Implementación del Modelo Real-ESRGAN para Imágenes Satelitales

Jhoe Gabriel Chirinos Sullcany

2 de julio de 2024

Resumen–Este estudio se centra en implementar un sistema de super-resolución rápido, eficiente y fácil de usar, específicamente diseñado para imágenes satelitales. Mediante el uso del modelo Real-ESRGAN, nuestro objetivo es optimizar la arquitectura, ajustar los parámetros y validar la efectividad a través de un riguroso proceso de entrenamiento y evaluación con conjuntos de datos reales. Buscamos mejorar significativamente la calidad de las imágenes, ofreciendo soluciones escalables que pueden aplicarse en diversas áreas, como el análisis de imágenes satelitales, el monitoreo ambiental, la planificación urbana, la gestión de desastres y la seguridad nacional. Este estudio aspira a potenciar avances en la toma de decisiones y eficiencias operativas mediante imágenes satelitales más nítidas y detalladas.

Palabras clave–Super-resolución de imágenes satelitales, Real-ESRGAN, Validación de rendimiento, Mejora de calidad de imagen, Red Generativa Adversarial

Abstract–This study focuses on implementing a fast, efficient, and user-friendly super-resolution system specifically designed for satellite images. Utilizing the Real-ESRGAN model, our objective is to optimize the architecture, adjust the parameters, and validate the effectiveness through a rigorous process of training and evaluation with real datasets. We aim to significantly enhance image quality, providing scalable solutions applicable in various areas such as satellite image analysis, environmental monitoring, urban planning, disaster management, and national security. This study aspires to advance decision-making and operational efficiencies through sharper and more detailed satellite images.

Keywords–Satellite image super-resolution, Real-ESRGAN, Performance validation, Image quality enhancement, Generative Adversarial Network

1 Introducción

La super-resolución [1] de imágenes es un campo en expansión que busca mejorar la calidad de las imágenes de baja resolución aumentando el número de píxeles en la imagen resultante. Sin embargo, este aumento no se realiza de manera aleatoria; se lleva a cabo de manera inteligente con el objetivo de recuperar la mayor cantidad de información posible para que la imagen final se asemeje lo más posible a la realidad. Este enfoque busca hacer que las imágenes sean más nítidas y claras, ya que en aplicaciones como la monitorización del medio ambiente, la inteligencia militar y la cartografía, la distinción de objetos pequeños es crucial.

En el contexto de imágenes satelitales [2], cuya resolución espacial limitada presenta desafíos adicionales, la super-resolución se convierte en una herramienta indispensable para mejorar la calidad de la información visual obtenida. Las imágenes satelitales son capturas de la Tierra o de otros cuerpos celestes obtenidas desde satélites artificiales en órbita. Son utilizadas en una amplia gama de aplicaciones debido a su capacidad para proporcionar una vista global y detallada de la superficie terrestre.

Buscamos que estas imágenes permitan una evaluación precisa de la salud de los ecosistemas, el seguimiento del crecimiento de los cultivos, la respuesta efectiva a desastres, la planificación urbana sostenible y la vigilancia estratégica. Las ventajas de la super-resolución de imágenes satelitales son significativas. No solo mejora la calidad de las

• E-mail de contacto: 1601291@uab.cat
• Mención realizada: Computación
• Trabajo tutoritzado por: Robert Benavente Vidal (Departamento de Ciencias de la Computación)
• Curs 2023/24

imágenes, permitiendo obtener imágenes más nítidas y claras, sino que también aumenta la información disponible al capturar más detalles y características de interés. Además, representa una alternativa más económica en comparación con la adquisición de nuevas imágenes de alta resolución, lo que resulta en una reducción de costos a la hora de tratar estas imágenes.

Sin embargo, la super-resolución de imágenes satelitales también presenta algunos desafíos. Entre ellos se encuentran la generación de artefactos, como halos o bordes irregulares, durante el proceso de super-resolución. Existe también el riesgo de pérdida de información importante en la imagen en ciertos casos. Además, el costo computacional puede ser un desafío significativo, especialmente para imágenes de gran tamaño. Estos desafíos deben abordarse de manera efectiva para garantizar que los beneficios de la super-resolución superen sus limitaciones.

En este contexto, nuestro objetivo es desarrollar un sistema de super-resolución de imágenes satelitales que sea rápido, eficiente y fácil de usar en diversas situaciones. Este sistema no solo mejorará la calidad de las imágenes satelitales, sino que también tendrá un impacto positivo en una amplia gama de aplicaciones. Desde la creación de mapas más detallados y precisos en cartografía hasta la evaluación del estado de los cultivos y la detección de contaminación en el medio ambiente, nuestra tecnología de super-resolución ofrecerá soluciones efectivas y versátiles para una variedad de desafíos en múltiples campos.

2 Estado del Arte

En esta sección, llevaremos a cabo una investigación exhaustiva de técnicas en el ámbito de la super-resolución de imágenes. Analizaremos diversas iniciativas para comprender los enfoques y métodos utilizados en el aumento de la resolución de imágenes. Este análisis nos permitirá establecer de manera precisa los objetivos de nuestro proyecto.

