

Super-resolution of satellite images using GAN: A scheme based on training with aerial images

Magda Alexandra Trujillo-Jiménez^{1,5}[0000-0001-5506-3496], Francisco Iaconis²[0000-0002-2373-5793],
Debora Pollicelli³[0009-0006-8771-2538], Gisela Noelia Revollo Sarmiento⁴[0000-0002-1532-5428],
Claudio Delrieux¹[0000-0002-2727-8374]

¹ Departamento de Ciencias e Ingeniería de la Computación, Universidad Nacional del Sur - CONICET, Bahía Blanca, Argentina

² Instituto de Física del Sur, Departamento de Física, Universidad Nacional del Sur - CONICET, Bahía Blanca, Argentina

³ Laboratorio de Investigación en Informática, Departamento de Informática, Facultad de Ingeniería, Universidad Nacional de la Patagonia San Juan Bosco, Puerto Madryn, Argentina

⁴ Instituto de Ecorregiones Andinas (INECOA - CONICET), Facultad de Ingeniería, Universidad Nacional de Jujuy, Jujuy, Argentina

⁵ Instituto Patagónico de Ciencias Sociales y Humanas, Centro Nacional Patagónico CCT CENPAT CONICET, Puerto Madryn, Argentina
mtrujillo@cenpat-conicet.gob.ar

Abstract. Satellite images often have limitations in terms of spatial resolution and, in many cases, high acquisition costs, which restricts their use in applications such as urban monitoring, land management, and wildlife studies. This work proposes an innovative approach that uses high-resolution aerial images to train a super-resolution model based on Generative Adversarial Networks. In particular, the ESRGAN (Enhanced Super-Resolution Generative Adversarial Network) model is adapted, optimizing its parameters in order to improve its computational efficiency and reduce training times. The model trained with aerial images is then evaluated on low-resolution satellite clips, analyzing its performance at x2 and x4 scale factors using structural, perceptual, and chromatic metrics (SSIM-Y, MS-SSIM, LPIPS, and CIEDE2000). The results show clear visual improvements, with greater sharpness, better edge definition, and consistent recovery of urban structures and terrain features. Quantitatively, the x2 scale achieves the highest values, while the x4 scale maintains stable and useful performance for practical applications. These findings demonstrate the feasibility of transferring super-resolution capability from aerial images to satellite images, even under spectral and geometric differences between domains. Overall, this work establishes a solid foundation for the development of low-cost, high-impact satellite super-resolution models and opens up future lines of research aimed at expanding training data, incorporating domain adaptation techniques, and exploring specific architectures for satellite sensors.

Keywords: Adversarial Generative Networks, Satellite Imagery, Aerial Imagery.

Super-resolución de imágenes satelitales usando GAN: Un esquema basado en entrenamiento con imágenes aéreas

Resumen. Las imágenes satelitales suelen presentar limitaciones en su resolución espacial y, en muchos casos, altos costos de adquisición, lo que restringe su uso en aplicaciones como el monitoreo urbano, la gestión territorial y el estudio de fauna. Este trabajo propone un enfoque innovador que aprovecha imágenes aéreas de alta resolución para entrenar un modelo de super-resolución basado en Redes Generativas Adversarias. En particular, se adapta el modelo ESRGAN (Enhanced Super-Resolution Generative Adversarial Network), optimizando sus parámetros con el fin de mejorar su eficiencia computacional y reducir los tiempos de entrenamiento. El modelo entrenado con imágenes aéreas se evalúa posteriormente sobre recortes satelitales de baja resolución, analizando su desempeño en factores de escala x2 y x4 mediante métricas estructurales, perceptuales y cromáticas (SSIM-Y, MS-SSIM, LPIPS y CIEDE2000). Los resultados muestran mejoras visuales claras, con mayor nitidez, mejor definición de bordes y una recuperación coherente de estructuras urbanas y elementos del terreno. Cuantitativamente, la escala x2 alcanza los valores más altos, mientras que la escala x4 mantiene un rendimiento estable y útil para aplicaciones prácticas. Estos hallazgos demuestran la viabilidad de transferir la capacidad de super-resolución desde imágenes aéreas hacia imágenes satelitales, incluso bajo diferencias espectrales y geométricas entre dominios. En conjunto, este trabajo establece una base sólida para el desarrollo de modelos de super-resolución satelital de bajo costo y alto impacto, y abre futuras líneas de investigación orientadas a ampliar los datos de entrenamiento, incorporar técnicas de domain adaptation y explorar arquitecturas específicas para sensores satelitales.

Palabras clave: Redes Generativas Adversarias, Imágenes Satelitales, Imágenes Aéreas.

1 Introducción

La resolución espacial de las imágenes satelitales constituye un parámetro determinante para el éxito de múltiples aplicaciones en teledetección, incluyendo el monitoreo urbano, la planificación territorial, la gestión de recursos naturales, la agricultura de precisión y la conservación ambiental. En contextos más recientes, esta necesidad se ha extendido a tareas críticas como el monitoreo de infraestructura, la detección temprana de desastres naturales, la identificación de edificaciones irregulares o la vigilancia de áreas afectadas por deforestación y minería ilegal (Li et al., 2023, Liu et al., 2015, Wang et al., 2019). En los últimos años, la creciente disponibilidad de sensores ópticos y de radar de apertura sintética (SAR) ha impulsado el uso de la teledetección como fuente esencial para la observación terrestre continua. Sin embargo, debido a limitaciones técnicas y costos asociados, muchas de estas imágenes presentan una resolución insuficiente para abordar tareas que requieren un alto nivel de detalle (Shan et al., 2018, Lanaras et al., 2018).

