, permitiéndonos contrastar su desempeño en condiciones reales de ruido y desenfoque. Finalmente, el estudio cierra demostrando

cómo estas arquitecturas permiten transformar capturas de baja fidelidad en productos con valor técnico, facilitando un monitoreo más preciso, permitiendo así una mayor autonomía técnica en el aprovechamiento de datos geoespaciales.

1. Objetivos

1.1 Objetivo general

Implementar un algoritmo de superresolución de imágenes satelitales basado en aprendizaje profundo, garantizando su validez a través de métricas estandarizadas de calidad y rendimiento.

1.2 Objetivos específicos

Diseñar un protocolo de preprocesamiento de datos que permita preparar la base de imágenes satelitales para el entrenamiento y validación del algoritmo.

Adaptar un algoritmo de superresolución del estado del arte, considerando criterios de precisión y rendimiento computacional.

Entrenar el algoritmo seleccionado en el lenguaje Python aplicando el protocolo de preprocesamiento definido.

Evaluar el desempeño del algoritmo, sobre la base de datos de imágenes preparadas, mediante métricas estandarizadas de calidad de imagen como PSNR, MSE y SSIM.

2. Marco teórico y estado del arte

Este capítulo establece los fundamentos científicos y técnicos de la investigación. Se expone la motivación operativa detrás del procesamiento de imágenes satelitales, el modelo matemático que describe su degradación desde la óptica de Fourier, y la evolución de las arquitecturas de aprendizaje profundo para resolver este problema.

2.1 Motivación operativa y el problema de la resolución

La superresolución de imagen (SR) se define como el proceso de estimar una imagen de alta resolución (I_{HR}) a partir de una única observación de baja resolución (I_{LR}). En la teledetección moderna, la resolución espacial es un factor crítico; mientras que los sistemas de alta precisión presentan costos elevados de despliegue, las constelaciones de satélites pequeños suelen sacrificar la calidad óptica debido a restricciones físicas de diseño.

Para que un modelo de SR sea aplicable, debe entrenarse con datos que reflejen el mundo físico. En esta investigación, se utilizan parches de imágenes satelitales ópticas RGB con una resolución nativa de 50 [cm/píxel], provenientes del conjunto DeepGlobe (Demir et al., 2018). Esta alta fidelidad permite simular la mejora de diferentes categorías de satélites en el mundo real mediante factores de escala matemáticos, como se detalla en la Tabla 1.

Tabla 1

Relación entre factores de escala y aplicaciones en sensores satelitales.

Escala	Sensor Origen	Resolución	Aplicación Técnica
x2	Multiespectral Media Res.	Alta Fidelidad	Homogenización Sentinel-2.
x4	Comercial Res. Media	Sub-métrica (VHR)	Mejora PlanetScope a WV-3.
x16	Nanosatélites / NewSpace	Alta Res. Urbana	Reconstrucción extrema.

2.2 Fundamentación física: Respuesta al impulso del sistema óptico

La formación de imágenes en un sistema satelital no es una transferencia directa y perfecta de la radiancia de la escena al plano del sensor. Físicamente, este proceso está gobernado por los fenómenos de difracción y las limitaciones de la apertura del sistema óptico. De acuerdo con la óptica de Fourier (Goodman, 2005), un sistema de adquisición puede modelarse como un filtro lineal e invariante en el espacio (LSI) que actúa sobre el campo electromagnético incidente.

La fundamentación matemática detallada sobre la propagación del campo, difracción en la pupila y definición de la PSF (Point Spread Function) mediante la óptica de Fourier se encuentra documentada en el Apéndice B.

2.2.1 Caracterización de núcleos de degradación suministrados

Para el desarrollo de esta investigación, se emplean núcleos de PSF caracterizados y suministrados por el grupo de investigación. Estos insumos representan perturbaciones reales del entorno orbital en plataformas NewSpace, permitiendo que el entrenamiento de los modelos de superresolución se realice bajo condiciones de degradación operativa:

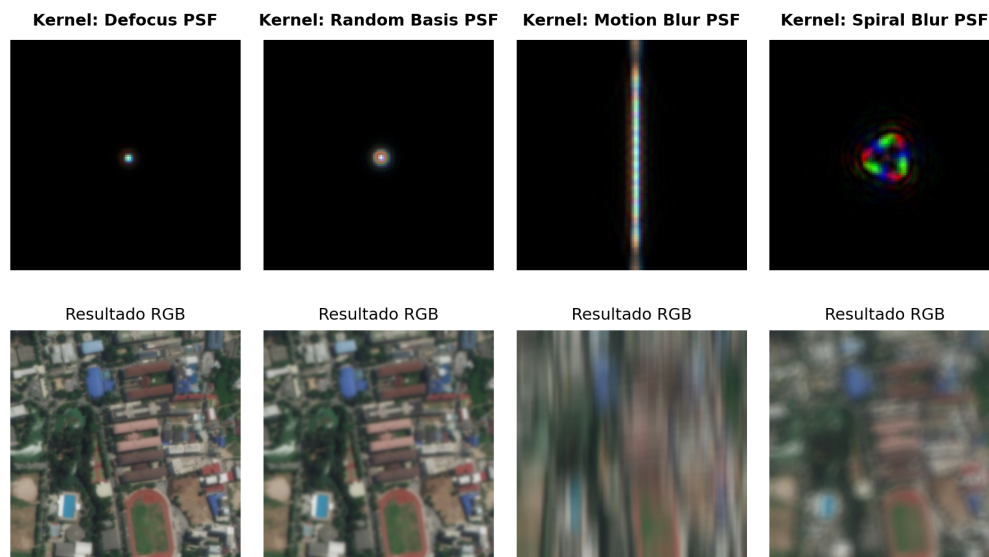
- **Desenfoque Óptico (Defocus):** Representa el desajuste del plano focal debido a la dilatación térmica de los elementos ópticos en el vacío. Físicamente, actúa como un filtro pasa-bajo que

atenúa las frecuencias espaciales más altas de la escena original.

- **Desenfoque por Movimiento (Motion Blur):** Insumo que integra el efecto de arrastre direccional resultante del desplazamiento relativo del satélite (aprox. 7.5 km/s) respecto a la superficie terrestre durante el tiempo de integración del sensor.
- **Desenfoque Vibracional (Spiral):** Núcleo que captura las micro-vibraciones mecánicas (jitter) provenientes de componentes internos como giroscopios y motores de reacción. Representa patrones de degradación no lineales que afectan la coherencia estructural.
- **Aberraciones Aleatorias (Random Basis):** Conjunto de núcleos estocásticos que representan la suma de distorsiones impredecibles, incluyendo turbulencias atmosféricas remanentes e imperfecciones microscópicas de manufactura.

Figura 1

Impacto visual de los núcleos de degradación suministrados



2.3 Modelado matemático de la degradación y formación de imágenes

Una vez establecida la naturaleza física de la PSF, el proceso completo de adquisición de una imagen satelital se describe mediante un modelo matemático de degradación lineal. Este modelo

actúa como el operador directo que transforma la escena real de alta resolución en la observación digital limitada que recibe el modelo de deep learning.

De acuerdo con el formalismo estándar en procesamiento digital de señales e imágenes, el flujo de degradación se define mediante la Ecuación 2.1:

$$I_{LR} = (I_{HR} * K) \downarrow_s + \eta \quad (2.1)$$

donde I_{LR} representa la observación de baja resolución obtenida por el sensor, I_{HR} es la radiancia teórica de la escena original o Ground Truth, K denota el núcleo de la PSF que actúa como operador de convolución para filtrar las altas frecuencias, \downarrow_s indica el operador de diezmado bajo un factor de escala s que modela el muestreo discreto del sensor, y η representa la incertidumbre estadística o ruido estocástico aditivo intrínseco al hardware electrónico.

2.3.1 Caracterización estadística del ruido estocástico (η)

Mientras que el núcleo K representa una degradación determinística ligada a la óptica, el término η introduce el componente estocástico o aleatorio al sistema. En el modelado de imágenes satelitales, este componente es fundamental para representar la incertidumbre inherente al proceso de cuantificación y transmisión electrónica.

El ruido η se parametriza generalmente como un ruido aditivo, compuesto por diversas fuentes físicas: el ruido de lectura (read noise), originado durante la conversión de carga a voltaje en el sensor; el ruido de disparo (shot noise), vinculado a la naturaleza discreta de los fotones; y el ruido, causado por la agitación de electrones en los circuitos a bajas temperaturas orbitales. Al integrar este término en la Ecuación 2.1, se garantiza que los modelos de aprendizaje profundo desarrollen una capacidad de discriminación robusta. Esto impide que la red neuronal amplifique artefactos electrónicos o interprete errores de muestreo como texturas reales de la superficie terrestre, un desafío crítico en la reconstrucción de señales de alta frecuencia.

2.3.2 *La superresolución como problema inverso*

Desde una perspectiva matemática, el proceso de transformar una imagen de alta resolución (I_{HR}) en una de baja resolución (I_{LR}) es directo y predecible. El verdadero desafío técnico surge al intentar realizar el camino inverso: la superresolución, que busca reconstruir una aproximación (\hat{I}_{HR}) contando únicamente con una observación limitada.

Este se clasifica como un problema mal condicionado, debido a que la pérdida de información durante el desenfoco y la reducción de tamaño genera una ambigüedad inherente. En términos simples, diferentes escenas reales de alta fidelidad podrían verse exactamente iguales al ser capturadas por un sensor de baja resolución, lo que hace imposible encontrar una solución única mediante métodos tradicionales. En este escenario, los modelos de Deep Learning funcionan como un regularizador inteligente; es decir, utilizan el conocimiento extraído de millones de ejemplos previos para decidir cuál es la reconstrucción más probable, permitiendo recuperar texturas y estructuras geográficas que la física del hardware no logró registrar.

2.4 Estado del Arte y evolución de la Superresolución (SR)

El campo de la Superresolución ha experimentado una transformación radical en la última década. El objetivo ha pasado de ser una simple suavización de píxeles a una reconstrucción inteligente de texturas y bordes, impulsada principalmente por el avance del aprendizaje profundo.

La evolución histórica detallada desde los métodos de interpolación clásica hasta los pioneros del aprendizaje profundo (SRCNN y RCAN) se encuentra documentada en el Apéndice C.

2.4.1 *La era de los Transformers en visión: SwinIR y modelos de atención global*

Aunque las redes convolucionales transformaron el campo de la superresolución, su diseño basado en filtros locales limita su capacidad para capturar relaciones de largo alcance dentro de una imagen. Las convoluciones operan sobre vecindades pequeñas de píxeles, lo que dificulta que la red comprenda la estructura global de objetos extensos, una limitación que se hace evidente en la teledetección al procesar grandes áreas geográficas. Para superar esta restricción, se han integrado arquitecturas basadas en Transformers, las cuales utilizan el mecanismo de auto-atención (Self-Attention) para modelar interdependencias entre parches distantes de la imagen, permitiendo una reconstrucción más coherente y rica en detalles.

Un hito fundamental en esta transición es la arquitectura SwinIR (Liang et al., 2021), la cual adapta el Swin Transformer para tareas de restauración de imágenes. La innovación de este modelo reside en el uso de ventanas locales desplazadas (Shifted Windows) para calcular la atención de manera jerárquica. Este enfoque permite que la red mantenga una eficiencia computacional comparable a las convoluciones, pero con la ventaja de que, al desplazar las ventanas entre capas sucesivas, la información puede fluir a través de los límites de cada parche. De este modo, SwinIR logra extraer características que consideran el contexto global de la escena, resultando en bordes mucho más definidos y una reducción significativa de artefactos en comparación con los métodos puramente convolucionales.

2.5 Fundamentos de optimización y funciones de pérdida en superresolución

El entrenamiento de modelos de superresolución se fundamenta en la minimización de una función de costo que cuantifica la discrepancia entre la reconstrucción estimada y la imagen de referencia. La selección de estas funciones es crítica, ya que determina qué atributos de la señal (bordes, texturas o valores de píxel) serán priorizados por el algoritmo durante el proceso de aprendizaje.

2.5.1 *Funciones de pérdida híbridas y preservación estructural*

En el procesamiento digital de imágenes, la función de pérdida más simple es el Error Cuadrático Medio (MSE); sin embargo, en tareas de superresolución, este enfoque tiende a producir resultados excesivamente suavizados al penalizar de forma desproporcionada las grandes desviaciones. Por esta razón, el marco teórico contemporáneo favorece el uso de funciones de pérdida híbridas que combinan diferentes dominios de la señal. Una de las variantes más robustas es la pérdida de Charbonnier, una aproximación diferenciable de la norma $L1$ que ofrece una mayor estabilidad numérica y una mejor preservación de los bordes al ser menos sensible a los valores atípicos.