Dividiremos las técnicas más populares y efectivas en dos campos generales según en qué están basadas.

2.1 Técnicas Basadas en Interpolación

Las técnicas basadas en interpolación [3] son métodos tradicionales para aumentar la resolución de una imagen. Aunque son menos sofisticadas que las técnicas basadas en aprendizaje automático, son rápidas y fáciles de implementar.

- **Interpolación Bicúbica**
  La Interpolación Bicúbica [3] utiliza una fórmula polinómica para estimar nuevos píxeles a partir de los valores de píxeles vecinos. Es una técnica más avanzada que la interpolación bilineal y proporciona resultados más suaves.
  Ventajas : Buena calidad para ampliaciones moderadas.
  Limitaciones : Puede generar artefactos de suavizado y no es eficaz para ampliaciones grandes.

- **Interpolación de Vecino Más Cercano [3]**
  La Interpolación de Vecino Más Cercano [3] asigna a cada nuevo píxel el valor del píxel más cercano en la imagen original. Es el método más simple y rápido.
  Ventajas : Fácil de implementar y computacionalmente eficiente.
  Limitaciones : Resultados en imágenes pixeladas y con baja calidad visual.

- **EBSR (Evidence-Based Super Resolution [3])**
  La EBSR (Evidence-Based Super Resolution [3]) utiliza evidencia de múltiples imágenes de baja resolución para generar una imagen de alta resolución. Este enfoque se basa en la fusión de información redundante presente en las imágenes de entrada para reconstruir detalles precisos en la imagen de salida.
  Ventajas : Capacidad para generar detalles precisos al combinar información de múltiples imágenes.
  Limitaciones : Requiere varias imágenes de entrada y puede ser computacionalmente intensivo.

2.2 Técnicas Basadas en Aprendizaje Automático

2.2.1 Redes Neuronales Convolucionales(CNN)

Las redes neuronales convolucionales (CNN) [4] son una clase de redes neuronales especialmente adecuadas para el procesamiento de datos con una estructura de cuadrícula, como las imágenes. Las CNN utilizan capas convolucionales que aplican filtros sobre las imágenes para extraer características importantes, y son especialmente eficaces para tareas de visión por computadora, como la super-resolución.

- **SRCNN (Super-Resolution Convolutional Neural Network)**
  SRCNN [5] es una de las primeras redes neuronales convolucionales aplicadas a la super-resolución de imágenes. Consiste en tres capas: una capa de convolución para extracción de características, una capa de no linealidad para mapear las características extraídas a un espacio de alta resolución y una capa final para reconstruir la imagen de alta resolu-ción.
  Ventajas : Simple y efectiva para mejoras básicas de resolución.
  Limitaciones : La arquitectura simple limita la capacidad para capturar detalles muy finos.

- **VDSR (Very Deep Super Resolution)**
  VDSR [6] es una red más profunda que SRCNN, con 20 capas convolucionales. Utiliza técnicas de aprendizaje profundo para mejorar la capacidad de la red para capturar detalles de alta frecuencia.
  Ventajas : Mejor rendimiento en comparación con SRCNN.
  Limitaciones : Mayor complejidad computacional y tiempo de entrenamiento.

- **EDSR (Enhanced Deep Super Resolution)**
  EDSR [7] mejora VDSR eliminando las capas de normalización batch (Batch Normalization), lo cual simplifica la red y mejora su rendimiento. Utiliza una arquitectura residual profunda para lograr resultados de

alta calidad.
Ventajas : Alto rendimiento en competiciones de super-resolución y gran capacidad para capturar detalles finos.
Limitaciones : Alta demanda computacional.

2.2.2 Redes Generativas Adversativas (GAN)

Las GAN [8] son una clase de modelos de aprendizaje automático compuesta por dos redes neuronales: un generador y un discriminador, que compiten entre sí. El generador crea imágenes de alta resolución a partir de imágenes de baja resolución, mientras que el discriminador trata de distinguir entre imágenes reales y generadas. Este enfoque permite a las GAN producir imágenes más realistas.

- **SRGAN (Super-Resolution Generative Adversarial Network)**
  SRGAN [9] es una red que utiliza una arquitectura GAN para generar imágenes de alta resolución a partir de imágenes de baja resolución. El generador está basado en una arquitectura CNN, mientras que el discriminador evalúa la autenticidad de las imágenes generadas.
  Ventajas : Capacidad para generar detalles de alta calidad y mejorar la percepción visual.
  Limitaciones : La estabilidad del entrenamiento de GAN puede ser un desafío.

- **ESRGAN (Enhanced Super-Resolution Generative Adversarial Network)**
  ESRGAN [10] es una versión mejorada de SRGAN, con una arquitectura residual y módulos de atención que permiten un mejor rendimiento. ESRGAN logra imágenes más nítidas y detalladas en comparación con SRGAN.
  Ventajas : Mejor calidad de imagen y detalles finos.
  Limitaciones : Requiere un entrenamiento complejo y ajustado.

- **ProGAN (Progressive Growing of GANs)**
  ProGAN [11] es una técnica que entrena redes GAN de manera progresiva, comenzando con una resolución baja y aumentando gradualmente hasta alcanzar la resolución deseada. Esto mejora la estabilidad del entrenamiento y la calidad de las imágenes generadas.
  Ventajas : Estabilidad mejorada y calidad de imagen superior.
  Limitaciones : Mayor tiempo de entrenamiento debido al enfoque progresivo.