La disponibilidad de sensores como Landsat-9 y Sentinel-2, que ofrecen resoluciones de 10 a 30 m/píxel, ha democratizado el acceso a datos de observación terrestre, aunque su capacidad para tareas de alta precisión sigue siendo limitada. Por otro lado, plataformas comerciales como PlanetScope, Pleiades-Neo y WorldView-3/4 alcanzan resoluciones inferiores a 0,5 m, pero su uso se ve restringido por los costos elevados, tiempos de revisita largos y limitaciones de acceso (Kang et al., 2022). Este compromiso entre resolución espacial, cobertura espectral y temporal constituye un desafío para la observación de la Tierra (Wang et al., 2023). A ello se suman las dificultades relacionadas con la calidad radiométrica, el ruido atmosférico y la variabilidad geométrica entre sensores, que impactan la comparabilidad y consistencia de las imágenes multitemporales (Galar et al., 2020; Salgueiro et al., 2021).

Esto ha llevado a la exploración de técnicas de super-resolución (Yang et al., 2019, Karwowska et al., 2022), que buscan mejorar la calidad de las imágenes mediante el procesamiento computacional. Los enfoques tradicionales de super-resolución, basados en interpolación bilineal, bicúbica o Lanczos, presentan baja capacidad de reconstrucción de texturas y bordes finos (Vivone et al., 2021). Asimismo, los métodos de pan-sharpening, aunque útiles para fusionar información pancromática y multiespectral, tienden a introducir distorsiones espectrales cuando los sensores presentan discrepancias en la respuesta radiométrica o en la geometría orbital (Lanaras et al., 2018). El auge del Deep Learning revolucionó este campo con la aparición de modelos convolucionales como SRCNN (Dong et al., 2016), EDSR (Lim et al., 2017), RCAN (Zhang et al., 2018) y más recientemente, SwinIR (Liang et al., 2022). Estas arquitecturas, capaces de aprender representaciones jerárquicas de los patrones espaciales, lograron mejoras significativas en la reconstrucción de detalles estructurales. Sin embargo, su desempeño depende en gran medida de la disponibilidad de datasets de entrenamiento bien alineados y de alta calidad, un recurso escaso en teledetección (Pineda et al., 2020).

Entre estas técnicas, las Redes Generativas Adversarias (GANs) han demostrado un gran potencial debido a su capacidad para generar detalles de alta calidad a partir de imágenes de baja resolución (Ledig et al., 2017, Karwowska et al., 2022).

En este contexto, el modelo ESRGAN (Enhanced Super-Resolution Generative Adversarial Network) se ha posicionado como una de las arquitecturas más efectivas para tareas de super-resolución en imágenes generales (Wang et al., 2018, 2021). No obstante, su aplicación en dominios específicos, como las imágenes satelitales, requiere adaptaciones y entrenamientos especializados.

Este trabajo propone un enfoque novedoso que utiliza imágenes aéreas de alta resolución, más precisamente ortofotos para entrenar este modelo, con el objetivo de transferir esta capacidad de super-resolución a imágenes satelitales de baja resolución de la misma zona geográfica. Las imágenes aéreas, al tener una resolución significativamente mayor, permiten generar pares de imágenes de alta y baja resolución ideales para el entrenamiento del modelo. Los resultados preliminares, muestran una mejora significativa en la calidad de las imágenes aéreas y satelitales, lo que valida el potencial de esta metodología.

2 Materiales y Métodos

2.1 Área de estudio

Para este trabajo se consideró como área de estudio la ciudad de Lleida, situada en la comunidad autónoma de Cataluña, al noreste de España. Se trata de una zona con una combinación de áreas urbanas densas, zonas periurbanas y sectores agrícolas, lo que la convierte en un entorno adecuado para evaluar la capacidad del modelo de super-resolución en diferentes tipos de superficie. El relieve predominantemente llano y la distribución heterogénea de edificaciones, calles y vegetación ofrecen una diversidad de patrones espaciales útiles para el entrenamiento y validación del modelo. La elección de Lleida como área de estudio se debió principalmente a la disponibilidad de imágenes aéreas ortorectificadas de alta calidad provenientes del Instituto Cartográfico y Geológico de Cataluña (ICGC).



Fig. 1. Vista geoespacial de la ciudad de Lleida que muestra el área de estudio utilizada para el análisis. Se destacan las principales zonas urbanas y las características espaciales relevantes para este trabajo (Fuente: Instituto Cartográfico y Geológico de Cataluña ICGC).

2.2 Datos de entrenamiento y pre-procesamiento

Se utilizó un conjunto de datos compuesto por 2 imágenes aéreas tipo ortofotos de la ciudad de Lleida. Estas imágenes fueron obtenidas mediante vuelos de drones y posteriormente ortorectificadas para eliminar distorsiones geométricas producidas por el relieve y el ángulo de captura, garantizando así una representación precisa y uniforme de la superficie terrestre (Fernández-Lozano et al., 2026, Abel et al., 2019, Garcia Diaz, 2023).

Las imágenes tienen una resolución espacial de 25 cm el pixel y fueron descargadas en formato RGB (banda roja, verde y azul), utilizando el complemento “Open ICGC” (ICGC, 2024), del Sistema de Información Geográfico QGIS (<https://qgis.org/>). Las ortofotos son imágenes aéreas verticales corregidas geométricamente para mantener una escala uniforme en toda su extensión, lo que permite representar con precisión la superficie terrestre. Una ortofoto en color captura información del espectro visible,

combinando las bandas RGB (Rojo, Verde y Azul) para generar una imagen en "color natural", que refleja la apariencia real del paisaje.

Las imágenes originales, de tamaños de 15715x12261 y 13480x14564 píxeles, fueron procesadas para generar 1418 recortes de alta y baja resolución (ver Fig. 2). Para ello, se extrajeron recortes de 512x512 píxeles como referencia de alta resolución. Luego, cada recorte fue escalado a una versión de baja resolución de 128x128 píxeles, utilizando el método de interpolación por área, que preserva mejor los detalles al reducir el tamaño de la imagen. Esta reducción se realizó dividiendo las dimensiones de la imagen por un factor de escala de x2 y x4.