Para complementar la reconstrucción espacial, se integran operadores de gradiente y análisis en el dominio de la frecuencia. Las funciones de pérdida de bordes, basadas en operadores diferenciales como Sobel o Laplacian, actúan como filtros pasa-alto que fuerzan al modelo a recuperar las altas frecuencias espaciales. Asimismo, la inclusión de pérdidas en el dominio de la frecuencia, calculadas mediante la Transformada Rápida de Fourier (FFT), asegura la coherencia espectral de la imagen. Este enfoque multiobjetivo se expresa matemáticamente como una suma ponderada:

$$\mathcal{L}_{total} = \alpha \mathcal{L}_{espacial} + \beta \mathcal{L}_{bordes} + \gamma \mathcal{L}_{frecuencia} \quad (2.2)$$

donde α , β y γ representan los coeficientes de ponderación que equilibran la importancia de cada dominio según los requerimientos de la aplicación final.

2.5.2 *Dinámica del aprendizaje y regularización estocástica*

La optimización de arquitecturas complejas como los Transformers requiere esquemas de tasa de aprendizaje que faciliten la exploración del espacio de búsqueda sin caer en mínimos locales

prematurados. La teoría de optimización moderna propone el uso de fases de calentamiento (warmup) seguidas de estrategias de decaimiento como el Cosine Annealing. Este último se fundamenta en reducir la tasa de aprendizaje siguiendo una curva cosenoidal, lo que permite realizar ajustes finos en los pesos a medida que el modelo se acerca a la convergencia.

Otro pilar teórico en la estabilidad de redes neuronales profundas es el Promedio Móvil Exponencial (EMA). Esta técnica consiste en mantener una copia de los pesos del modelo que se actualiza lentamente como un promedio de los estados anteriores. Desde el punto de vista de la generalización, el EMA actúa como un suavizado en el paisaje de pérdida (loss landscape), reduciendo el impacto de las fluctuaciones estocásticas del entrenamiento y produciendo un modelo final más robusto ante datos no vistos. Finalmente, el uso de precisión mixta (Mixed Precision) se justifica teóricamente como una optimización del flujo de tensores que aprovecha diferentes niveles de cuantificación para acelerar el cálculo sin perder la estabilidad numérica del gradiente.

2.6 Métricas de evaluación de calidad

Para cuantificar la fidelidad del mapeo, se emplean dos métricas estandarizadas que evalúan la reconstrucción desde perspectivas distintas: el error absoluto y la percepción estructural.

2.6.1 PSNR (*Peak Signal-to-Noise Ratio*)

Se basa en el Error Cuadrático Medio (MSE) y mide la potencia del error introducido por el proceso de superresolución frente a la señal máxima (Gonzalez y Woods, 2018). Se mide en decibelios (dB).

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I_{HR}(i, j) - I_{SR}(i, j)]^2 \quad (2.3)$$

$$PSNR = 10 \cdot \log_{10} \left(\frac{L^2}{MSE} \right) \quad (2.4)$$

Donde L representa el valor máximo dinámico de los píxeles (ej. $L = 255$ en imágenes de 8 bits) y m, n son las dimensiones espaciales.

2.6.2 SSIM (*Structural Similarity Index*)

A diferencia del PSNR, el SSIM modela la percepción del sistema visual humano (Wang et al., 2004). Mide la correlación de los patrones de píxeles evaluando tres términos: luminancia (l), contraste (c) y estructura (s).

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (2.5)$$

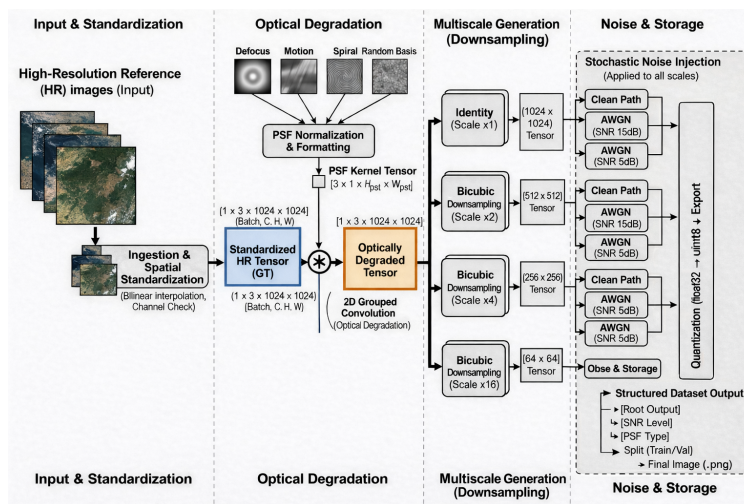
Donde μ representa las medias, σ^2 las varianzas, σ_{xy} la covarianza entre las ventanas x y y , y c_1, c_2 son constantes de estabilidad numérica. Un valor más cercano a 1 indica una mayor fidelidad estructural de los objetos geográficos reconstruidos.

3. Diseño del protocolo de preprocesamiento de datos

Tras comprender en el capítulo anterior las limitaciones ópticas de los satélites, el primer desafío técnico consiste en replicar estas fallas en un entorno controlado. La Figura 2 presenta el esquema general de nuestra metodología del preprocesamiento de datos, detallando las etapas y componentes técnicos específicos que conforman nuestro flujo de trabajo. A continuación, se describe cómo transformamos las imágenes RGB de alta resolución del dataset DeepGlobe para emular la degradación y el ruido, creando así el conjunto de datos fundamental que servira de base para nuestras redes neuronales.

Figura 2

Arquitectura del pipeline para la generación de datos sintéticos y degradación multiescala.



Nota. La figura describe el flujo computacional desde la entrada de imágenes en alta resolución hasta la exportación del dataset en formato .png.

3.1 Acondicionamiento del dataset base

Los datos utilizados en esta investigación provienen del conjunto público DeepGlobe Road Extraction Dataset (Demir et al., 2018). Este repositorio fue construido originalmente para el desarrollo de algoritmos de extracción de redes viales y destaca por poseer una resolución espacial nativa de [50 cm/píxel]. La Tabla 2 presenta la magnitud y distribución original de las instancias provistas por los autores antes de cualquier modificación.

Tabla 2

Distribución original de imágenes satelitales en el DeepGlobe Dataset.

Partición original	Cantidad de Imágenes	Contiene máscara de segmentación
Entrenamiento (Training)	6226	Sí
Validación (Validation)	1243	No
Prueba (Test)	1101	No
Total disponible	8570	

La selección de este conjunto se fundamenta en su alta fidelidad óptica y la enorme riqueza de sus texturas espaciales. Las capturas abarcan una gran variedad de entornos, desde intrincadas geometrías urbanas hasta densas zonas agrícolas y topografías rurales, lo cual resulta idóneo para entrenar la capacidad de generalización de un modelo de superresolución (véase la Figura 3).

Figura 3

Muestra representativa de la diversidad topográfica y urbana presente en el conjunto de datos DeepGlobe.



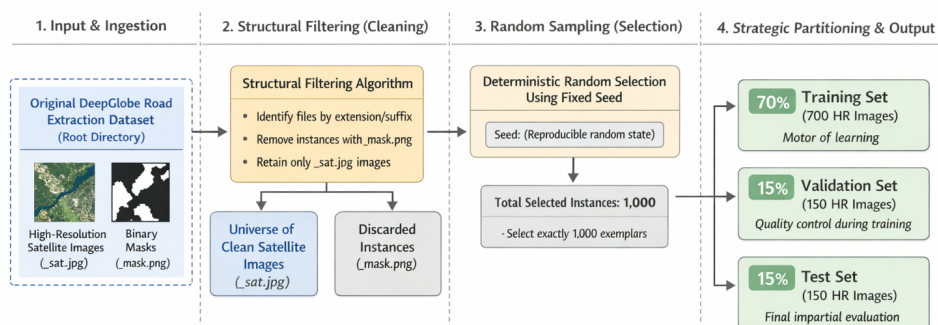
Como paso inicial en el preprocesamiento, se realizó una depuración estructural del repositorio. Dado que el objetivo de esta investigación es la reconstrucción de resolución y no la clasificación de píxeles, se eliminaron todas las máscaras binarias de anotación vial. De esta forma, las imágenes RGB originales quedaron establecidas de manera exclusiva como la verdad de campo o imágenes de Alta Resolución (I_{HR}).

Posteriormente, se procedió a delimitar el volumen de trabajo. Debido a la masiva carga computacional que exige el entrenamiento de redes neuronales profundas sobre imágenes de 1024×1024 píxeles, se seleccionó aleatoriamente una muestra representativa de 1000 imágenes. Buscando un equilibrio óptimo entre representatividad estadística y viabilidad técnica en memoria VRAM,

esta muestra se segmentó mediante una partición estricta: el 70 % (700 imágenes) se destinó para el entrenamiento de los pesos sinápticos; el 15 % (150 imágenes) para el conjunto de validación, actuando como control intermedio; y el 15 % restante (150 imágenes) se aisló totalmente para las pruebas de evaluación final.

Figura 4

Pipeline de preparación y partición del dataset.



3.2 Degradación multiescala

El preprocesamiento constituye la fase crítica para garantizar la robustez del modelo de aprendizaje profundo. Habiendo definido el modelo matemático unificado en el Capítulo 1, en esta sección se detalla su implementación. A diferencia de los protocolos de entrenamiento convencionales que validan sus modelos bajo submuestreos ideales, nuestro enfoque integra una degradación física multiescala que emula las aberraciones de un sistema óptico real.

Se diseñó un pipeline de degradación sintética que transforma las imágenes de alta resolución (HR) en tensores de baja resolución (LR) mediante cuatro etapas secuenciales:

- 1. Acondicionamiento del Tensor Base:** El flujo inicia garantizando la consistencia del formato RGB. Se aplica una estandarización mediante interpolación para obtener un tensor HR con dimensiones fijas de 1024×1024 píxeles. Los datos se convierten a precisión de punto flotante (float32) y se normalizan al rango radiométrico $[0, 1]$, asegurando la estabilidad numérica

para las operaciones en la GPU.

2. **Convolución Óptica (PSF):** Para emular las limitaciones físicas, se implementa una operación de convolución 2D agrupada. Se aplican los diferentes núcleos de degradación (Defocus, Motion, Spiral y Random Basis) a cada uno de los canales de la imagen RGB. Es imperativo que cada kernel sea previamente normalizado ($\sum PSF = 1$) para preservar la energía total de la escena y no alterar el brillo global de la imagen.
3. **Submuestreo Espacial (Downsampling):** Tras obtener el tensor degradado ópticamente, se utiliza una interpolación bicúbica para reducir las dimensiones espaciales. Se generan cuatro rutas de datos diferenciadas: una escala de identidad ($\times 1$) para preservación de rasgos, y tres escalas de reducción que simulan distintas plataformas satelitales a $\times 2$ (512×512), $\times 4$ (256×256) y $\times 16$ (64×64).
4. **Inyección Estocástica y Exportación:** Como etapa final, se aplica una inyección de Ruido Gaussiano Blanco Aditivo (AWGN) para emular las interferencias electrónicas. Se evalúan tres niveles por cada escala: una ruta limpia sin ruido agregado, y dos rutas con ruido controlado a relaciones señal-ruido (SNR) de 15 dB y 5 dB. El conjunto final se cuantiza y se exporta estructuradamente en formato `.png`.

3.3 Validación y resultados del protocolo de degradación

Para verificar la eficacia del pipeline computacional y asegurar que los tensores resultantes imitan fielmente las limitaciones de los sensores satelitales, se realizó una inspección cualitativa y estructural sobre las muestras exportadas.

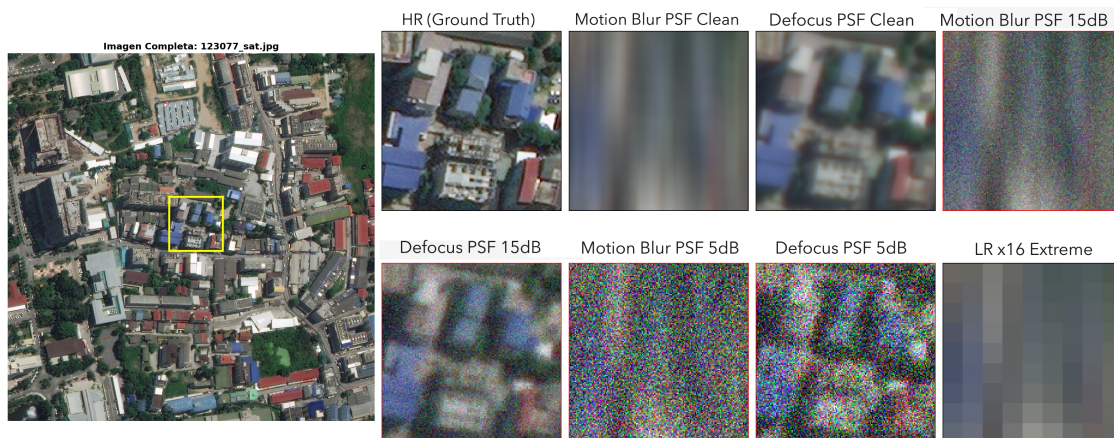
3.3.1 Inspección visual multiescala

La reducción espacial mediante interpolación bicúbica, combinada con las funciones de dispersión de punto (PSF), generó una pérdida de información de alta frecuencia matemáticamente coherente. Como se evidencia en la Figura 5, los bordes geométricos de las vías y las texturas complejas de los tejados (presentes en la verdad de campo I_{HR}) sufren un severo suavizado direccional y esférico a medida que aumenta el factor de escala (s).