2.2.3 Redes Basadas en Transformadores

Los transformadores [12] son una arquitectura que ha demostrado ser muy eficaz en tareas de procesamiento de secuencias, y recientemente se han adaptado para tareas de visión por computadora, incluyendo la super-resolución.

- **SwinIR (Swin Transformer for Image Restoration)**
  SwinIR [13] utiliza el Swin Transformer, una arquitectura basada en transformadores para visión por computadora. SwinIR aplica esta arquitectura a la tarea de restauración de imágenes, incluyendo super-resolución.
  Ventajas : Capacidad de modelar relaciones de largo alcance en imágenes y alta calidad de restauración
  Limitaciones : Complejidad computacional y requerimientos de memoria.

- **DeiT (Data-efficient Image Transformers)**
  Aunque originalmente diseñado para clasificación de imágenes, DeiT [14] se puede adaptar para super-resolución aplicando técnicas de aprendizaje eficiente con pocos datos y utilizando la arquitectura de transformadores.
  Ventajas : Eficiencia en el uso de datos y alta capacidad de generalización.
  Limitaciones : Necesidad de adaptar el modelo para tareas específicas de visión por computadora.

3 MODELO REAL-ESRGAN

En este apartado nos centraremos en explicar más en detalle el modelo que escogimos para su implementación, que sería el Real-ESRGAN [15] y las razones que lo hacen relevante para nuestro proyecto.

3.1 Descripción General del Modelo

Real-ESRGAN ha sido meticulosamente diseñado para abordar los desafíos únicos de mejora de imágenes del mundo real, las cuales frecuentemente exhiben una variedad de imperfecciones como ruido y artefactos de compresión. Esta variante mejorada de ESRGAN está específicamente adaptada para elevar la calidad de imágenes reales, diferenciándose de ESRGAN que se centra principalmente en mejorar imágenes de baja resolución.

Un aspecto distintivo de Real-ESRGAN radica en su capacidad para adaptarse a una variedad de escenarios prácticos. Desde la restauración de fotografías antiguas hasta la mejora de imágenes de vigilancia y la recuperación de detalles en imágenes médicas, su flexibilidad lo convierte en una herramienta invaluable. Al abordar de manera efectiva las degradaciones inherentes a estas imágenes del mundo real, Real-ESRGAN se destaca por producir resultados de alta calidad que superan notablemente la percepción visual.

Las razones principales para seleccionar Real-ESRGAN para nuestro proyecto son su robustez y versatilidad. Su capacidad para mejorar significativamente la calidad de imágenes degradadas lo hace ideal para aplicaciones críticas donde la precisión y el detalle son esenciales. Además, su eficacia en diversos contextos prácticos asegura que el modelo pueda ser utilizado en múltiples campos, aportando mejoras sustanciales en la calidad de las imágenes procesadas.

3.2 Arquitectura de Real-ESRGAN

En la figura 1 podemos observar la arquitectura de Real-ESRGAN, conocida como Arquitectura Residual con Atención, se basa en el éxito de su predecesora, ESRGAN, pero introduce varias mejoras para aumentar la calidad de las imágenes generadas. A continuación, se describen los componentes clave y sus funciones:

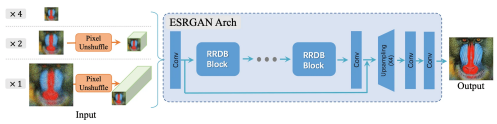

Fig. 1: Red generadora del modelo Real-ESRGAN, muy parecida a la arquitectura de su red predecesora ESRGAN.
Imagen tomada de https://arxiv.org/pdf/2107.10833

3.2.1 Generador

El generador en Real-ESRGAN es una red neuronal profunda que utiliza mecanismos avanzados para producir imágenes de alta resolución a partir de imágenes de baja resolución. Los elementos clave de su arquitectura son:

- **Bloques Densos Residuales en Residuales (RRDB)** Utilizan conexiones residuales para facilitar el flujo de gradientes durante el entrenamiento, lo que permite entrenar redes más profundas sin problemas de desvanecimiento del gradiente. Como podemos ver en la parte izquierda de la figura 2 cada bloque residual contiene varias capas convolucionales seguidas de una activación no lineal. Cuantas mas capas y conexiones haya entre los bloques mas aumenta el rendimiento del modelo.

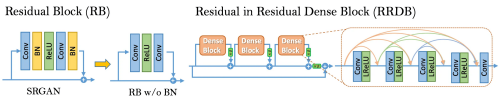

Fig. 2: Imagen que muestra cómo se elimina la Batch Normalization de los bloques usados en ESRGAN, seguida de la estructura de los bloques densos residuales en residuales.
Imagen tomada de https://arxiv.org/pdf/2107.10833

- **Upsampling** Utiliza técnicas avanzadas de upsampling, como convoluciones transpuestas y sub-pixel convolutions, para aumentar la resolución de la imagen de manera eficiente y efectiva, evitando artefactos comunes en métodos de upsampling más simples.