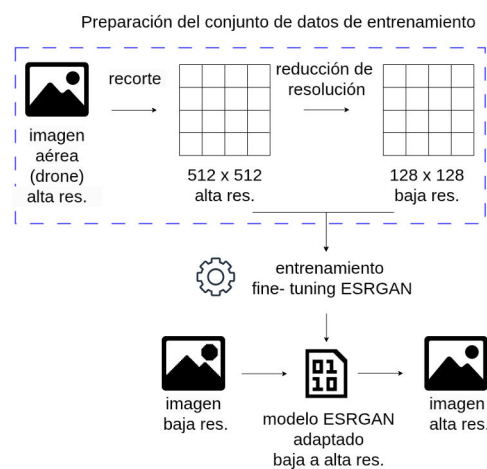


Fig. 2. Diagrama de flujo de pre procesamiento de datos para el entrenamiento.

2.3 Datos de prueba

Para el proceso de prueba, se utilizaron imágenes multiespectrales SuperView-1 con aproximadamente un 10% de cobertura de nubes. Las bandas utilizadas fueron azul (450–520 nm), verde (520–590 nm) y rojo (630–690 nm) con una resolución espacial de 2 m. Se realizaron recortes en RGB de 615 x 615 píxeles.

Es fundamental que el entrenamiento se realice con una gran cantidad y diversidad de imágenes, de modo que incluso los patrones menos frecuentes estén estadísticamente representados. En caso contrario, por ejemplo en contextos donde existan particularidades como cultivos experimentales con líneas de siembra circulares en lugar de longitudinales, el modelo tenderá a reconstruir patrones mayoritarios (e.g., líneas longitudinales), introduciendo sesgos. Cabe destacar que, dado que la escala de interpolación de píxeles es menor que los patrones espaciales de interés, esta limitación no representa un obstáculo crítico para la calidad general de los resultados.

2.4 Descripción del modelo Real-ESRGAN

Para mejorar la resolución de las imágenes satelitales se utilizó el modelo Real-ESRGAN (Enhanced Super-Resolution Generative Adversarial Network)

propuesto por Wang et al. (2021), una extensión del modelo ESRGAN original (Wang et al., 2018). Este modelo está basado en Redes Generativas Adversarias (GANs) y fue diseñado específicamente para tareas de super-resolución de imágenes en contextos con degradaciones reales y complejas.

La arquitectura de Real-ESRGAN mantiene el mismo generador utilizado en ESRGAN, compuesto por 16 bloques residuales densos (Residual-in-Residual Dense Blocks, RRDB). Estos bloques permiten preservar la información a través de múltiples niveles de profundidad, mejorando la reconstrucción de detalles finos y reduciendo artefactos no deseados en las imágenes generadas. Además, se emplea un factor de escala $\times 4$, logrando aumentar la resolución espacial de las imágenes originales (ver Fig. 3).

El modelo incorpora un proceso denominado pixel-unshuffle, que reduce el tamaño espacial de las imágenes de entrada mientras aumenta el número de canales. Este procedimiento permite que la mayor parte de los cálculos se realicen en un espacio de menor resolución, disminuyendo así el uso de memoria GPU y los requerimientos computacionales durante el entrenamiento (Wang et al., 2021).

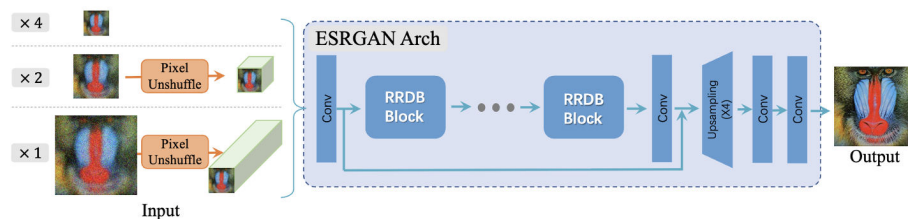


Fig. 3. Arquitectura general del modelo Real-ESRGAN, basada en una red generativa adversaria con 16 bloques residuales densos (RRDB) en el generador. El modelo incorpora pixel-unshuffle, conexiones residuales y normalización espectral para mejorar la estabilidad y la calidad de la super-resolución (Wang et al., 2021).

A diferencia de ESRGAN, Real-ESRGAN sustituye el discriminador de tipo VGG-style por una arquitectura U-Net, con conexiones skip connections que facilitan la propagación de gradientes y reducen el problema de desvanecimiento. Esta estructura proporciona una realimentación por píxel, permitiendo al discriminador generar valores de “realidad” locales y entregar información más precisa al generador. De esta manera, el discriminador no solo diferencia estilos globales, sino también texturas locales con mayor fidelidad.

Para estabilizar el entrenamiento y mitigar los artefactos excesivamente agudos generados por la dinámica adversaria, se aplica regularización mediante normalización espectral. Además, se incorpora una función de pérdida perceptual, basada en características extraídas de una red VGG pre-entrenada, que optimiza la similitud visual entre las imágenes generadas y las de referencia de alta resolución. En conjunto, estas mejoras permiten que Real-ESRGAN produzca resultados más realistas, nítidos y visualmente coherentes, incluso cuando las imágenes de entrada provienen de degradaciones complejas o no modeladas idealmente, como es el caso de las imágenes satelitales de baja resolución empleadas en este trabajo.