En la escala $\times 4$, la huella urbana aún es vagamente distinguible, simulando la captura de un satélite comercial de resolución media. Sin embargo, en la reducción extrema a $\times 16$, los píxeles colapsan en bloques homogéneos, provocando fenómenos de aliasing y pérdida total de la firma estructural de las carreteras.

Figura 5

Efectos visuales del pipeline de degradación multiescala sobre parches del DeepGlobe.



Nota. La matriz ilustra un parche de alta resolución (I_{HR}) a la izquierda, y su evolución al ser sometido a diferentes núcleos PSF.

3.4 Análisis del Impacto Estocástico (SNR)

Adicional a la pérdida de resolución espacial, la inyección del Ruido Gaussiano Blanco Aditivo (AWGN) validó la robustez del entorno de pruebas.

A un nivel de relación señal-ruido de 15 dB, las interferencias electrónicas son perceptibles pero permiten que la topología general de la imagen sobreviva. No obstante, al someter los tensores al nivel más agresivo de 5 dB, la varianza del ruido enmascara casi por completo la varianza estructural de la imagen. Este escenario de degradación extrema (escala $\times 16$ con SNR de 5 dB) establece un límite crítico de información; reconstruir bordes viales a partir de este punto requiere que las arquitecturas de aprendizaje profundo (EDSR y SwinIR) no solo interpolen colores, sino que infieran y generen geometrías a partir de un contexto global severamente corrompido.

4. Selección y Arquitectura de los Modelos de Superresolución

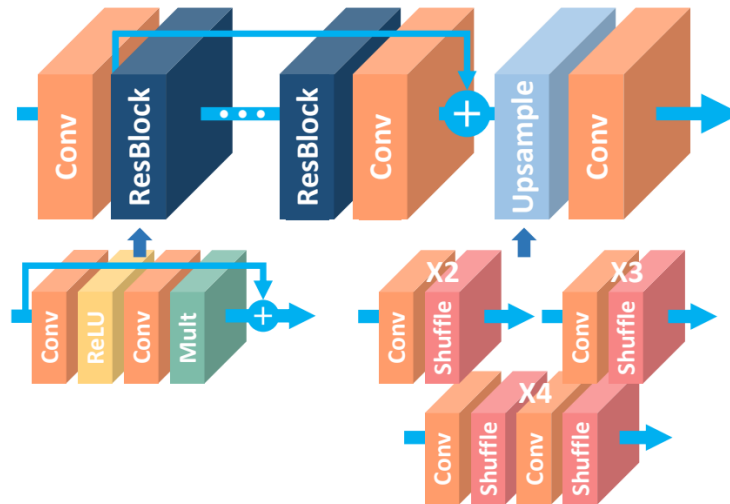
Con los datos sintéticos ya preparados y caracterizados, el siguiente paso lógico es construir el motor matemático capaz de revertir esta degradación. Este capítulo detalla la estructura del EDSR, que servirá como nuestra línea base convolucional, y del SwinIR, un modelo avanzado basado en Transformers que busca superar las limitaciones clásicas.

4.1 Arquitectura Base: Enhanced Deep Residual Networks (EDSR)

Para establecer un punto de comparación riguroso, se selecciona el modelo EDSR como línea base representativa de las Redes Neuronales Convolucionales (CNNs). La innovación central de esta arquitectura es la supresión total de las capas de Normalización por Lotes (Batch Normalization o BN) en sus bloques residuales (Lim et al., 2017).

Figura 6

Arquitectura del modelo EDSR evidenciando la estructura de bloques residuales sin capas de normalización.



En la teledetección, aplicar BN resulta contraproducente porque altera artificialmente la distribución de colores y la luminancia absoluta, destruyendo la firma radiométrica original de la imagen satelital. Al eliminar estas capas, el EDSR no solo preserva el contraste físico real del terreno, sino que reduce el consumo de memoria en aproximadamente un 40 %. Esta eficiencia permite construir una red más profunda que, mediante el uso exclusivo de operaciones de convolución (3×3) y activaciones ReLU, extrae características topográficas de alta frecuencia. Los resultados de este modelo establecerán el "piso de rendimiento" frente a aberraciones como el desenfoque óptico y el ruido del lente.

4.2 Arquitectura Avanzada: Swin Image Restoration (SwinIR)

Para superar las limitaciones locales de las CNNs, se implementa SwinIR, un modelo avanzado basado en Vision Transformers (ViT) diseñado para la restauración de imágenes, el cual es capaz de modelar dependencias espaciales de largo alcance (Liang et al., 2021). A diferencia de las convoluciones clásicas que procesan parches de información de forma aislada, este enfoque resulta

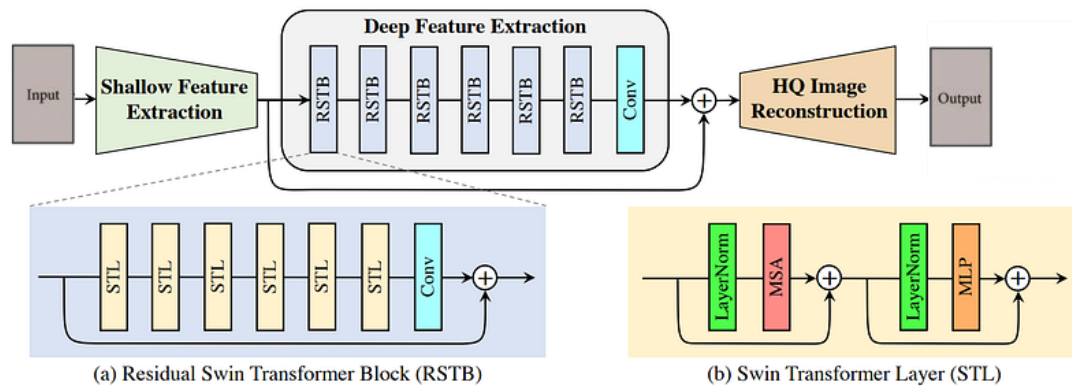
crucial en la teledetección, donde estructuras continuas como redes viales, complejos industriales o ríos requieren de un análisis global del contexto visual para ser reconstruidas fielmente tras la degradación orbital.

La arquitectura de este modelo se articula algorítmicamente en tres módulos secuenciales:

1. **Extracción Superficial (Shallow Feature Extraction):** El proceso inicia utilizando una capa convolucional estándar que actúa como puente. Su función es estabilizar el mapeo inicial de los píxeles de baja resolución hacia un espacio de características de alta dimensión, preservando la coherencia espacial y estructural básica antes de someter los datos a las capas de atención.
2. **Extracción Profunda (Deep Feature Extraction):** Constituye el núcleo analítico de la red y está compuesto por una serie de bloques RSTB (Residual Swin Transformer Blocks). En esta etapa, la arquitectura abandona la convolución tradicional y la sustituye por mecanismos de auto-atención anidados en conexiones residuales. Esta estructura garantiza que los gradientes matemáticos fluyan eficientemente a lo largo de la red durante el entrenamiento, evitando la pérdida de información topográfica de baja frecuencia.
3. **Módulo de Reconstrucción de Alta Calidad:** Tras refinar las características latentes, la red emplea la técnica de Sub-pixel Convolution (también conocida como Pixel Shuffle) para proyectar la información nuevamente al espacio de los píxeles en la resolución objetivo dictada por la escala ($x2, x4, x16$). Finalmente, una conexión de salto global suma la imagen original interpolada al residuo generado por el Transformer, mitigando los típicos artefactos visuales de "damero".

Figura 7

Arquitectura del modelo SwinIR destacando el bloque RSTB y el mecanismo de atención por ventanas desplazadas.



El avance matemático más significativo que viabiliza esta topología es su mecanismo de Atención por Ventanas Desplazadas (Shifted Window Attention). En un Transformer convencional, calcular la interacción matemática de cada píxel contra todos los demás en una imagen satelital densa (1024×1024) resulta computacionalmente prohibitivo. SwinIR mitiga este obstáculo dividiendo la captura satelital en ventanas locales discretas (por ejemplo, matrices de 8×8 píxeles) y calculando la atención internamente. Para evitar que la información quede aislada dentro de cada bloque, la red desplaza sistemáticamente la cuadrícula de partición en las capas sucesivas. Este desplazamiento estratégico obliga a que las nuevas ventanas crucen los límites previos, permitiendo que la información geométrica viaje transversalmente por toda la imagen con un costo computacional lineal. Es esta capacidad de conectar lógicamente extremos opuestos de la imagen lo que otorga al SwinIR la superioridad para recuperar geometrías complejas que el EDSR no logra resolver.

4.3 Justificación del Análisis Comparativo

La selección de EDSR y SwinIR no obedece únicamente a una progresión histórica en la visión artificial, sino a la necesidad de evaluar un compromiso crítico (trade-off) en la teledetección

operativa: el balance entre el costo computacional y la fidelidad de reconstrucción.

En un escenario real de monitoreo satelital, donde constelaciones como PlanetScope generan terabytes de información diaria, la velocidad de inferencia y el consumo de memoria son factores determinantes. En este aspecto, el modelo EDSR ostenta una superioridad teórica en términos de eficiencia. Al basarse en operaciones convolucionales estrictamente locales y haber eliminado la normalización por lotes, su arquitectura es ligera y rápida. Es capaz de procesar grandes volúmenes de datos con una huella de memoria reducida, ofreciendo una mejora sustancial en los bordes básicos de las imágenes de baja resolución.

Por el contrario, el modelo SwinIR representa el extremo opuesto del espectro. Su superioridad teórica radica en la fidelidad topográfica. Al utilizar la atención por ventanas desplazadas, el modelo posee una comprensión global del contexto, lo que le permite discernir con mucha mayor precisión entre el ruido estocástico y las verdaderas texturas de alta frecuencia (como el asfalto o las geometrías arquitectónicas). Sin embargo, esta capacidad de razonamiento espacial^{exige} un costo computacional exponencialmente mayor durante el entrenamiento y una demanda de memoria significativamente más alta en la etapa de inferencia.

Por lo tanto, la justificación de este análisis comparativo radica en cuantificar si la brecha tecnológica justifica el salto de paradigma. El objetivo empírico de esta investigación es determinar si el incremento en calidad visual y métricas de error (PSNR y SSIM) que promete el SwinIR frente a perturbaciones físicas complejas (como la PSF asimétrica y el Motion Blur) es lo suficientemente significativo como para justificar el sacrificio en eficiencia computacional que representa abandonar la línea base convolucional del EDSR.

5. Entrenamiento de los Modelos

Una vez definidas las arquitecturas de red, es necesario someterlas a un proceso de aprendizaje iterativo. Este capítulo describe la dinámica empírica de entrenamiento en Python, detallando la configuración de los hiperparámetros, las funciones de pérdida (L1/L2) y cómo ambos modelos ajustaron sus pesos al interactuar con el protocolo de degradación.

5.1 Evaluación Preliminar de las Funciones de Degradación

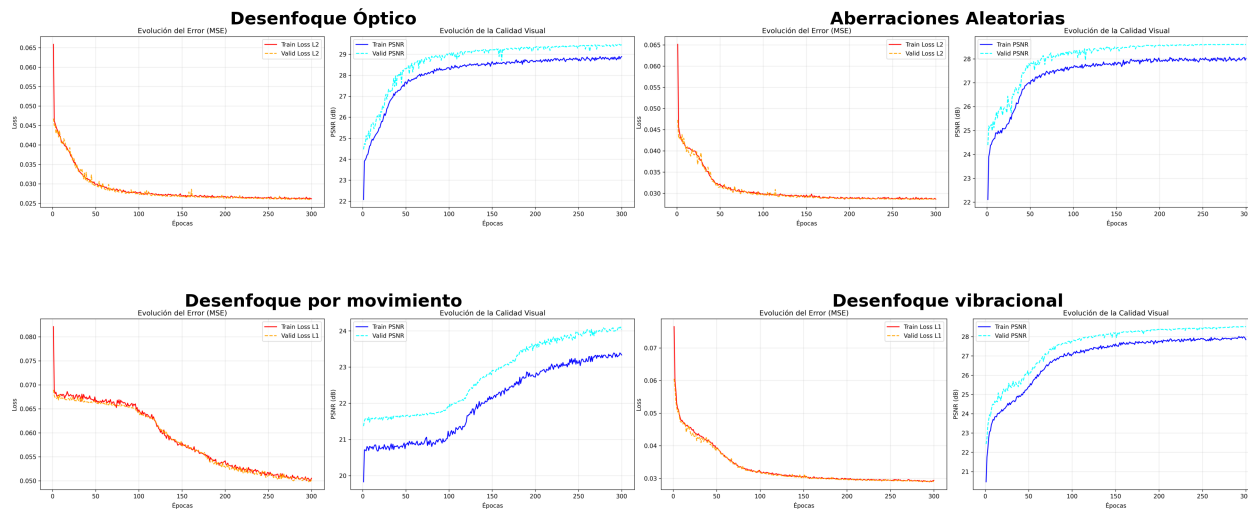
Durante la fase exploratoria de la investigación, la arquitectura base convolucional (EDSR) fue sometida a un entrenamiento preliminar frente a las cuatro funciones de dispersión de punto (K) modeladas en el marco teórico: desenfoque gaussiano (Focus), desenfoque de barrido (Línea), desenfoque vibracional (Espiral) y aberraciones aleatorias.