3.2.2 Discriminador

El discriminador en Real-ESRGAN se emplea para diferenciar entre imágenes reales y generadas dentro de un marco adversarial. Este está basado en la arquitectura U-Net, reconocida por su efectividad en tareas como segmentación y generación de imágenes. Podemos ver una represtación visual en la figura 3.

- **Red Convolucional Profunda:** Utiliza múltiples capas convolucionales para analizar y extraer características a diferentes niveles de la imagen, desde patrones locales hasta estructuras globales. Cada capa convolucional se acompaña de activaciones no lineales (como Leaky ReLU) y técnicas de normalización (como Spectral Normalization) para mejorar la estabilidad del entrenamiento.

- **Mecanismos de Atención:** Similar al generador, el discriminador también puede incorporar mecanismos de atención para enfocarse en características críticas de la imagen, mejorando su capacidad para distinguir entre imágenes reales y generadas.

- **Pérdidas Adversariales:** Utiliza la pérdida adversarial clásica basada en GANs (Generative Adversarial Networks), donde el objetivo es maximizar la capacidad del discriminador para distinguir entre imágenes reales y generadas, mientras que el generador intenta minimizar esta capacidad.

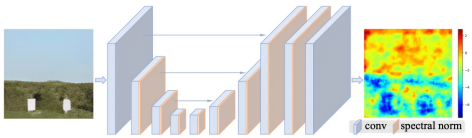

Fig. 3: Arquitectura del discriminador U-Net seguida de la normalización espectral
Imagen tomada de https://arxiv.org/pdf/2107.10833

4 Objetivos del Proyecto

En el desarrollo de este proyecto, nos proponemos alcanzar una serie de objetivos clave que asegurarán el éxito de nuestra investigación. Estos objetivos abarcan desde la configuración inicial del entorno hasta la validación final del modelo con imágenes satelitales. A continuación, se detallan los objetivos específicos que debemos cumplir:

- **Implementación del Modelo Real-ESRGAN** Configurar el entorno de desarrollo y asegurar la disponibilidad de las bibliotecas y dependencias necesarias para trabajar con el modelo Real-ESRGAN.

- **Entrenamiento y Validación con Dataset Conocido** Utilizar un dataset conocido, idealmente uno ampliamente utilizado y bien documentado en el campo de la superresolución de imágenes, para entrenar el modelo Real-ESRGAN y validar los resultados obtenidos. Esto garantizará que la implementación y el entrenamiento sean correctos y produzcan resultados esperados.

- **Implementación y Mejora de Parámetros del Modelo Real-ESRGAN** Realizar ajustes en la arquitectura de la red neuronal, optimizar hiperparámetros y explorar técnicas de preprocesamiento de datos para maximizar la eficacia en la mejora de la calidad de las imágenes.

- **Evaluación de la Efectividad del Modelo Real-ESRGAN**
  Utilizar métricas objetivas como MSE [16], PSNR [17] o SSIM [18] para evaluar el rendimiento del modelo. Además, llevar a cabo evaluaciones visuales comparativas entre las imágenes originales y las generadas por el modelo resultante para garantizar una evaluación completa y precisa de su desempeño.

- **Entrenamiento con Imágenes Satelitales**
  Una vez validada la implementación con el dataset conocido, proceder con el entrenamiento del modelo utilizando conjuntos de datos de imágenes satelitales de baja resolución y sus contrapartes de alta resolución.

- **Validación de Resultados con Imágenes Satelitales**
  Realizar una validación exhaustiva de los resultados del modelo entrenado con imágenes satelitales para garantizar su eficacia en la mejora de la calidad de estas imágenes. Esto incluirá la evaluación de métricas objetivas y evaluaciones visuales comparativas.

5 Experimento Inicial

A continuación, detallamos los pasos seguidos para la implementación, entrenamiento y validación del modelo Real-ESRGAN con un conjunto de datos de prueba, concretamente el dataset DIV2K [19] . Este primer experimento fue una conexión inicial con el modelo para verificar que todo funcionase correctamente. Utilizamos nuestro ordenador, equipado con una tarjeta gráfica RTX 3060, para llevar a cabo el entrenamiento.

5.1 Dataset DIV2K

Utilizamos el dataset DIV2K, reconocido ampliamente en el campo de la super-resolución debido a que contiene una gran variedad de imágenes de todo tipo de materiales, formas y colores como podemos observar en la figura 4.

Fig. 4: Muestra de los diferentes tipos de imágenes que puede haber en el dataset DIV2K

Este dataset consta de 800 imágenes de entrenamiento y 100 de validación. La mayoría de las imágenes de entrenamiento tienen una resolución de 1020x924, mientras que las imágenes ground truth tienen una resolución de 2040x1356, lo que facilita el upscale a 2 veces la resolución.

Inicialmente, teníamos la intención de generar un modelo que pudiese hacer un upscaling de 4 veces mayor la imagen original. Sin embargo, como era una prueba inicial y las imágenes ya venían por defecto con un downscale bicúbico a la mitad de la resolución (x2), decidimos mantener esta configuración.