2.5 Entrenamiento del modelo

Dado que Real-ESRGAN fue originalmente desarrollado y optimizado para imágenes de propósito general, y ha sido posteriormente adaptado a otros dominios, como imágenes médicas (Nandal et al., 2024) e imágenes degradadas (Zhu et al., 2023), en este trabajo se realizaron modificaciones específicas en el proceso de entrenamiento para ajustarlo al dominio de imágenes aéreas y satelitales. Para optimizar el entrenamiento y el uso de recursos computacionales, se ajustaron varios parámetros del modelo. El tamaño del batch se configuró en dos imágenes en la primera fase y en una en la fase optimizada, reduciendo el consumo de memoria. Se habilitó el uso de GPU para acelerar el proceso de entrenamiento mediante paralelización, y se fijaron 100.000 iteraciones para balancear tiempo de cómputo y calidad de resultados. El modelo utilizó funciones de pérdida combinadas (L1, perceptual y adversaria), mejorando tanto la reconstrucción estructural como la calidad visual de las imágenes generadas.

Tabla 1. Resumen de los hiperparámetros y configuraciones empleadas durante el entrenamiento del modelo ESRGAN

Parámetro	Valor	Parámetro	Valor
<i>Recursos de cómputo</i>		<i>Pérdidas y pesos</i>	
gpu_ids	0, 1, 2, 3, 4	pixel_criterion	L1
batch_size	1	pixel_weight	0.01
dev_ratio	0.01	feature_criterion	L1
<i>Hiperparámetros del generador (G)</i>		feature_weight	1
lr_G	1	gan_type	RaGAN
weight_decay_G	0	gan_weight	5
beta1_G	0.9	D_update_ratio	1
beta2_G	0.99	D_init_iters	0
<i>Hiperparámetros del discriminador (D)</i>		<i>Frecuencias de salida</i>	
lr_D	1	print_freq	100
weight_decay_D	0	val_freq	1000
beta1_D	0.9	save_freq	10000
beta2_D	0.99		
<i>Esquema de aprendizaje</i>		<i>Tamaños de recortes</i>	
lr_scheme	MultiStepLR	crop_size	0.85
niter	1000	lr_size	128
warmup_iter	-1	hr_size	512
lr_steps	[50000]		
lr_gamma	0.5		

El modelo fue entrenado utilizando los parámetros detallados en la Tabla 1, con una tasa de aprendizaje inicial de 1×10^{-4} para el generador y el discriminador, esquema MultiStepLR con decaimiento al llegar a las 50.000 iteraciones, y un total de 100.000 iteraciones de entrenamiento. Durante el entrenamiento se combinaron pérdidas de contenido (L1), perceptuales y adversariales (RaGAN), con pesos definidos para favorecer la nitidez sin sacrificar estabilidad.

El entrenamiento se realizó utilizando un servidor equipado con un procesador Intel Core i9-13900KF y una GPU NVIDIA RTX 3080 Ti de 12 GB de VRAM GDDR6 (ASUS TUF). El código desarrollado para este estudio, incluyendo los scripts de entrenamiento, evaluación y generación de resultados, se encuentra disponible públicamente en GitHub: <https://github.com/aletrujim/Super-Reso>.

2.6 Métodos de evaluación y comparación

Para evaluar el desempeño del modelo en la tarea de super-resolución, se utilizaron métricas cuantitativas y perceptuales ampliamente empleadas en el área de procesamiento de imágenes. Estas métricas permiten medir tanto la similitud estructural y la fidelidad visual entre las imágenes generadas y las imágenes de referencia, como la calidad percibida por el observador.

En primer lugar, se calculó la métrica SSIM (Structural Similarity Index Measure) sobre el canal de luminancia (Y) de las imágenes. SSIM evalúa la similitud estructural entre ambas imágenes, considerando luminancia, contraste y estructura. Esta métrica se calcula sobre el canal Y dado que es el más relevante para la percepción humana de la calidad visual (Wang et al., 2004). Además, se incorporó la métrica MS-SSIM (Multi-Scale Structural Similarity Index), que extiende SSIM a múltiples escalas espaciales. Esto permite evaluar la consistencia estructural y de contraste en diferentes niveles de detalle, proporcionando una medida más robusta de la calidad perceptual (Wang et al., 2003).

Para capturar diferencias perceptuales no lineales difíciles de reflejar mediante métricas tradicionales, se empleó la métrica LPIPS (Learned Perceptual Image Patch Similarity) (Zhang et al., 2018). LPIPS utiliza redes neuronales pre-entrenadas (en este caso, AlexNet) para comparar representaciones profundas entre la imagen generada y la de referencia. Valores más bajos indican mayor similitud perceptual.

Finalmente, se calculó el índice de diferencia de color CIEDE2000, basado en el espacio de color CIE Lab, que cuantifica las discrepancias cromáticas promedio entre ambas imágenes (Sharma et al., 2005). Esta métrica es especialmente útil para detectar variaciones en la reproducción del color, aspecto relevante en el contexto de imágenes aéreas y satelitales.

Estas cuatro métricas permiten una evaluación complementaria y equilibrada del rendimiento del modelo, considerando tanto la precisión estructural como la calidad visual percibida en las imágenes generadas. Para la presentación de los resultados, se evaluaron nueve recortes de imágenes aéreas y nueve recortes de imágenes satelitales, seleccionados para representar diferentes texturas, contrastes y niveles de detalle.

3 Resultados

En esta sección se presentan los resultados obtenidos durante el proceso de entrenamiento y evaluación del modelo Real-ESRGAN adaptado al dominio de imágenes aéreas y satelitales. Se incluye tanto un análisis cualitativo del rendimiento del modelo a lo largo de distintas etapas del entrenamiento como una evaluación cuantitativa basada en métricas estandarizadas de calidad de imagen. La combinación de ambos enfoques permite valorar de manera integral la capacidad del modelo para recuperar detalles de alta resolución y reproducir estructuras espaciales relevantes del terreno.