Como se documenta en la Figura 8, se evaluó la evolución del error y la calidad visual para cada perturbación. Con el objetivo de optimizar los recursos computacionales y establecer un análisis comparativo riguroso con el modelo avanzado (SwinIR), se determinó acotar la experimentación final exclusivamente a los dos escenarios que representaron los límites operativos del sistema.

En consecuencia, se seleccionó el kernel que permitió la convergencia más estable y el mayor rendimiento de reconstrucción (desenfoque gaussiano o Focus), y el kernel que indujo el peor escenario de convergencia y mayor inestabilidad (desenfoque de barrido o Línea). Esta selección garantiza que la comparativa posterior evalúe las cotas superior e inferior del desempeño del pipeline propuesto.

Figura 8

Evaluación preliminar de convergencia del modelo EDSR frente a cuatro funciones de degradación distintas, utilizada para la selección metodológica de los casos de prueba.



5.2 Configuración del Entorno y Hardware

Con las arquitecturas algorítmicas definidas y el protocolo de simulación del núcleo K establecido, el desarrollo teórico converge en la ejecución empírica. Este capítulo documenta el proceso iterativo de aprendizaje supervisado, detallando cómo cada modelo interactúa con el conjunto de datos para ajustar sus pesos internos. A continuación, se especifican los recursos de hardware, se justifica el diseño de una función de pérdida híbrida unificada, y se desglosan las estrategias de optimización específicas que demandan la arquitectura convolucional (EDSR) frente a la complejidad topológica del Transformer (SwinIR).

El procesamiento algorítmico de imágenes submétricas bajo factores de escala extremos (hasta $\times 16$) exige una gestión estricta de la memoria y alto rendimiento de cómputo paralelo. El entorno se configuró bajo las siguientes especificaciones técnicas:

- **Framework:** Desarrollo en Python utilizando la biblioteca PyTorch (Paszke et al., 2019),

aprovechando su generación dinámica de grafos.

- **Hardware de Cómputo:** Entrenamiento acelerado mediante una GPU NVIDIA GeForce RTX 4070 Ti Super (16 GB VRAM).
- **Gestión de Datos:** Implementación de DataLoaders con multiprocesamiento (`num_workers=2` y memoria anclada `pin_memory`) para recortes espaciales aleatorios (`crops`) de 512×512 píxeles, aplicando un riguroso aumento de datos (rotaciones y variaciones de intensidad) para prevenir el sobreajuste.

Tabla 3

Especificaciones del entorno de desarrollo y hardware.

Componente / Recurso	Especificación Técnica
Unidad de Procesamiento Gráfico (GPU)	NVIDIA GeForce RTX 4070 Ti Super
Memoria de Video (VRAM)	16 GB GDDR6X
Lenguaje de Programación	Python 3.10
Framework de Aprendizaje Profundo	PyTorch (con soporte CUDA)
Aceleración de Precisión Mixta	Habilitada (<code>torch.amp.autocast</code>)

5.3 Diseño de las Funciones de Pérdida Híbridas

Para garantizar la fidelidad topográfica de los objetos geográficos y evitar la difuminación de texturas, se descartó el uso exclusivo del Error Cuadrático Medio (L2), dada su tendencia a promediar píxeles y generar resultados suavizados. En su lugar, se diseñaron funciones de pérdida híbridas (CombinedLoss) adaptadas a la naturaleza matemática de cada arquitectura, buscando un equilibrio entre la precisión del píxel y la continuidad geométrica de los elementos espaciales.

5.3.1 *Función de pérdida para la arquitectura convolucional (EDSR)*

Para el modelo base (EDSR), se implementó una función bidimensional estandarizada orientada a la preservación espacial directa. El término principal de reconstrucción paramétrica recae sobre el Error Absoluto Medio (L1). A diferencia del L2, el L1 penaliza las diferencias de forma lineal, lo que permite mantener la nitidez global de las altas frecuencias y ha demostrado ser significativamente menos sensible a los valores atípicos introducidos por el ruido.

Para complementar esta precisión a nivel de píxel, se integró una segunda componente denominada penalización de bordes (Edge Loss). Esta penalización hace uso de operadores convolucionales espaciales, específicamente filtros de Sobel direccionales, para extraer los gradientes horizontales y verticales tanto de la predicción como de la imagen original. Al calcular la divergencia entre ambos gradientes, ponderada con un multiplicador de importancia $\alpha = 0,2$, se asegura que el modelo convolucional no solo converja hacia el valor de luminancia correcto, sino que recupere fielmente la geometría y estructura de las topografías analizadas.

5.3.2 *Función de pérdida para el modelo de atención (SwinIR)*

Por otro lado, la arquitectura basada en mecanismos de atención (SwinIR) requiere restricciones matemáticas más complejas. Dado que los Vision Transformers procesan la imagen mediante la partición en ventanas (Windows), es fundamental evitar la formación de artefactos de cuadrícula durante la reconstrucción. Por ello, la función de pérdida se extendió a tres componentes para estabilizar el entrenamiento profundo. En primer lugar, se reemplazó el L1 tradicional por la función de pérdida de Charbonnier, configurada con un peso de $\lambda_1 = 0,7$ y una constante de estabilización numérica $\epsilon = 10^{-3}$. Al ser una aproximación diferenciable y continua del error absoluto, esta función facilita la propagación ininterrumpida de gradientes a través de las capas de auto-atención.

En segundo lugar, se mantuvo la penalización de bordes (Edge Loss) mediante filtros de

Sobel con un peso de $\lambda_2 = 0, 2$. No obstante, dado que esta arquitectura avanzada utiliza precisión mixta automática y escalado de gradientes para optimizar el consumo de memoria en la GPU, la implementación de esta componente se condicionó para forzar el cálculo topográfico estrictamente en formato de punto flotante de 32 bits (`float32`). Esta medida preventiva es crucial para evitar desbordamientos o colapsos numéricos durante la optimización.

Finalmente, para contrarrestar la severa degradación óptica impuesta por la PSF, se introdujo una penalización espectral en el dominio de la frecuencia (Frequency Loss) con un peso de $\lambda_3 = 0, 1$. Esta métrica aplica la Transformada Rápida de Fourier (FFT) bidimensional ortogonal a los tensores, evaluando el error absoluto entre las magnitudes frecuenciales de la predicción y la imagen original. El objetivo físico de esta tercera componente es obligar a la arquitectura SwinIR a recuperar con máxima exactitud las texturas más finas que usualmente desaparecen bajo el núcleo de degradación óptico.

5.4 Dinámicas de Entrenamiento

Los detalles sobre la configuración de hiperparámetros, optimizadores, planificadores de tasa de aprendizaje y dinámicas específicas de entrenamiento para ambas arquitecturas se encuentran documentados en el Apéndice D.

5.5 Monitoreo de Convergencia y Prevención de Sobreajuste

El entrenamiento de arquitecturas profundas frente a datos altamente ruidosos (*PSF_5*) conlleva un riesgo inminente de sobreajuste (overfitting), donde la red memoriza el ruido en lugar de aprender la estructura subyacente. Para mitigar esto, se implementó un protocolo de Parada Temprana (Early Stopping).

Durante el ciclo de entrenamiento, el modelo es evaluado constantemente contra un conjunto de validación independiente. El algoritmo monitorea la métrica de calidad (PSNR); si esta no

presenta una mejora superior a un delta de 0,005 dB durante una ventana de paciencia de 40 épocas consecutivas, el entrenamiento se interrumpe automáticamente. Además, el sistema guarda exclusivamente los pesos de la época que alcanzó el rendimiento máximo, garantizando que el modelo final sea la versión con mayor capacidad de generalización topográfica.

5.6 Protocolo de Evaluación Cuantitativa

Para cuantificar la fidelidad de la reconstrucción y comparar objetivamente el rendimiento entre la línea base (EDSR) y el modelo avanzado (SwinIR), se definieron dos métricas estandarizadas en la literatura de procesamiento digital de imágenes, las cuales se calculan sobre el espacio de color RGB:

1. **Relación Señal a Ruido Máxima (PSNR):** Evalúa el error absoluto a nivel de píxel entre la predicción y la imagen original (ground truth). Aunque es susceptible a pequeñas variaciones de brillo, un PSNR alto (medido en decibelios, dB) es el primer indicador de que la red logró revertir la degradación matemática del núcleo K .
2. **Índice de Similitud Estructural (SSIM):** A diferencia del PSNR, el SSIM modela la percepción visual humana. Mide tres componentes fundamentales de la topografía satelital: la luminancia, el contraste y, lo más importante, la correlación estructural (los bordes de las vías y edificios). Un valor de SSIM cercano a 1,0 certifica que la red no solo coloreó correctamente la imagen, sino que recuperó la geometría original.

5.7 Consideraciones de Complejidad Computacional y Memoria

El análisis empírico del entrenamiento reveló el costo real de implementar mecanismos de atención global en la teledetección. Mientras que la arquitectura EDSR procesa lotes de imágenes con un consumo de VRAM moderado, el modelo SwinIR+LIIF satura rápidamente los 16 GB de la GPU (RTX 4070 Ti Super).

Para viabilizar la convergencia del Transformer bajo el factor de escala extremo de $\times 16$, fue necesario implementar Lotes Adaptativos (Adaptive Batching). Para resoluciones de entrada muy pequeñas (32×32 píxeles), el tamaño del lote se mantuvo en 4 imágenes. Sin embargo, a medida que el tamaño de entrada aumentaba para escalas menores (ej. $\times 2$, con parches de 256×256), el lote debió reducirse a una única imagen por iteración. Este análisis de complejidad será fundamental en el siguiente capítulo para determinar si la mejora métrica del SwinIR justifica su exigencia de hardware en un escenario de monitoreo satelital real.

Tabla 4

Matriz de experimentación y variables de degradación orbital.

Variable Física Simulada	Niveles de Perturbación Evaluados
Factor de Escala (Submuestreo)	$\times 2$, $\times 4$, $\times 16$ (Satélites)
Niveles de Ruido Térmico	Limpio (Clean), Leve (15 dB), Extremo (5 dB)
Función de Dispersión de Punto (PSF)	Desenfoque Óptico (Focus), Desenfoque por Movimiento (Línea)

6. Evaluación del Desempeño y Validación Experimental

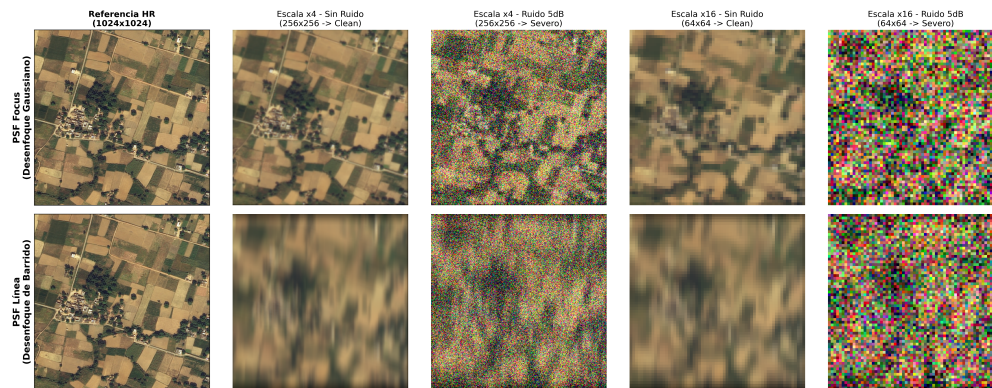
El propósito de esta fase experimental es determinar el rendimiento objetivo de las arquitecturas propuestas. El presente capítulo contrasta el desempeño de la línea base convolucional (EDSR) frente al modelo avanzado (SwinIR), cuantificando la fidelidad en la recuperación de texturas satelitales a través de métricas de calidad estandarizadas frente a distintos niveles de degradación estocástica.

6.1 Validación del Protocolo de Preparación de Datos

Los resultados del primer objetivo evidencian la eficacia del entorno de simulación para replicar la degradación multifactorial de las imágenes satelitales. La Figura 9 detalla el impacto combinado del submuestreo drástico (x4, x16) y las aberraciones ópticas del sensor. Específicamente, la inyección de ruido severo (SNR de 5 dB) provoca una pérdida crítica de la fidelidad espectral y espacial de los píxeles. Este nivel de destrucción visual corrobora la exigencia de implementar modelos avanzados de superresolución, los cuales son estrictamente necesarios para inferir la estructura original a partir de entradas severamente degradadas.

Figura 9

Tablero comparativo de degradación multiescala y variabilidad radiométrica.



6.2 Análisis Comparativo EDSR vs. SwinIR

6.2.1 Convergencia y Funciones de Pérdida

El análisis detallado de la convergencia, la estabilidad numérica y la evolución de las funciones de pérdida para ambas arquitecturas frente a los distintos núcleos de degradación se encuentra documentado en el Apéndice E. En resumen, el modelo SwinIR demostró una convergencia notablemente más estable que el EDSR frente a ruidos severos.