En cuanto a la preparación de datos, solo tuvimos que generar el fichero de metadatos. Este fichero lista las rutas a las imágenes ground truth y a las imágenes de baja resolución del dataset DIV2K. Es esencial para organizar y acceder eficientemente a los datos durante el entrenamiento del modelo.

5.2 Métricas de Evaluación

- **MSE (Mean Squared Error)**
  Esta métrica calcula el promedio de las diferencias al cuadrado entre los valores de píxeles de la imagen original y la imagen generada. Un valor bajo de MSE indica una menor diferencia entre las dos imágenes, lo que sugiere una mejor calidad de la imagen generada. Por ejemplo, un MSE de 0 entre dos imágenes significaría que son idénticas.
  La fórmula del MSE es:

$$\text{MSE} = \frac{1}{NM} \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} (I(i,j) - K(i,j))^2 \quad (1)$$

  Donde I es la imagen original, K es la imagen generada, y N y M son las dimensiones de las imágenes.

- **PSNR (Peak Signal-to-Noise Ratio)**
  El PSNR mide la relación entre la potencia de la señal de la imagen original y la potencia del ruido que afecta a la calidad de la imagen. Se expresa en decibelios (dB). Cuanto mayor sea el valor de PSNR, menor será la cantidad de ruido presente en la imagen generada en comparación con la original. En caso de comparar dos imágenes idénticas, el PSNR sería infinito debido a una división por 0.
  La fórmula del PSNR es:

$$\text{PSNR} = 20 \cdot \log_{10} \left(\frac{\text{MAX}_I}{\sqrt{\text{MSE}}} \right) \quad (2)$$

  Donde MAX_I es el valor máximo posible de un píxel en la imagen.

- **SSIM (Structural Similarity Index Measure)**
  Esta métrica evalúa la similitud estructural entre la imagen original y la imagen generada. Se basa en la percepción visual humana y considera la luminancia, el contraste y la estructura de la imagen. Un valor de SSIM cercano a 1 indica una alta similitud estructural entre las dos imágenes.
  La fórmula del SSIM es:

$$\text{SSIM}(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (3)$$

  Donde μ_x y μ_y son las medias de x y y, σ_x^2 y σ_y^2 son las varianzas de x y y, σ_{xy} es la covarianza de x y y, y C_1 y C_2 son constantes para estabilizar la división.

5.3 Resultado de las Primeras Pruebas

A continuación, veremos los resultados en la tabla 1 , seguidamente de las reflexiones que obtuvimos de la primera prueba de implementación del modelo.

TABLA 1: RESULTADOS DE LAS MÉTRICAS DEL MODELO CORTO DE REAL-ESRGAN

Modelo Corto de Real-ESRGAN	
Iteraciones	70000
Épocas	174
Tiempo(h)	7
PSNR(Average)	24.51
SSIM(Average)	0.73
MSE(Average)	278.41

- **PSNR (Average): 24.51**
  Indica una calidad razonablemente buena en las imágenes generadas, aunque aún no alcanza niveles subóptimos (por encima de 30 dB) para la super-resolución, sugiriendo potencial de mejora con entrenamientos adicionales.

- **SSIM (Average): 0.73**
  El modelo muestra buena capacidad para preservar estructuras y detalles de las imágenes originales. Mejorar este valor podría aumentar la similitud perceptual entre las imágenes generadas y las originales, mejorando la calidad visual

- **MSE (Average): 278.41**
  Indica diferencias notables entre las imágenes generadas y las originales. Un MSE más bajo sería preferible, señalando errores menores en la reconstrucción de las imágenes.

En conclusión, los resultados obtenidos son aceptables, considerando el entrenamiento relativamente corto en comparación con las recomendaciones estándar. Aunque hay margen para mejorar, estos resultados indican que el modelo tiene potencial y puede beneficiarse de entrenamientos más prolongados para alcanzar un rendimiento superior.

5.4 Estudio de casos

En esta sección, analizaremos casos específicos para evaluar el rendimiento del modelo corto de Real-ESRGAN, centrándonos en los resultados extremos de las métricas PSNR, SSIM y MSE. Estos análisis revelarán las fortalezas y debilidades del modelo en diferentes escenarios, ofreciendo una comprensión más visual y completa de sus resultados.

5.4.1 Mejores Casos

La figura 5 presenta los siguientes valores métricos destacados: PSNR de 34.38, SSIM de 0.95 y MSE de 23.66.

El PSNR de sugiere una relación señal-ruido bastante alta, superando las 30 dB pero no llegando a los 40 dB que se consideran óptimos para la mejora de imágenes. El SSIM indica una alta similitud estructural entre ambas imágenes, mostrando una buena conservación de detalles y estructuras. Por último, el MSE señala que existen muy pocas diferencias puntuales entre las dos imagenes.

Fig. 5: Imagen que tiene los mejores valores de las métricas

Sin embargo, a pesar de estos buenos resultados métricos, la presencia predominante de un fondo negro en esta imagen limita su idoneidad como representación visual del modelo.

Por lo tanto, nos fijaremos mas al detalle en la segunda imagen con los mejores valores de PSNR y MSE para obtener una representación más adecuada y completa del rendimiento del modelo.