3.1 Rendimiento del modelo

Para evaluar la evolución del aprendizaje y el efecto de los checkpoints sobre la calidad de las reconstrucciones, se generaron salidas intermedias en tres momentos claves del entrenamiento: 5.000, 50.000 y 100.000 iteraciones. Las imágenes de tipo display (ver Figura 3) permiten visualizar el progreso del modelo al reconstruir los recortes de entrada (LQ) y compararlos tanto con la versión resultante del modelo de super-resolución (SR) como con el ground truth (GT).

En las primeras iteraciones (5.000), las imágenes SR presentan bordes suavizados y un nivel de detalle aún incipiente, lo que es coherente con una etapa temprana de aprendizaje en la que predomina el ajuste de la pérdida de contenido. A las 50.000 iteraciones, ya se observa una mejora notable en la recuperación de texturas, patrones agrícolas y límites entre parcelas, producto de la influencia progresiva de las pérdidas perceptuales y adversariales. Finalmente, en el checkpoint de 100.000 iteraciones, el modelo alcanza reconstrucciones más nítidas y coherentes con la estructura del terreno observada en el GT, ofreciendo detalles más finos en caminos, edificaciones y líneas de cultivo. Estas observaciones cualitativas reflejan la convergencia gradual del modelo y el aporte complementario de cada componente de la función de pérdida. Los resultados cuantitativos se presentan en las siguientes subsecciones.

3.2 Evaluación del modelo con imágenes aéreas

La evaluación cuantitativa del modelo evidencia diferencias claras en el rendimiento entre las escalas de ampliación $\times 2$ y $\times 4$. Para la escala $\times 2$, las métricas estructurales presentan los valores más altos: SSIM-Y y MS-SSIM muestran mayores niveles de coherencia estructural y preservación de luminancia con respecto a las imágenes de referencia, indicando que el modelo logra mantener de forma consistente la geometría de objetos urbanos, la continuidad de bordes y la textura general de superficies (ver Figura 4, Tabla 2).

En términos perceptuales, los valores de LPIPS son significativamente más bajos, lo que sugiere que la reconstrucción posee menor divergencia perceptual respecto al ground truth, especialmente en regiones complejas como áreas vegetadas o límites entre infraestructura y suelo expuesto. La métrica CIEDE2000, asociada a diferencias cromáticas, también presenta menores valores en $\times 2$, evidenciando que el modelo mantiene mejor la coherencia de color y evita derivaciones tonales no deseadas. Esto se observa especialmente en zonas donde pequeños desbalances suelen amplificarse, como sombras, techos metálicos o vegetación con variaciones espectrales. Para la escala $\times 4$, todas las métricas muestran un rendimiento ligeramente inferior, lo cual es

esperable dada la mayor dificultad de reconstruir información a partir de una resolución base más reducida. En esta escala, se observa un descenso moderado en SSIM-Y y MS-SSIM, indicando una menor, aunque aún estable, capacidad para recuperar estructuras finas.



Fig. 4. Comparación visual del desempeño del modelo en distintos checkpoints (5.000, 50.000 y 100.000), utilizada para evaluar la estabilidad del entrenamiento y la calidad de la reconstrucción.

El aumento en los valores de LPIPS refleja que las diferencias perceptuales se vuelven más evidentes, especialmente en texturas altamente detalladas. Asimismo, CIEDE2000 registra un incremento leve, señalando una mayor sensibilidad del modelo a variaciones cromáticas cuando se requiere extrapolar detalles no presentes

en la imagen de entrada. A pesar de estas diferencias, los resultados en x4 continúan siendo consistentes y visualmente plausibles (ver Figura 5). El modelo demuestra una capacidad robusta para generar mejoras notables incluso en escenarios de reconstrucción más exigentes, manteniendo una relación equilibrada entre fidelidad estructural, percepción visual y estabilidad cromática.

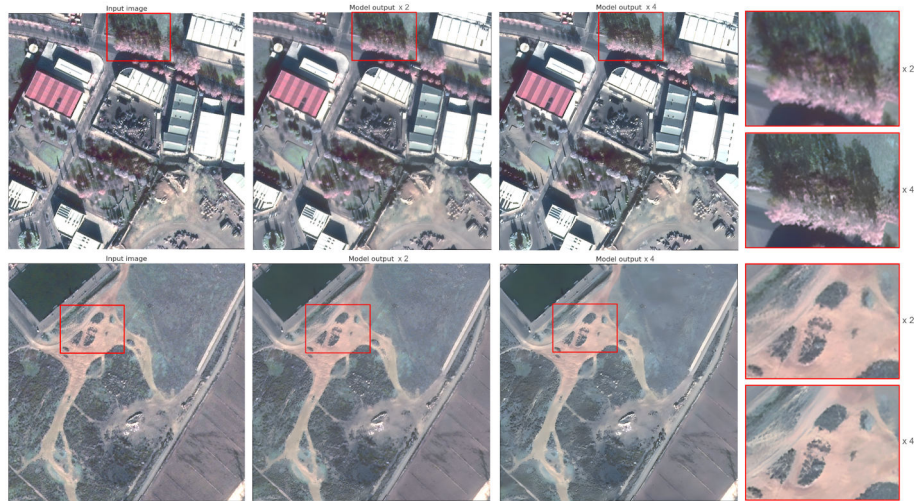


Fig. 5. Ejemplos de super-resolución para imágenes aéreas utilizando el modelo propuesto con factores de escala x2 y x4. Se muestra la imagen de entrada y las salidas generadas por el modelo, junto con ampliaciones de regiones de interés (recuadros rojos) que permiten apreciar la mejora en el nivel de detalle y la reconstrucción de texturas finas.

Tabla 2. Estadísticos descriptivos de las métricas de evaluación calculadas sobre las imágenes aéreas utilizadas para validar el modelo.