6.2.2 Evaluación Cuantitativa

En esta sección se analiza el desempeño numérico de las arquitecturas implementadas mediante el uso de las métricas estandarizadas PSNR y SSIM. El objetivo es determinar la capacidad de reconstrucción de la red vial frente a la degradación sintética que emula las condiciones de un nanosatélite en órbita.

La evaluación se realiza sobre el conjunto de prueba del dataset DeepGlobe, contrastando la línea base convolucional (EDSR) frente a la arquitectura de atención global (SwinIR). Los

resultados se desglosan en función del factor de escala ($x2$, $x4$, $x16$) y el nivel de ruido inyectado (Limpio, 5 dB y 15 dB). Este análisis permite cuantificar no solo la precisión de los modelos, sino también su estabilidad estadística (media y desviación estándar) ante condiciones de baja señal-ruido.

Como se observa en la Tabla 5, existe una correlación directa entre el incremento del ruido y la degradación de las métricas, siendo el escenario de $x16$ con 5 dB el punto crítico de operación. En este nivel, la superioridad del SwinIR se vuelve estadísticamente significativa, logrando preservar la continuidad de los bordes donde el EDSR presenta un colapso estructural.

Tabla 5

Resultados cuantitativos comparativos para EDSR y SwinIR bajo diferentes degradaciones (PSF Gaussiana y Direccional).

PSF	Ruido	Escala	PSNR (dB)		SSIM	
			EDSR	SwinIR	EDSR	SwinIR
Gaussiano	Limpio	x2	32.69 ± 2.34	32.87 ± 2.38	0.8132 ± 0.0402	0.8151 ± 0.0397
		x4	29.13 ± 3.05	29.30 ± 3.02	0.6943 ± 0.0756	0.7001 ± 0.0733
		x16	23.69 ± 3.64	23.88 ± 3.64	0.5619 ± 0.1245	0.5647 ± 0.1235
	15 dB	x2	25.50 ± 3.09	26.51 ± 2.99	0.6012 ± 0.1008	0.6199 ± 0.0910
		x4	24.53 ± 3.19	25.05 ± 3.15	0.5786 ± 0.1120	0.5873 ± 0.1072
		x16	22.14 ± 3.39	22.16 ± 3.37	0.5500 ± 0.1272	0.5500 ± 0.1269
	5 dB	x2	23.72 ± 3.23	24.21 ± 3.15	0.5676 ± 0.1175	0.5671 ± 0.1108
		x4	22.88 ± 3.26	22.88 ± 3.30	0.5579 ± 0.1229	0.5504 ± 0.1220
		x16	20.84 ± 3.23	20.83 ± 3.22	0.5433 ± 0.1288	0.5411 ± 0.1282
Direccional	Limpio	x2	23.60 ± 3.12	26.01 ± 2.86	0.5954 ± 0.1031	0.6434 ± 0.0838
		x4	25.12 ± 3.08	25.90 ± 3.09	0.6075 ± 0.1001	0.6187 ± 0.0968
		x16	22.83 ± 3.57	22.93 ± 3.56	0.5541 ± 0.1266	0.5552 ± 0.1263
	15 dB	x2	21.76 ± 3.41	22.31 ± 3.45	0.5488 ± 0.1270	0.5512 ± 0.1251
		x4	21.83 ± 3.43	21.98 ± 3.43	0.5492 ± 0.1273	0.5484 ± 0.1266
		x16	21.07 ± 3.37	21.08 ± 3.38	0.5442 ± 0.1292	0.5436 ± 0.1290
	5 dB	x2	21.33 ± 3.39	21.59 ± 3.42	0.5452 ± 0.1283	0.5385 ± 0.1264
		x4	21.11 ± 3.34	21.25 ± 3.39	0.5447 ± 0.1285	0.5381 ± 0.1275
		x16	20.29 ± 3.18	20.31 ± 3.20	0.5408 ± 0.1294	0.5389 ± 0.1290

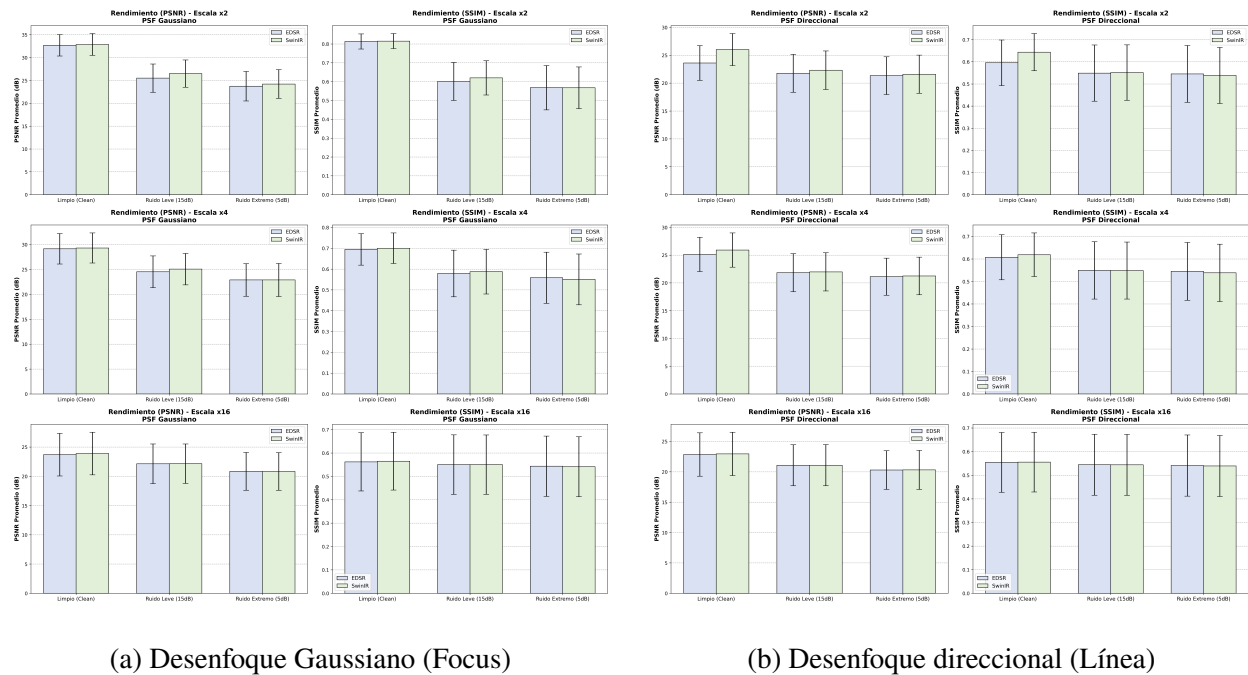
Nota. Los valores en negrita indican el mejor desempeño para cada métrica en el escenario evaluado del conjunto de test.

Para complementar el análisis numérico tabulado, la Figura 10 ilustra gráficamente el comportamiento estadístico de ambas arquitecturas en el escenario de mayor exigencia topológica (escala x16, emulación de nanosatélite). En los diagramas de barras, la altura representa el rendimiento promedio (PSNR y SSIM), mientras que las líneas verticales sobre cada barra indican la desviación estándar (σ).

Visualmente, se hace evidente la degradación del modelo base (EDSR) a medida que aumenta el ruido, presentando barras de error significativamente largas en 15 dB; esto denota una alta inestabilidad algorítmica y fallos impredecibles en la reconstrucción. En contraste, la arquitectura SwinIR no solo mantiene métricas promedio superiores, sino que exhibe barras de error notablemente más cortas en todos los umbrales de ruido, confirmando que su mecanismo de atención global proporciona una inferencia estructuralmente estable y confiable frente a perturbaciones estocásticas severas.

Figura 10

Evaluación del rendimiento sobre el conjunto de test bajo condiciones de degradación.



6.2.3 Inspección Visual y Discusión

Dada la extensividad de los experimentos realizados, el análisis cualitativo y la inspección visual exhaustiva para todas las escalas espaciales (x2, x4 y x16) frente a las distintas condiciones de degradación se documentan de manera detallada en el Apéndice F. Esta selección busca ilustrar el comportamiento de los modelos propuestos bajo tres condiciones críticas de operación: la capacidad de recuperación de nitidez pura, ante aberraciones asimétricas de captura y como se comporta ante el colapso de la información topológica.

7. Conclusiones

La implementación de algoritmos de superresolución basados en aprendizaje profundo demostró ser una alternativa viable y técnicamente sólida para superar las limitaciones ópticas intrínsecas de las plataformas satelitales compactas. A través de un proceso metodológico estructurado en cuatro fases, se logró validar cuantitativamente la capacidad de las redes neuronales para recuperar información espacial de alta frecuencia a partir de observaciones severamente degradadas, superando en la mayoría de los escenarios experimentales a los métodos convencionales de interpolación utilizados como línea base.

El protocolo de preprocesamiento diseñado constituyó el fundamento metodológico que dotó de validez física a toda la investigación. Al incorporar modelos matemáticos de degradación orbital reales, como núcleos PSF que emulan el desenfoque óptico y el motion blur derivado de la velocidad orbital, junto con la inyección de ruido, se logró trascender las limitaciones de los bancos de prueba convencionales que emplean únicamente submuestreo bicúbico ideal. El resultado fue un entorno de simulación que refleja fielmente las condiciones de captura de un satélite en órbita terrestre baja, garantizando que los modelos entrenados aprendan representaciones robustas y transferibles a escenarios operativos reales.

La selección y adaptación de EDSR y SwinIR permitió abordar el problema de superresolución satelital desde dos paradigmas distintos del aprendizaje profundo. La configuración del EDSR, optimizada para preservar la firma radiométrica de las imágenes satelitales, estableció un punto de comparación riguroso y computacionalmente eficiente. Por su parte, la adaptación del SwinIR con un decodificador de representación continua y una función de pérdida híbrida que penaliza simultáneamente el error espacial, la pérdida de bordes y la coherencia espectral, demostró que

las arquitecturas basadas en mecanismos de atención global son especialmente adecuadas para revertir degradaciones asimétricas, donde la comprensión del contexto global de la imagen resulta indispensable para la reconstrucción fiel de estructuras geográficas continuas.

El entrenamiento supervisado de ambas arquitecturas en Python con PyTorch evidenció que la complejidad de los modelos de atención global exige no solo mayor capacidad de cómputo, sino también estrategias de optimización más sofisticadas para alcanzar una convergencia estable. Mientras el EDSR convergió de manera adecuada con una configuración de hiperparámetros relativamente estándar, el SwinIR requirió la integración de técnicas avanzadas como el calentamiento de la tasa de aprendizaje, el decaimiento cosenoidal y el promedio móvil exponencial de pesos para estabilizar el comportamiento de los mecanismos de auto-atención. Esta diferencia en la dinámica de entrenamiento constituye en sí misma un resultado relevante, pues cuantifica el costo operativo real de adoptar arquitecturas Transformer en aplicaciones de teledetección a escala industrial.

La evaluación comparativa mediante métricas estandarizadas confirmó que la superioridad de una arquitectura sobre otra no es absoluta, sino dependiente del tipo y severidad de la degradación aplicada. En condiciones de degradación simétrica moderada, ambos modelos ofrecen un rendimiento comparable, mientras que ante degradaciones asimétricas que simulan el movimiento orbital, el SwinIR demostró una ventaja significativa gracias a su capacidad para modelar dependencias espaciales de largo alcance. No obstante, uno de los hallazgos más relevantes de la evaluación fue que, bajo condiciones de ruido extremo en factores de escala altos, la interpolación bicúbica superó numéricamente a ambos modelos de aprendizaje profundo en términos de PSNR. Este resultado, lejos de representar una limitación del enfoque propuesto, pone de manifiesto que cuando la señal de entrada está tan severamente corrompida que apenas contiene información estructural aprovechable, el suavizado pasivo de la bicúbica minimiza el error matemático generando un promedio uniforme, mientras que los modelos neuronales son penalizados por intentar reconstruir geometrías a partir de una señal casi inexistente. Ante este escenario, el SSIM demostró ser una métrica más fiel a la coherencia estructural real percibida, subrayando la importancia de evaluar la calidad de reconstrucción desde perspectivas complementarias y no depender exclusivamente del

PSNR como criterio de desempeño.

8. Recomendaciones

A partir de los resultados experimentales y las limitaciones identificadas durante el desarrollo de esta investigación, se proponen las siguientes líneas de trabajo futuro para optimizar y expandir la aplicación del aprendizaje profundo en la superresolución de imágenes satelitales:

8.1 Ampliación y validación del modelo de degradación

Se sugiere expandir el pipeline de preprocesamiento sintético incorporando variables físicas dinámicas del entorno orbital. Específicamente, se recomienda modelar perturbaciones atmosféricas como la dispersión de Rayleigh, la presencia de aerosoles y las variaciones de iluminación intra-orbital. Asimismo, para cuantificar la brecha de dominio, es imperativo validar los modelos entrenados sintéticamente frente a pares de imágenes reales capturadas simultáneamente por constelaciones de distinta resolución nativa, evaluando la transferencia del aprendizaje a sensores operativos.