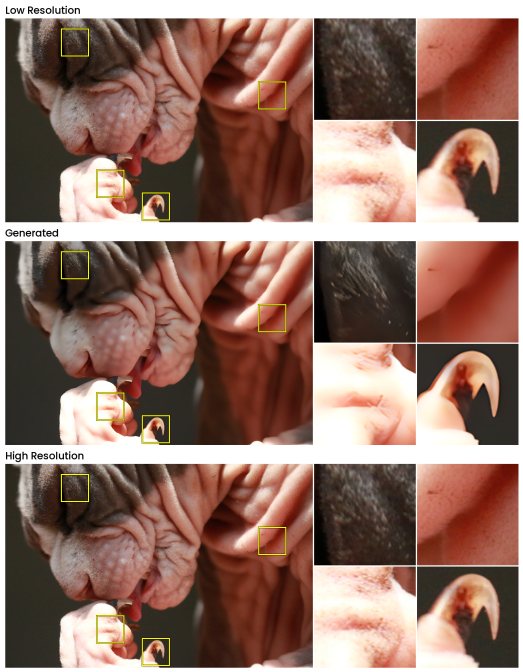

Fig. 6: Imagen que tiene los segundos mejores valores de las métricas

La figura 6 presenta los siguientes valores métricos destacados: PSNR de 32.31, SSIM de 0.85 y MSE de 38.16.

El PSNR indica una relación señal-ruido aceptable, que está dentro de un rango estándar para la mejora de imágenes pero sigue sin alcanzar los 40 dB. El SSIM señala la gran similitud estructural entre ambas imágenes, y por último, el MSE señala que todavía existen puntos donde no acaban de igualarse, pese a eso sigue indicando un buen resultado.

Esta imagen si que puede llegar a ser representativa ya que muestra como se diferencian varias texturas como podrian ser el pelo, las uñas y la piel . También muetra diferentes colores, sombras y tonos, que se llegan a apreciar claramente.

5.4.2 Peor Caso

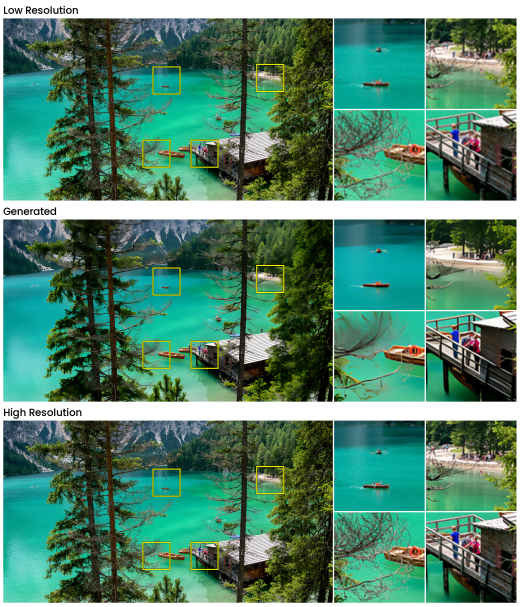

Fig. 7: Imagen que tiene los peores valores de las métricas

La figura 7 presenta los siguientes valores métricos: PSNR de 19.36, SSIM de 0.48 y MSE de 753.85.

Los valores métricos de PSNR, SSIM y MSE indican una calidad deficiente en la imagen procesada en comparación con la original. El PSNR muestra una alta cantidad de ruido y pérdida de calidad, mientras que el SSIM revela una baja similitud estructural y falta de preservación de detalles importantes, por último El MSE señala una significativa pérdida de información con errores cuadráticos medios elevados.

En resumen, estos resultados reflejan una baja calidad y similitud entre la imagen generada y la original, destacando áreas críticas donde el modelo necesita mejoras para obtener resultados más precisos y fieles.

Al analizar visualmente la imagen generada en la figura 7 , se observan desafíos significativos en el modelo. Las ramas muestran un notable nivel de difuminación, dificultando la recreación precisa. La identificación de personas en el muelle también resulta complicada, y se observa un cambio abrupto de colores en la superficie del lago, en lugar de una transición suave o difuminada esperada.

5.5 Conclusiones del Experimento Inicial

El modelo corto de Real-ESRGAN muestra resultados prometedores, destacando en algunos casos por su buena similitud estructural y alta relación señal-ruido, lo cual beneficia la conservación de detalles y la fidelidad visual. Sin embargo, se observa la limitación de representación visual debido a la presencia de un fondo negro en una de las imágenes.

Por otro lado, el peor caso exhibe valores muy bajos de PSNR y SSIM, indicando una pérdida significativa de calidad y detalles en comparación con la imagen original. Esto subraya áreas críticas de mejora como la preservación de detalles en elementos específicos como ramas, objetos pequeños y transiciones suaves de colores en el paisaje.

En resumen, aunque los resultados actuales son aceptables y muestran el potencial del modelo Real-ESRGAN, se destaca la necesidad de un entrenamiento más extenso para mejorar tanto la precisión como la fidelidad en la reproducción de imágenes. Este hallazgo sugiere que un entrenamiento prolongado podría elevar el rendimiento del modelo, mejorando la calidad general de las imágenes procesadas y reduciendo las discrepancias respecto a las originales.