Escala	Métrica	Media	Desvío Std	Mínimo	Máximo
x2	SSIM-Y	0,812	0,044	0,740	0,865
	MS-SSIM	0,969	0,005	0,961	0,977
	LPIPS	0,137	0,026	0,098	0,171
	CIEDE2000	3,206	0,334	2,382	3,557
x4	SSIM-Y	0,779	0,064	0,701	0,866
	MS-SSIM	0,964	0,012	0,948	0,983
	LPIPS	0,187	0,044	0,111	0,255
	CIEDE2000	3,979	0,420	3,457	4,755

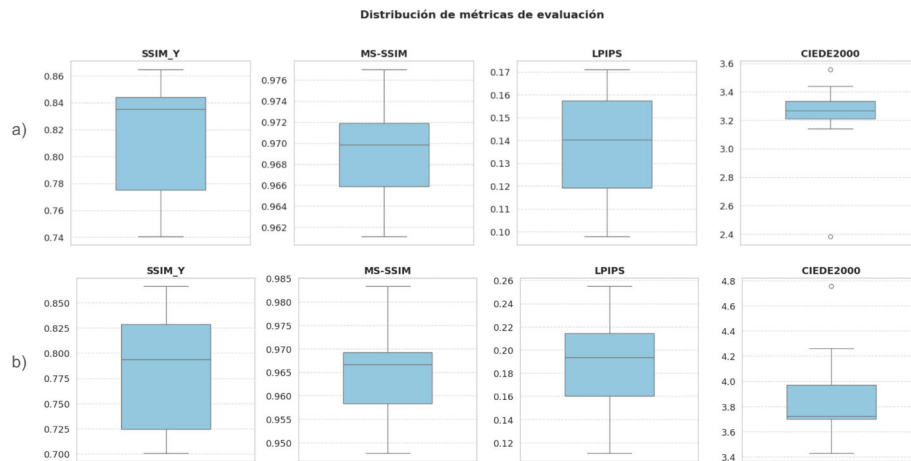


Fig. 6. Distribución estadística de las métricas de evaluación para los tres recortes de imágenes aéreas, ilustrada mediante diagramas de caja que permiten comparar la consistencia del modelo. (a) Resultados para escala de super-resolución x2. (b) Resultados para escala de super-resolución x4. Se presentan las distribuciones de SSIM-Y y MS-SSIM (mayores valores indican mejor preservación estructural), junto con LPIPS y CIEDE2000 (menores valores reflejan menor error perceptual y cromático).

3.3 Evaluación del modelo en imágenes satelitales

La evaluación del modelo sobre los recortes de imágenes satelitales demuestra que, aun habiendo sido entrenado únicamente con imágenes aéreas, es capaz de mejorar de forma notable la resolución y la calidad visual de las imágenes satelitales. Para ambas escalas de super-resolución se observa un comportamiento estable y coherente, con mejoras perceptibles en la reconstrucción de detalles (ver Figura 6, tabla 3). Las imágenes generadas presentan contornos más definidos y una mayor nitidez en comparación con las versiones de baja resolución, especialmente en estructuras urbanas como edificios, techos y calles, y en elementos del terreno como áreas de vegetación.

Para la escala x2, el desempeño general es sólido: los valores de SSIM-Y se encuentran mayormente en el rango 0.66 – 0.85, indicando una buena recuperación estructural del contenido original. Esta tendencia se refuerza con los valores de MS-SSIM, que presentan poca variabilidad y se mantienen por encima de 0.96 en casi todos los casos, lo que refleja una preservación estable de estructuras a múltiples escalas. En términos perceptuales, LPIPS muestra valores bajos ($\approx 0.09 - 0.34$), con una mediana alrededor de 0.18, lo cual evidencia una mejora perceptible en texturas y bordes. Por último, los errores cromáticos evaluados mediante CIEDE2000 se concentran entre 2 y 3.2, con un único caso atípico cercano a 5, asociado a variaciones más marcadas en regiones de alta heterogeneidad espectral.

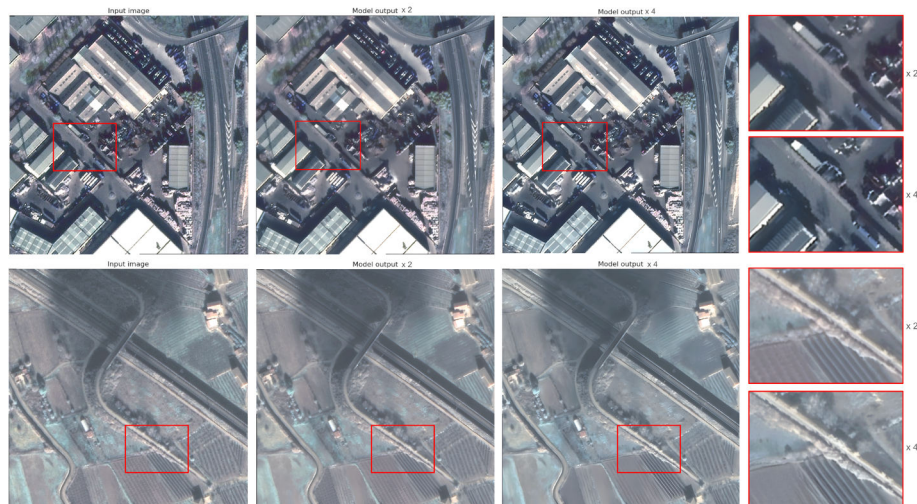


Fig. 7. Comparación visual entre recortes de imágenes satelitales de entrada y las generadas por el modelo de super-resolución para factores de escala x2 y x4. Los recuadros rojos indican las áreas ampliadas a la derecha, donde puede observarse la mejora en la definición de bordes, texturas y estructuras finas lograda por el modelo en ambos niveles de escala.

Tabla 3. Estadísticos descriptivos de las métricas de evaluación calculadas sobre las imágenes satelitales utilizadas para validar el modelo.