8.2 Optimización arquitectónica y eficiencia computacional

Considerando la rápida saturación de la memoria VRAM observada al implementar arquitecturas basadas puramente en auto-atención bajo factores de escala extremos, se recomienda investigar el desarrollo de topologías híbridas. Integrar capas convolucionales en las etapas iniciales de extracción de características permitiría reducir la dimensionalidad espacial de los tensores antes de aplicar los mecanismos de ventanas desplazadas, logrando un equilibrio entre la eficiencia de

memoria y la fidelidad topográfica. Adicionalmente, resulta pertinente explorar técnicas de cuantificación de pesos y poda de red para evaluar la degradación de la métrica SSIM frente a la reducción del costo computacional.

8.3 Evolución de las métricas de evaluación

Los resultados experimentales demostraron que bajo condiciones de ruido extremo y submuestreos severos, la métrica PSNR presenta un comportamiento paradójico al favorecer el suavizado pasivo de la interpolación bicúbica. Por tanto, se recomienda prescindir del PSNR como métrica principal en etapas de alta degradación y transicionar hacia métricas de percepción visual profunda. De igual forma, se sugiere implementar una evaluación orientada a tareas; esto implica medir el rendimiento de la superresolución cuantificando el incremento o decremento en la precisión de algoritmos posteriores, tales como la segmentación semántica de redes viales o la detección automatizada de infraestructura urbana.

8.4 Proyección hacia el procesamiento en órbita

Finalmente, de cara a la aplicabilidad directa en misiones de la industria NewSpace, se recomienda realizar un estudio de viabilidad para la implementación de estos algoritmos directamente en el hardware de la plataforma satelital. Procesar la superresolución en órbita permitiría optimizar el ancho de banda del enlace de bajada, transmitiendo únicamente imágenes mejoradas o parches de interés. Este análisis debe centrarse en los requerimientos de potencia y disipación térmica del hardware de inferencia espacial frente a las severas restricciones que caracterizan a los nanosatélites.

Referencias Bibliográficas

- Chen, Y., Liu, S., y Wang, X. (2021). Learning continuous image representation with local implicit image function. En *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 8628-8638). <https://arxiv.org/abs/2012.09161>
- Demir, I., Koperski, K., Lindenbaum, D., Pang, G., Huang, J., Basu, S., Hughes, F., Tuia, D., y Raskar, R. (2018). Deepglobe 2018: A challenge to parse the earth through satellite images. En *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 172-181). <https://arxiv.org/abs/1805.06561>
- Dong, C., Loy, C. C., He, K., y Tang, X. (2016). Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(2), 295-307.
- Elad, M., y Feuer, A. (1997). Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images. *IEEE Transactions on Image Processing*, 6(12), 1646-1658. <https://doi.org/10.1109/83.650118>
- Gonzalez, R. C., y Woods, R. E. (2018). *Digital image processing* (4ta ed.). Pearson.
- Goodman, J. W. (2005). *Introduction to Fourier Optics* (3rd ed.). Roberts and Company Publishers.
- Keys, R. (1981). Cubic convolution interpolation for digital image processing. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 29(6), 1153-1160. <https://doi.org/10.1109/TASSP.1981.1163711>
- Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., y Timofte, R. (2021). SwinIR: Image restoration using swin transformer. En *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 1833-1844). <https://arxiv.org/abs/2108.10257>

- Lim, B., Son, S., Kim, H., Nah, S., y Mu Lee, K. (2017). Enhanced deep residual networks for single image super-resolution. En *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 136-144). <https://arxiv.org/abs/1707.02921>
- Poghosyan, A., y Golkar, A. (2017). CubeSat evolution: Analyzing CubeSat capabilities for conducting science missions. *Progress in Aerospace Sciences*, 88, 59-83. <https://doi.org/10.1016/j.paerosci.2016.11.002>
- Wang, Z., Bovik, A. C., Sheikh, H. R., y Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4), 600-612.
- Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., y Fu, Y. (2018). Image super-resolution using very deep residual channel attention networks. En *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 286-301). <https://arxiv.org/abs/1807.02758>

Apéndices

Apéndice A: Repositorio del Proyecto

Para garantizar la transparencia metodológica y la reproducibilidad de esta investigación, el código fuente desarrollado ha sido publicado de forma abierta. En el repositorio se encuentran los scripts de Python utilizados para el preprocesamiento matemático de las imágenes, el entrenamiento de las arquitecturas (EDSR y SwinIR), y la evaluación cuantitativa de los resultados.

El material y las instrucciones de uso pueden consultarse en el siguiente enlace de GitHub:

<https://github.com/jarajara25/SISR>

Apéndice B: Demostración Matemática de la Degradación Óptica

Propagación del campo y difracción en la pupila

Considérese una onda monocromática proveniente de una fuente puntual en el infinito, tal como un objeto observado desde la órbita terrestre. Al incidir sobre la apertura de entrada del sistema óptico, denotada por la función de transparencia $P(u, v)$, el campo eléctrico es truncado por las dimensiones físicas de la apertura. Bajo la aproximación de difracción de Fraunhofer (campo lejano), la amplitud compleja del campo eléctrico en el plano de la imagen, denotada por $E_i(x, y)$, se define mediante la siguiente integral de difracción:

$$E_i(x, y) = \frac{1}{j\lambda z} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(u, v) \exp \left[-j \frac{2\pi}{\lambda z} (ux + vy) \right] du dv \quad (8.1)$$

En esta expresión, λ representa la longitud de onda de operación, z es la distancia focal del sistema óptico, y las variables (u, v) y (x, y) corresponden a las coordenadas espaciales en el plano de la pupila y el plano de la imagen, respectivamente. Esta formulación demuestra que el sistema óptico actúa como un procesador analógico que realiza inherentemente una Transformada de Fourier sobre la señal luminosa (Goodman, 2005), donde las frecuencias espaciales (f_x, f_y) están mapeadas a las coordenadas del plano imagen según las relaciones $f_x = x/(\lambda z)$ y $f_y = y/(\lambda z)$.

Definición de la Point Spread Function (PSF)

Dado que los sensores electrónicos (CCD/CMOS) responden a la intensidad del flujo radiante y no a la fase del campo eléctrico, la respuesta del sistema a una fuente puntual, conocida

como Función de Dispersión de Punto (PSF, por sus siglas en inglés), se define como el cuadrado de la magnitud de la amplitud compleja del campo en el plano focal:

$$PSF(x, y) = |E_i(x, y)|^2 = |\mathcal{F}\{P(u, v)\}|^2 \quad (8.2)$$

Donde $\mathcal{F}\{\cdot\}$ denota el operador de la Transformada de Fourier bidimensional, el cual describe matemáticamente la propagación de la luz desde el plano de la pupila hasta el plano de la imagen bajo la aproximación de difracción de Fraunhofer.

Apéndice C: Evolución Histórica de la Superresolución

Métodos de interpolación clásica

Antes de la consolidación de la inteligencia artificial, la superresolución se abordaba fundamentalmente como un problema de procesamiento digital de señales mediante técnicas de interpolación. Estos métodos estiman el valor de un nuevo píxel a partir de una vecindad de puntos conocidos utilizando funciones matemáticas predefinidas. Entre los enfoques más comunes se encuentra la interpolación bilineal, que promedia los cuatro píxeles más cercanos, y la interpolación bicúbica (Keys, 1981), que emplea una vecindad de dieciséis píxeles y una función polinómica de tercer grado.

A pesar de su eficiencia computacional y su uso extendido como estándar industrial para el redimensionamiento de imágenes, estos métodos presentan limitaciones críticas en aplicaciones de teledetección. Al basarse en coeficientes fijos, actúan esencialmente como filtros pasa-bajo que suavizan la imagen, eliminando los detalles de alta frecuencia y produciendo bordes difusos. Dado que estas técnicas no poseen la capacidad de sintetizar información inexistente en la captura original, en el contexto de esta investigación se consideran únicamente como una línea base o baseline para evaluar el rendimiento de modelos más sofisticados.

Pioneros del Aprendizaje Profundo: SRCNN

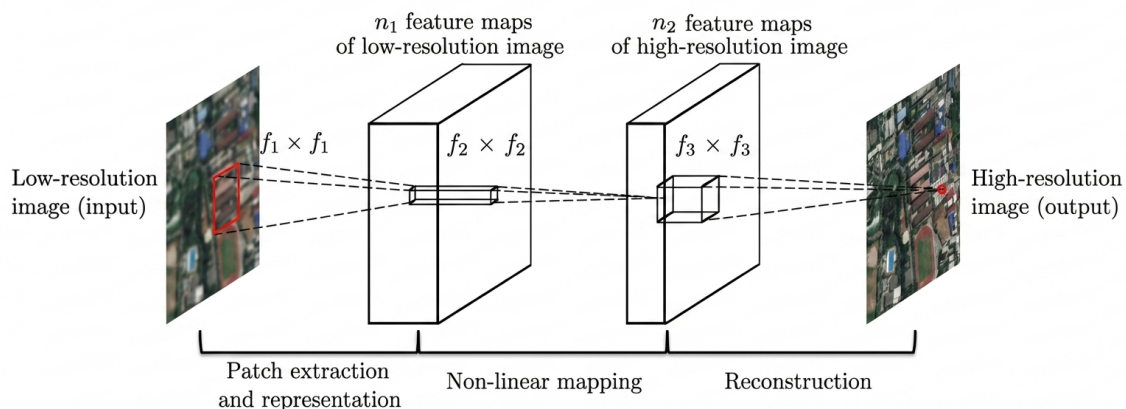
El cambio de paradigma hacia el aprendizaje automático se consolidó con la introducción de la red neuronal convolucional para superresolución, conocida como SRCNN (Dong et al., 2016).

Esta arquitectura fue la primera en demostrar que era posible aprender de manera directa, y de extremo a extremo, el mapeo complejo que existe entre una observación de baja resolución y su contraparte de alta fidelidad. A diferencia de los métodos clásicos, SRCNN sustituye los filtros manuales por capas de convolución cuyos pesos se optimizan durante el entrenamiento.

El funcionamiento de esta red se divide en tres etapas fundamentales que fluyen de manera continua. En la primera etapa, la red realiza una extracción de características locales o parches de la imagen de entrada. Posteriormente, mediante una capa de mapeo no lineal, estas características se transforman en representaciones de alta dimensión que vinculan estadísticamente ambos dominios de resolución. Finalmente, una última capa de reconstrucción ensambla estas representaciones para generar la imagen de salida mejorada. Aunque su diseño de tres capas es sencillo comparado con los estándares actuales, SRCNN sentó los fundamentos teóricos para el desarrollo de arquitecturas más profundas, al demostrar que la optimización basada en datos es superior a cualquier operador matemático estático.

Figura 11

Arquitectura de la red neuronal SRCNN

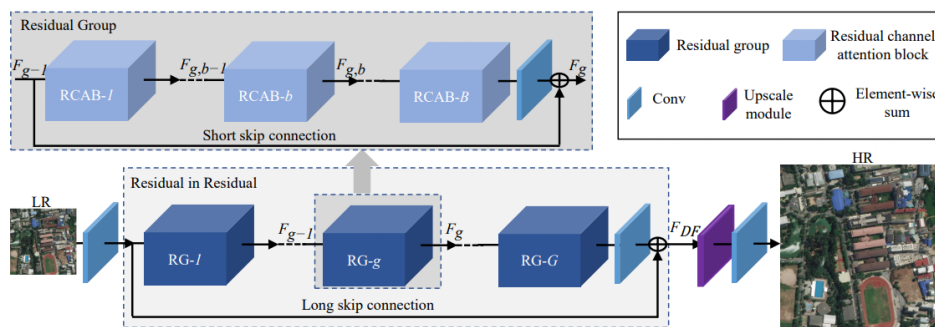


Arquitecturas residuales y de alto rendimiento: RCAN

Posteriormente a la red EDSR, la evolución de estos sistemas integró mecanismos de atención para priorizar la información más relevante dentro de los mapas de características. Un exponente destacado de esta tendencia es la arquitectura RCAN (Residual Channel Attention Networks) (Zhang et al., 2018). A diferencia de los modelos anteriores que trataban todos los canales de características de manera equitativa, RCAN introdujo el mecanismo de atención de canal, el cual permite a la red enfocarse en los componentes de la señal que contienen información estructural crítica mientras suprime el ruido o las características menos útiles. Esta capacidad de discriminación es especialmente valiosa en imágenes de teledetección, donde la variabilidad de texturas y el ruido electrónico exigen una selección inteligente de los datos para lograr una reconstrucción precisa de los bordes.

Figura 12

Arquitectura propuesta del modelo RCAN.



Apéndice D: Dinámicas y Configuración de Entrenamiento

Dinámica de Entrenamiento para la Línea Base: EDSR

La arquitectura puramente convolucional requiere un enfoque de optimización reactivo y constante. La dinámica se rigió por los siguientes hiperparámetros: Optimizador: Algoritmo Adam, estándar para el ajuste adaptativo en redes residuales profundas.