6 Experimento Final

Luego de confirmar la correcta implementación y validación del modelo, procedimos a utilizar el dataset de imágenes satelitales AID [20]. En este experimento, optamos por ejecutarlo en un servidor GPU proporcionado por el área de computación de la UAB, el cual estaba equipado con una tarjeta gráfica RTX 3090, considerablemente más potente que la utilizada previamente. Esta mejora en hardware permitió un entrenamiento mucho más rápido y eficiente.

Nuestro objetivo principal consiste en desarrollar un modelo utilizando las imágenes de entrenamiento del dataset AID y evaluar su rendimiento al aplicarlo para mejorar la resolución de las imágenes de validación.

Para este experimento, aumentamos significativamente las iteraciones de entrenamiento, pasando de 70,000 a 900,000, con el fin de investigar cómo esta variación afecta el rendimiento del modelo.

6.1 Dataset AID

El dataset AID está compuesto por un total de 10,000 imágenes satelitales, cada una con una resolución de 600x600 píxeles. Estas imágenes han sido seleccionadas para representar una amplia variedad de escenas geográficas y condiciones ambientales, proporcionando un conjunto de datos diverso y adecuado para aplicaciones de análisis y procesamiento de imágenes satelitales.

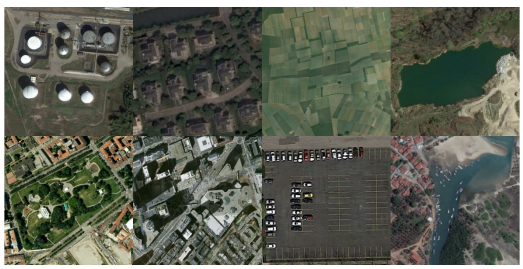

Fig. 8: Muestra de los diferentes tipos de imágenes que puede haber en el dataset AID

En la figura 8 podemos observar una muestra del dataset AID, viendo la variedad diferentes terrenos que contiene.

Esta vez creamos nuestro propio dataset utilizando imágenes del conjunto AID, el cual está organizado en 30 carpetas. Para el conjunto de entrenamiento, seleccionamos las primeras 100 imágenes de cada carpeta, resultando en un total de 3000 imágenes. Para el conjunto de validación, escogimos las siguientes 20 imágenes de cada carpeta, sumando un total de 600 imágenes.

Posteriormente, aplicamos un script para reducir la resolución de cada imagen en tres escalas diferentes: 0.75, 0.5 y 1/3 del tamaño original. Estas versiones reducidas de las imágenes, junto con las imágenes originales, son las que utilizamos finalmente para el entrenamiento, alcanzando así un total de 12000 imágenes para el conjunto de entrenamiento.

Para el conjunto de validación, dado que deseamos realizar un upscaling de 2 veces, generamos versiones de baja resolución que son la mitad del tamaño original, es decir, un total de 600 imágenes de 300x300 píxeles.

En resumen, el tamaño final del dataset es de 13200 imágenes en total, con 12000 imágenes utilizadas para el entrenamiento y 1200 imágenes para la validación.

6.2 Resultados del Experimento Final

TABLA 2: RESULTADOS DE LAS MÉTRICAS DEL MODELO LARGO DE REAL-ESRGAN

Modelo Largo de Real-ESRGAN	
Iteraciones	900000
Épocas	239
Tiempo(h)	80
PSNR(Average)	27.62
SSIM(Average)	0.80
MSE(Average)	143.81

- **PSNR (Average): 27.62**
  Indica una mejora significativa en la calidad de las imágenes generadas, alcanzando un nivel más cercano a los estándares subóptimos para super-resolución (por encima de 30 dB).

- **SSIM (Average): 0.80**
  Refleja una mejor capacidad del modelo para preservar detalles y estructuras, resultando en una similitud perceptual más alta entre las imágenes generadas y las originales.

- **MSE (Average): 143.81**
  Indica una reducción considerable en las diferencias entre las imágenes generadas y las originales, señalando una reconstrucción más precisa y menos errores.

Comparando con el modelo corto, el modelo largo de Real-ESRGAN muestra una mejora sustancial en todas las métricas evaluadas. Esto sugiere que el entrenamiento prolongado ha sido efectivo para mejorar el rendimiento del modelo, elevando la calidad general de las imágenes procesadas y reduciendo las discrepancias respecto a las originales.

6.3 Mejores Casos

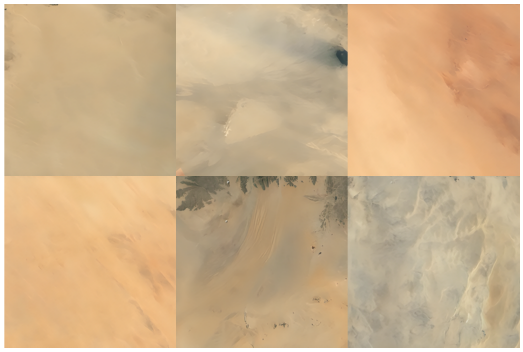

Fig. 9: Las seis imagenes con mejores valores de las métricas

Fig. 10: Imagen representativa del modelo con los mejores valores de las métricas

En la figura 9 se observan imágenes del desierto con mínimas estructuras definidas.