Escala	Métrica	Media	Desvío Std	Mínimo	Máximo
x2	SSIM-Y	0,755	0,096	0,577	0,852
	MS-SSIM	0,965	0,030	0,887	0,980
	LPIPS	0,180	0,084	0,092	0,342
	CIEDE2000	2,842	1,033	1,807	5,046
x4	SSIM-Y	0,729	0,120	0,519	0,847
	MS-SSIM	0,970	0,009	0,951	0,982
	LPIPS	0,216	0,082	0,107	0,319
	CIEDE2000	3,162	0,333	2,625	3,623

En la escala x4, como es esperable, se observa una ligera disminución del rendimiento debido a la mayor complejidad de reconstrucción desde resoluciones más bajas. SSIM-Y muestra un rango más amplio ($\approx 0.52 - 0.85$), aunque el modelo mantiene una buena recuperación estructural en la mayoría de los recortes. La métrica MS-SSIM continúa siendo estable ($\approx 0.95 - 0.98$), lo que sugiere que la pérdida estructural es acotada incluso a mayor escala. Los valores de LPIPS evidencian una degradación leve respecto a x2 ($\approx 0.11 - 0.32$), pero aún dentro de rangos considerados razonables para super-resolución x4. En cuanto a CIEDE2000, la

variabilidad aumenta ligeramente ($\approx 2.6 - 3.6$), reflejando el desafío adicional de mantener coherencia cromática al cuadruplicar la resolución.

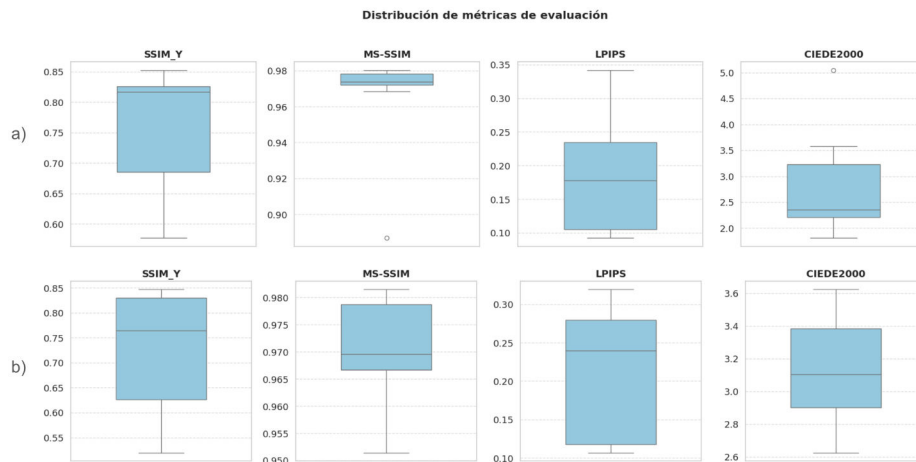


Fig. 8. Distribución de las métricas de evaluación calculadas sobre recortes de imágenes satelitales procesadas por el modelo. (a) Resultados para escala de super-resolución x2. (b) Resultados para escala de super-resolución x4. Las tendencias observadas muestran un rendimiento estable en ambas escalas, con una degradación moderada al pasar de x2 a x4, consistente con el aumento de la complejidad de la reconstrucción.

4 Discusión y Conclusiones

Los resultados obtenidos en este estudio revelan que el modelo de super-resolución entrenado exclusivamente con imágenes aéreas capturadas por drones es capaz de mejorar de forma consistente la calidad de imágenes satelitales de baja resolución, incluso a pesar de las diferencias en perspectiva, escala y condiciones de adquisición entre ambos tipos de datos. Tanto en la escala x2 como en x4, el modelo demuestra una capacidad estable para recuperar información estructural, aumentar la nitidez y mejorar la percepción visual de los recortes satelitales evaluados.

En términos cuantitativos, los valores de SSIM-Y y MS-SSIM muestran que el modelo mantiene una estructura interna coherente en las imágenes reconstruidas. La escala x2 alcanza los mejores resultados, lo que era esperable debido a la menor complejidad del proceso de reconstrucción. No obstante, la escala x4 presenta un rendimiento competitivo, preservando un equilibrio adecuado entre recuperación estructural, calidad perceptual (LPIPS) y estabilidad cromática (CIEDE2000), tal como reflejan las distribuciones de métricas y los valores estadísticos resumidos. Estas observaciones coinciden con la evaluación visual: las imágenes super-resueltas muestran una mayor claridad en estructuras urbanas como edificios, calles y techos, así como en detalles del terreno y la vegetación.

Un aspecto relevante es que, pese a no haber sido entrenado con imágenes satelitales, el modelo generaliza adecuadamente a este nuevo dominio. Esto sugiere que las características aprendidas a partir de imágenes aéreas como bordes, texturas urbanas y

patrones espaciales, son transferibles y útiles para mejorar datos satelitales. Sin embargo, también se identifican limitaciones asociadas a la diferencia espectral, ya que las imágenes aéreas suelen tener un espectro visible más uniforme, mientras que las imágenes satelitales pueden incorporar variaciones radiométricas propias del sensor. Estas diferencias pueden explicar algunas imperfecciones cromáticas o pequeñas inconsistencias observadas en ciertos escenarios.

Como líneas de trabajo futuras, se propone ampliar el conjunto de datos incorporando imágenes de distintas regiones geográficas, estaciones del año y condiciones atmosféricas, con el fin de mejorar la capacidad de generalización del modelo y reducir su dependencia del dominio aéreo original. Asimismo, resulta pertinente explorar arquitecturas alternativas de super-resolución, incluyendo modelos recientes basados en transformers u otros enfoques que puedan optimizar la recuperación de detalles finos en imágenes satelitales. Finalmente, se sugiere evaluar técnicas de domain adaptation que permitan ajustar de manera más precisa las representaciones internas del modelo a las características espectrales, geométricas y radiométricas propias de cada tipo de sensor, contribuyendo así a mejorar el desempeño en escenarios reales y heterogéneos.