1. **Tasa de Aprendizaje Reactiva:** Tasa inicial estática de 1×10^{-4} . Se implementó un programador de reducción por estancamiento (`ReduceLRonPlateau`), el cual disminuye la tasa a la mitad si el PSNR de validación no mejora durante 15 épocas consecutivas.
2. **Tamaños de Lote:** Asignación conservadora (1 a 2 imágenes por iteración) y estabilización matemática mediante el recorte de gradientes (`Gradient Clipping`) con una norma máxima de 5.0, mitigando el impacto de gradientes explosivos causados por el ruido extremo (`PSF_5`).

Dinámica de Entrenamiento Avanzado: SwinIR + LIIF

El entrenamiento del modelo basado en Vision Transformers requiere una orquestación algorítmica significativamente más sofisticada para alcanzar la convergencia, debido a la inestabilidad inherente de los mecanismos de atención en las etapas tempranas.

1. **Aceleración y Precisión Mixta (AMP):** Para optimizar la huella en los 16 GB de VRAM de la GPU, se habilitó el uso de tensores de menor precisión (`autocast`) de manera dinámica.
2. **Optimizador AdamW:** Se reemplazó el Adam tradicional por AdamW (Loshchilov & Hutter,

2017), aplicando un decaimiento de pesos (Weight Decay) de 1×10^{-4} . Esto actúa como un regularizador crucial para evitar que las cabezas de atención del Transformer memoricen el ruido en lugar de la estructura topográfica.

3. **Planificación de Tasa de Aprendizaje (Warmup + Cosine):**

- **Calentamiento (Warmup):** Incremento lineal de la tasa de aprendizaje durante las primeras 10 épocas hasta alcanzar el valor óptimo de 2×10^{-4} , estabilizando los gradientes iniciales.
- **Enfriamiento (Cosine Annealing):** Transición hacia una curva de decaimiento cosenooidal, reduciendo suavemente la tasa hasta un mínimo de 1×10^{-7} en las etapas finales del ciclo para refinar los detalles estructurales (finetuning).

4. **Promedio Móvil Exponencial (EMA):** Se implementó una sombra de los pesos de la red utilizando EMA con un factor de decaimiento de 0.999. Las métricas de validación y la inferencia final se evalúan sobre este modelo promediado, lo que incrementa drásticamente la robustez frente a perturbaciones estocásticas (PSF).

5. **Decodificador LIIF Implícito:** A diferencia de EDSR, la reconstrucción final omite el paso de escalado convolucional clásico, delegando la inferencia de píxeles continuos a un perceptrón multicapa en combinación con una interpolación bicúbica residual (Chen et al., 2021).

Tabla 6

Comparativa de hiperparámetros de entrenamiento: EDSR vs. SwinIR.

Parámetro	Arquitectura EDSR	Arquitectura SwinIR
Función de Pérdida	Híbrida (L_1 + Edge Loss)	Híbrida (Charbonnier + Edge + Frecuencia)
Algoritmo Optimizador	Adam ($\beta_1 = 0,9, \beta_2 = 0,999$)	AdamW (Weight Decay = 1×10^{-4})
Tasa de Aprendizaje	1×10^{-4}	2×10^{-4}
Planificador (LR)	<i>ReduceLRonPlateau</i>	<i>LinearLR</i> (Warmup) + <i>CosineAnnealingLR</i>
Tamaño de Lote	1 a 2 imágenes (Dinámico)	1 a 2 imágenes (Dinámico)
Épocas Máximas	300	300
<i>Early Stopping</i>	No aplica directamente	40 épocas

Apéndice E: Análisis de Convergencia y Funciones de Pérdida

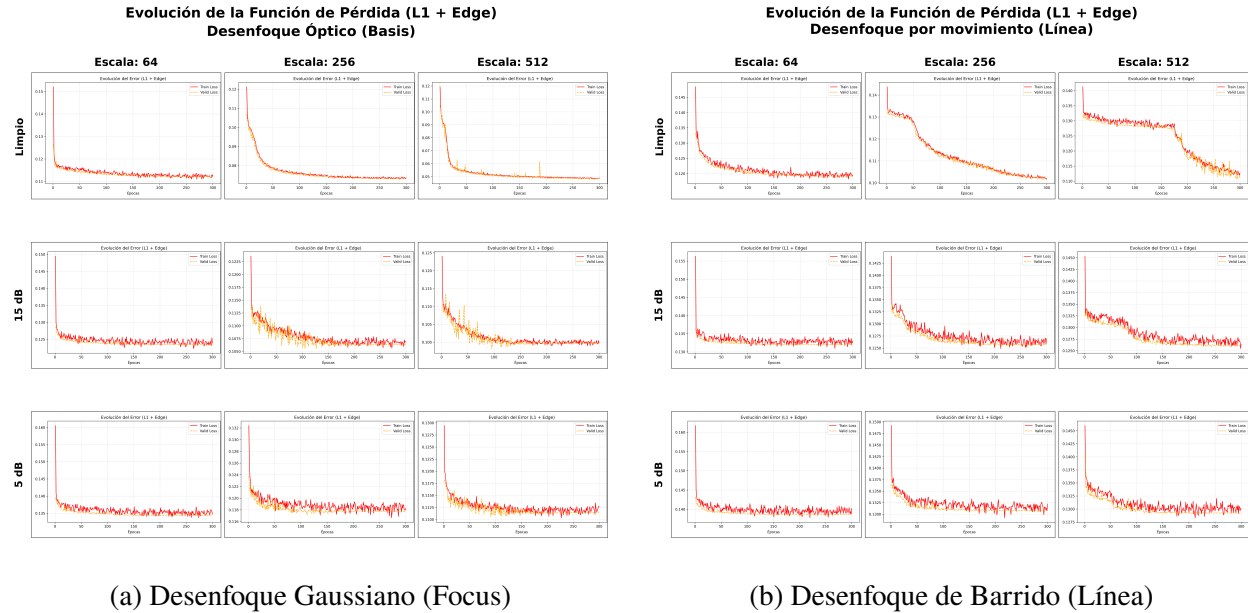
Convergencia de la Línea Base Convolutiva (EDSR)

La Figura 13 presentan la evolución del error para el modelo EDSR. En los escenarios de baja complejidad (escala x2, sin ruido), la magnitud de la pérdida disminuye de forma monotónica, indicando un aprendizaje adecuado de los patrones espaciales. Sin embargo, conforme se incrementa el factor de escala (x4, x16) y se introduce ruido estocástico (15 dB y 5 dB), se observa un aumento progresivo en la brecha entre las curvas de entrenamiento y validación.

En las condiciones de mayor degradación (escala x16, SNR 5 dB), la curva de validación del EDSR presenta oscilaciones de alta frecuencia y una tendencia a la divergencia en las épocas finales. Este comportamiento evidencia que las operaciones convolucionales locales tienden a sobreajustarse al ruido de la imagen de entrada en lugar de extraer las características topográficas subyacentes. La inestabilidad de la métrica se hace más pronunciada al evaluar la matriz correspondiente al PSF Línea, donde la asimetría del desenfoque direccional dificulta la minimización del error de validación.

Figura 13

Dinámica de optimización de la arquitectura convolucional (EDSR). Se contrasta la evolución de la función de pérdida frente al desenfocado gaussiano (a) y direccional (b)



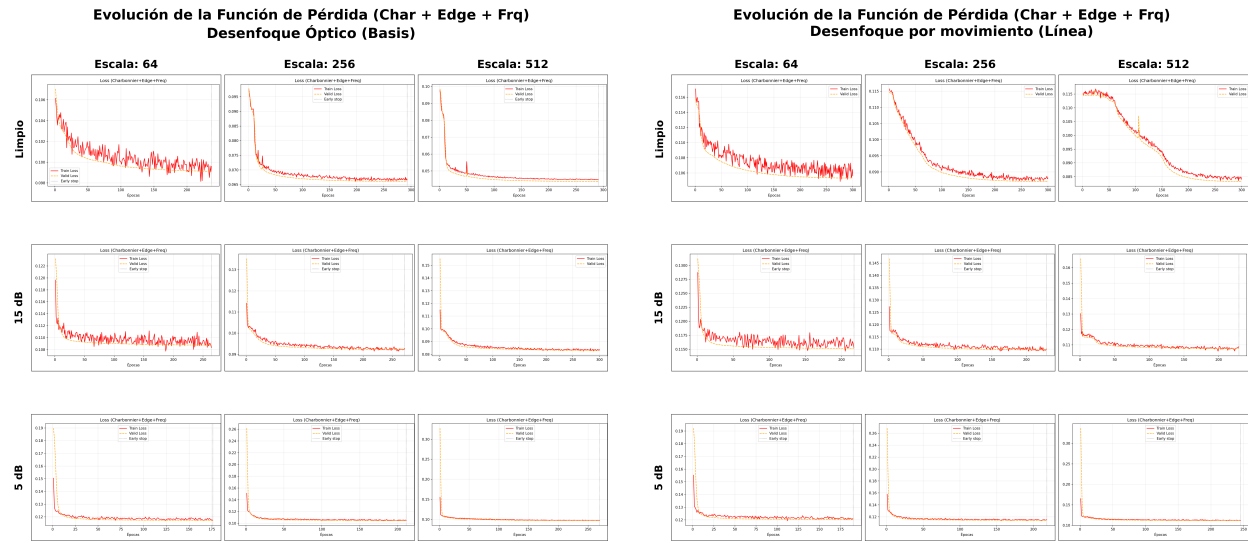
Convergencia del Modelo Avanzado (SwinIR)

En contraste, las Figura 14 exponen la dinámica de entrenamiento de la arquitectura SwinIR bajo las mismas condiciones experimentales. La inclusión de estrategias de optimización como el calentamiento de la tasa de aprendizaje (Warmup) previene la aparición de picos iniciales de error, logrando un descenso más uniforme en las primeras etapas.

A lo largo de todos los cuadrantes de las matrices, el modelo basado en Transformers mantiene una convergencia más estable que la línea base. Incluso en el escenario más crítico (escala x16, SNR 5 dB), la curva de validación no presenta signos de divergencia y logra un comportamiento asintótico horizontal. Este resultado cuantitativo indica que el procesamiento mediante ventanas desplazadas (Shifted Windows) proporciona un campo receptor global que permite a la red discriminar de manera más eficaz entre el ruido aleatorio y las estructuras geométricas reales, asegurando la estabilidad del entrenamiento independientemente del tipo de desenfoco aplicado.

Figura 14

Evolución de la pérdida del SwinIR para desenfoque gaussiano (a) y direccional (b).



(a) Desenfoque Gaussiano (Focus)

(b) Desenfoque de Barrido (Línea)

Apéndice F: Inspección Visual Multiescala

Análisis de Refinamiento Estructural en Escala x2

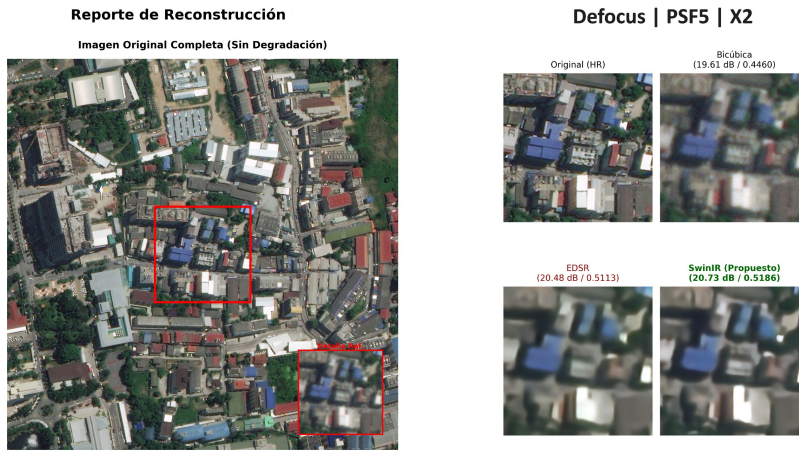
En la escala x2, el desafío principal radica en el refinamiento de bordes arquitectónicos. En la condición ideal (b), el modelo SwinIR alcanza una métrica superior de 31,16 dB, superando a EDSR(30,95 dB) en la definición de ángulos y contornos rectos. Mientras que la interpolación bicúbica (19,73 dB) entrega una imagen difuminada, el modelo propuesto logra una reconstrucción de alta frecuencia visualmente cercana a la verdad de campo, validando la eficacia de los mecanismos de atención para preservar la geometría urbana en condiciones óptimas.

No obstante, el rendimiento del sistema enfrenta su prueba más crítica en el nivel intermedio con movimiento. En concordancia con los hallazgos previos, el desenfoque direccional se identifica como la degradación más compleja de revertir para ambas arquitecturas debido a la asimetría en la pérdida de información. A pesar de ser el escenario donde el modelo presenta mayor dificultad para procesar la imagen reflejado en la caída de los valores absolutos de PSNR, el SwinIR logra aquí su mayor ventaja comparativa con 17,90 dB frente a los 16,06 dB de EDSR.