Bibliografía

- Liu, T., & Yang, X. (2015). Monitoring land changes in an urban area using satellite imagery, GIS and landscape metrics. *Applied geography*, 56, 42-54.
- Wang, D., Shao, Q., & Yue, H. (2019). Surveying wild animals from satellites, manned aircraft and unmanned aerial systems (UASs): A review. *Remote Sensing*, 11(11), 1308.
- Shan, J., Weng, Q., Ehlers, M., Quattrochi, D. A., Zhou, G., Stilla, U., ... & Anderson, S. (2018). *Urban remote sensing*. CRC press.
- Yang, W., Zhang, X., Tian, Y., Wang, W., Xue, J. H., & Liao, Q. (2019). Deep learning for single image super-resolution: A brief review. *IEEE Transactions on Multimedia*, 21(12), 3106-3121.
- K. Karwowska and D. Wierzbicki, "Using Super-Resolution Algorithms for Small Satellite Imagery: A Systematic Review," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 3292-3312, 2022, doi: 10.1109/JSTARS.2022.3167646.
- Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., ... & Shi, W. (2017). Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4681-4690).
- Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., ... & Change Loy, C. (2018). Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European conference on computer vision (ECCV) workshops* (pp. 0-0).
- Wang, X., Xie, L., Dong, C., & Shan, Y. (2021). Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 1905-1914).
- Institut Cartogràfic i Geològic de Catalunya ICGC. (s.f.). Ortofoto 25 cm v7r0 color. Infraestructura de Dades Espacials de Catalunya (IDE).

Arabboev, M., Begmatov, S., Rikhsivoev, M., Nosirov, K., & Saydiakbarov, S. (2024). A comprehensive review of image super-resolution metrics: classical and AI-based approaches. *Acta IMEKO*, 13(1), 1-8.

X. Wang, L. Xie, C. Dong, and Y. Shan, "Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2021, pp. 1905–1914.

Nandal, P., Pahal, S., Khanna, A., & Pinheiro, P. R. (2024). Super-resolution of medical images using real ESRGAN. *IEEE Access*.

Zhu, Z., Lei, Y., Qin, Y., Zhu, C., & Zhu, Y. (2023). IRE: improved image super-resolution based on real-ESRGAN. *IEEE Access*, 11, 45334-45348.

Abel Nájera Ramos, I., Vázquez Jiménez, R., Rocío Ramos Bernal, D. N., Gloria Rojas Sánchez, I., & Ana Ma Liborio Vicente, I. (2019). Propuesta Metodológica Para La Generación De Ortofotos Y Modelos Digitales De Elevación De Alta Resolución A Través De Vuelos Con Drones No fotogramétricos.

Fernández-Lozano, J., & Gutiérrez-Alonso, G. (2016). Aplicaciones geológicas de los drones. *Revista de la Sociedad Geológica de España*, 29(1), 89-105.

Garcia Diaz, L. (2023). Pueblos abandonados de Cataluña. Análisis espacial de su dinámica territorial y de los cambios en sus edificaciones.

Li, Y., Zhang, Y., Timofte, R., Van Gool, L., Yu, L., Li, Y., ... & Wang, X. (2023). NTIRE 2023 challenge on efficient super-resolution: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 1922-1960).

Lanaras, C., Bioucas-Dias, J., Galliani, S., Baltsavias, E., & Schindler, K. (2018). Super-resolution of Sentinel-2 images: Learning a globally applicable deep neural network. *ISPRS Journal of Photogrammetry and Remote Sensing*, 146, 305-319.

Kang, Y., Jang, E., Im, J., & Kwon, C. (2022). A deep learning model using geostationary satellite data for forest fire detection with reduced detection latency. *GIScience & Remote Sensing*, 59(1), 2019-2035.

Wang, S., Guan, K., Zhang, C., Zhou, Q., Wang, S., Wu, X., ... & Ma, Z. (2023). Cross-scale sensing of field-level crop residue cover: Integrating field photos, airborne hyperspectral imaging, and satellite data. *Remote Sensing of Environment*, 285, 113366.

Galar, M., Sesma, R., Ayala, C., Albizua, L., & Aranda, C. (2020). Super-resolution of sentinel-2 images using convolutional neural networks and real ground truth data. *Remote Sensing*, 12(18), 2941.

Salgueiro, L., Marcello, J., & Vilaplana, V. (2021). Single-image super-resolution of Sentinel-2 low resolution bands with residual dense convolutional neural networks. *Remote Sensing*, 13(24), 5007.

Vivone, G., Deng, L. J., Deng, S., Hong, D., Jiang, M., Li, C., ... & Plaza, A. (2024). Deep learning in remote sensing image fusion: Methods, protocols, data, and future perspectives. *IEEE Geoscience and Remote Sensing Magazine*.

Dong, C., Loy, C. C., & Tang, X. (2016, September). Accelerating the super-resolution convolutional neural network. In *European conference on computer vision* (pp. 391-407). Cham: Springer International Publishing.

Lim, B., Son, S., Kim, H., Nah, S., & Mu Lee, K. (2017). Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 136-144).

Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., & Fu, Y. (2018). Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European conference on computer vision (ECCV)* (pp. 286-301).

Liang, J., Zeng, H., & Zhang, L. (2022, October). Efficient and degradation-adaptive network for real-world image super-resolution. In European Conference on Computer Vision (pp. 574-591). Cham: Springer Nature Switzerland.

Pineda, F., Ayma, V., & Beltran, C. (2020). A generative adversarial network approach for super-resolution of Sentinel-2 satellite images. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43, 9-14.