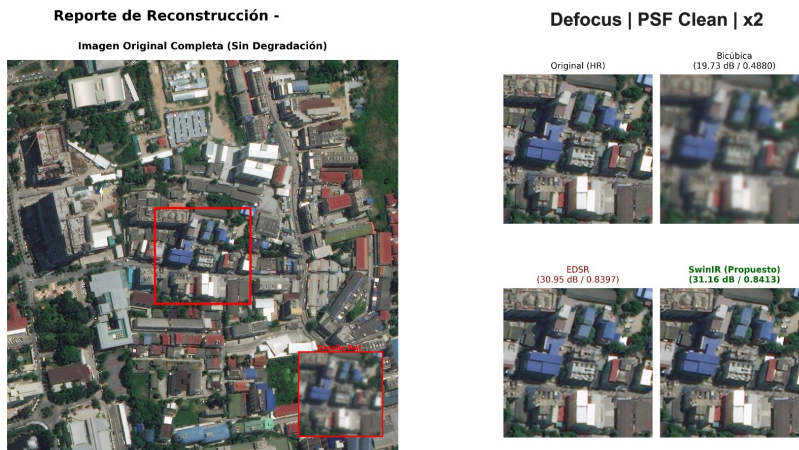
Mientras que la convolución local de EDSR presenta dificultades antes los movimientos lineales que resulta en un parche con pérdida total de contraste, el SwinIR logra compensar parcialmente la distorsión, rescatando la linealidad de las estructuras que de otro modo quedarían omitidas. Finalmente, incluso en ruido extremo (a), el modelo logra mantenerse por encima de los 20 dB, demostrando una robustez superior en tareas de filtrado y reconstrucción simultánea frente a la degradación.

Figura 15

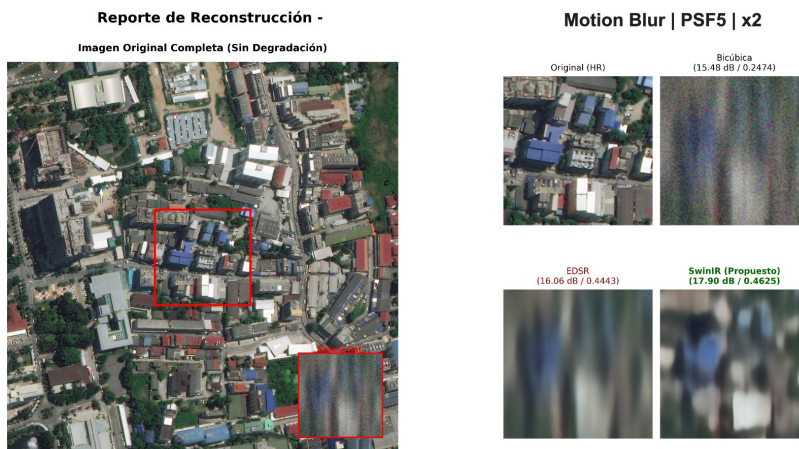
Comparativa a diferentes condiciones para escalamiento x2



(a) Condición ruido extremo



(b) Condición ideal



(c) Nivel de ruido intermedio con movimiento

Análisis de Estabilidad Estructural en Escala x4

Al incrementar el factor de escala a $x4$, se observa un fenómeno analítico crítico en la condición de ruido extremo (a), donde la interpolación Bicúbica (19,64 dB) logra superar numéricamente tanto al EDSR (19,54 dB) como al SwinIR (19,57 dB) en la métrica PSNR. Este comportamiento se explica porque, ante un ruido tan severo, el suavizado pasivo de la bicúbica reduce el error cuadrático medio de forma más efectiva que el intento de reconstrucción de los modelos de aprendizaje profundo, los cuales tienden a interpretar el ruido como información estructural. Sin embargo, cualitativamente la bicúbica entrega un parche operativamente nulo por su excesiva borrosidad, mientras que el SwinIR logra rescatar la coherencia geométrica de los edificios con el SSIM más estable de la comparativa.

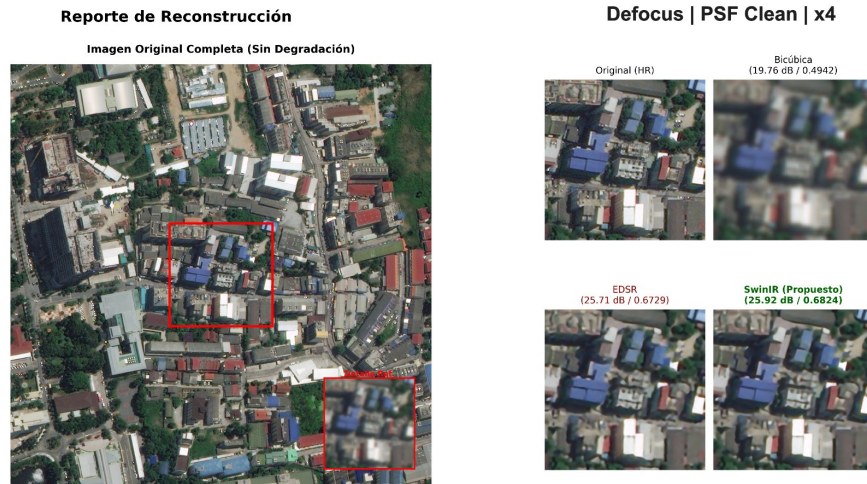
En la condición ideal (b), el SwinIR (25,92 dB) consolida su ventaja sobre el EDSR (25,71 dB) al definir bordes arquitectónicos con mayor limpieza y sin las oscilaciones de contorno típicas de las redes convolucionales al procesar escalas altas. Finalmente, el escenario de movimiento (c) reafirma que el desenfoque direccional (PSF línea) es la degradación que más dificulta el procesamiento. A pesar de la caída generalizada de las métricas, el SwinIR (17,50 dB) logra corregir parte del efecto de “arrastre” del sensor y rescatar la linealidad de las sombras, superando la tendencia del EDSR de generar parches borrosos y con artefactos de color que falsean la posición real de las estructuras urbanas.

Figura 16

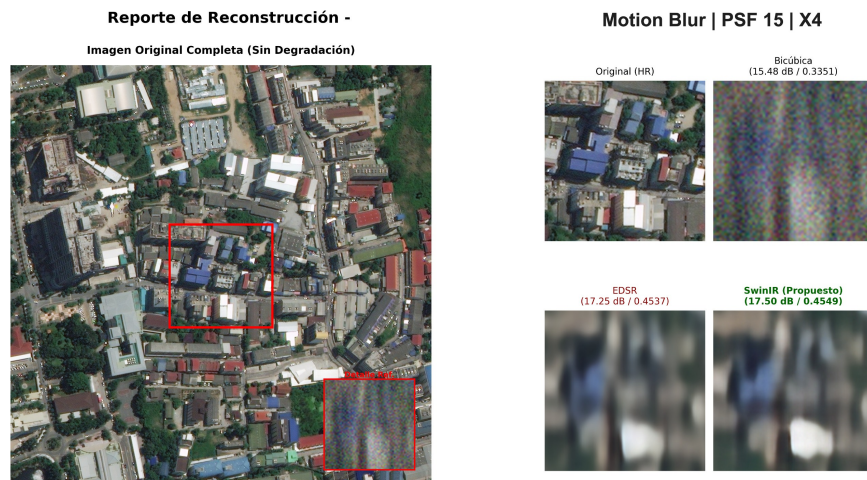
Comparativa a diferentes condiciones para escalamiento x4



(a) Condición ruido extremo



(b) Condición ideal



(c) Nivel de ruido intermedio con movimiento

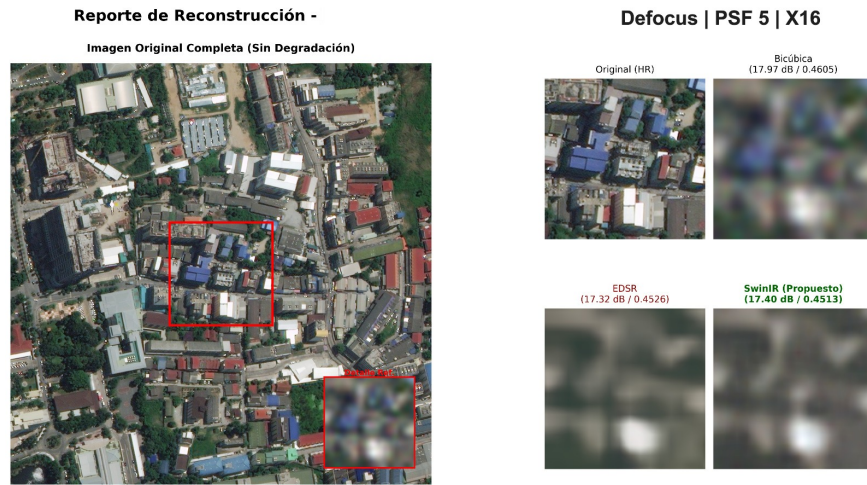
Análisis de Inferencia Topológica en Escala x16

La escala x16 representa el límite operativo del sistema, donde la pérdida de información espacial es masiva. En la condición de ruido extremo (a), se observa de forma acentuada la paradoja del PSNR: la interpolación Bicúbica (17,97 dB) supera numéricamente a los modelos de aprendizaje profundo. Este fenómeno se debe a que, ante una señal casi inexistente, el suavizado masivo de la bicúbica minimiza el error cuadrático medio al generar un promedio uniforme, mientras que los modelos intentan reconstruir formas que el ruido ha desplazado, siendo penalizados por cualquier desviación estructural. El SwinIR (17,40 dB), aunque con menor PSNR, logra rescatar una coherencia semántica mínima que permite intuir la presencia de volúmenes urbanos al ojo humano.

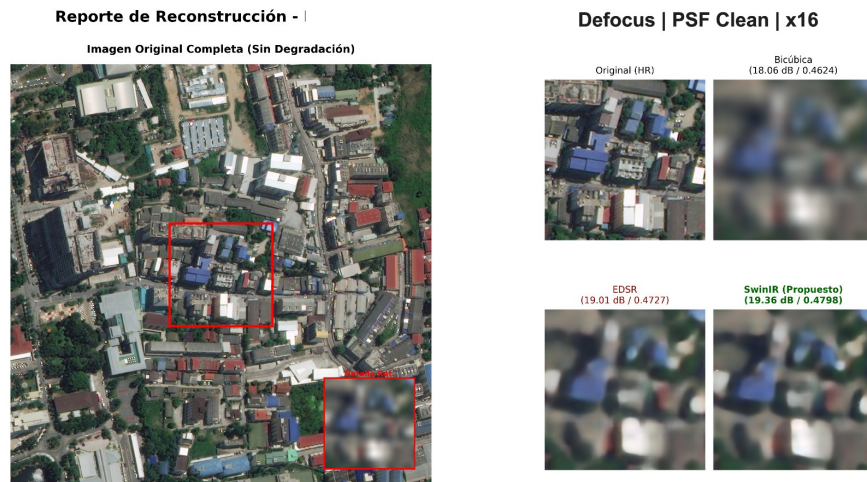
En la condición ideal (b), el SwinIR (19,36 dB) consolida su capacidad de inferencia, logrando interpretar la lógica espacial de los edificios con mayor claridad que el EDSR (19,01 dB). Sin embargo, el escenario de movimiento direccional (c) revela el límite tecnológico de la metodología. En este punto, el EDSR (16,39 dB) logra una ligera ventaja numérica sobre el SwinIR (16,32 dB), indicando que bajo degradaciones asimétricas extremas en factores de escala tan altos, la capacidad de atención del SwinIR alcanza un techo en la teoría de la información. Estos resultados confirman que en x16 el sistema transiciona de una tarea de reconstrucción a una de inferencia de contexto, donde la recuperación de microtexturas es físicamente inviable.

Figura 17

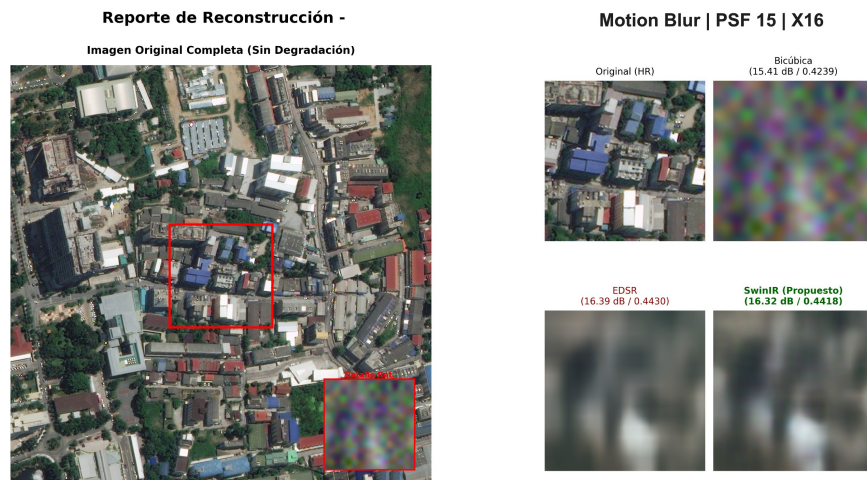
Comparativa a diferentes condiciones para escalamiento x16



(a) Condición ruido extremo



(b) Condición ideal



(c) Nivel de ruido intermedio con movimiento