Todos los casos muestran valores de PSNR superiores a 35, alcanzando hasta 39.6 en el mejor caso, indicando un rendimiento cercano a lo óptimo (40 dB) en términos de PSNR. Los valores de SSIM no bajan de 0.90, con un máximo de 0.95, lo que sugiere una alta similitud entre las imágenes generadas y las originales. Además, los valores de MSE son mínimos, sin superar los 15 y llegando a tan solo 7 en un caso, lo que implica diferencias insignificantes entre las imágenes.

Sin embargo, estos resultados no sonrepresentativos de la calidad general del modelo, ya que las imágenes analizadas no constan de diferentes texturas, formas y colores.

Por lo tanto, procederemos a analizar la siguiente imagen con los mejores valores de las métricas.

La figura 10 presenta los siguientes valores métricos destacados: PSNR de 34.30, SSIM de 0.92 y MSE de 24.11.

El PSNR aunque sigue sin llegar al valor óptimo nos indica que la relación señal-ruido es bastante alta, lo cual positivo para el modelo. El valor de SSIM representa una similitud muy alta, dando a entender que los detalles y estructuras se han conservado, y por último el MSE con un valor bastante bajo, muestra que hay pocos puntos donde la imàgenes difieran significativamente.

De cara al análisis visual, observamos en la figura 10 que las formas de los edificios, incluyendo la del campo que está justo en el centro, se han conservado de manera precisa sin la aparición de artefactos. Un punto a enfocar es que nuestro modelo tiende a suavizar los contornos como se puede observar en la figura 10.

6.4 Peor Caso

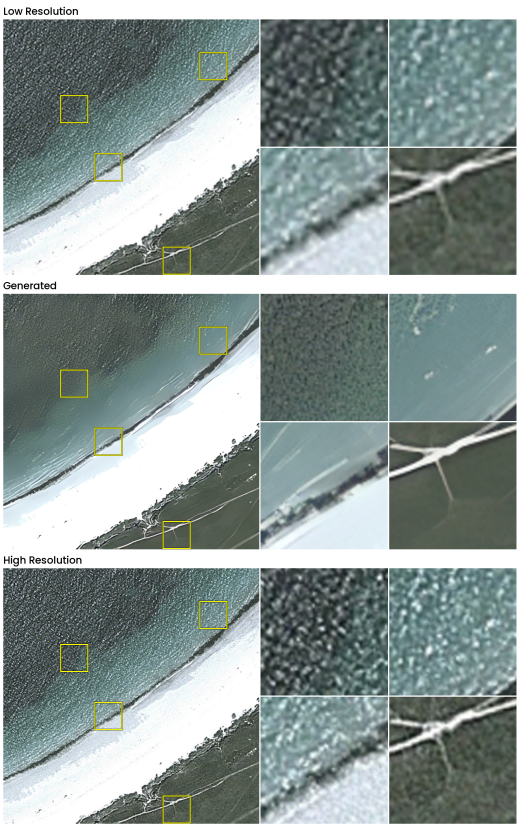

Fig. 11: Imagen representativa del modelo con las peores métricas

La figura 11 presenta los siguientes valores métricos: PSNR de 20.91 dB, SSIM de 0.52 y MSE de 527.09.

El PSNR se sitúa por debajo de los umbrales aceptables para la mejora de imágenes, indicando una considerable degradación de la señal y pérdida de calidad perceptible. Además, el SSIM refleja una baja similitud estructural, evidenciando una notable falta de conservación de detalles y estructuras esenciales. Por otro lado, el MSE revela una discrepancia significativa y un elevado promedio de errores cuadráticos medios, señalando una pérdida sustancial de información en la imagen procesada.

En resumen, estos valores métricos reflejan una calidad deficiente y similitud limitada entre la imagen generada y la original, resaltando áreas críticas que requieren mejoras para obtener resultados más precisos y fieles.

Al examinar visualmente la imagen de la playa en la figura 11, se observan desafíos notables en la reproducción del mar. Las formas de las olas han sido representadas como líneas blancas, lo cual indica dificultades en la generación de estructuras marítimas detalladas.

6.5 Conclusiones Experimento Final

El experimento final del modelo Real-ESRGAN, entrenado durante 900,000 iteraciones utilizando el dataset AID, mostró mejoras significativas en comparación con el modelo anterior. Se observaron notables mejoras en la similitud estructural, la conservación de detalles y la discrepancia con respecto a las imágenes originales.

Estos resultados destacaron la eficacia del entrenamiento prolongado para mejorar la calidad y precisión del modelo en la tarea de super-resolución de imágenes satelitales. Aunque los avances fueron prometedores, hubo oportunidades para optimizar aún más el rendimiento del modelo mediante estrategias adicionales de refinamiento y ajuste de parámetros.

7 Conclusiones del Proyecto

Este proyecto implementó y evaluó el modelo Real-ESRGAN para la super-resolución de imágenes utilizando los datasets DIV2K y AID. Los experimentos demostraron mejoras significativas en la calidad de las imágenes al incrementar las iteraciones de entrenamiento y el número de imágenes procesadas.

Inicialmente, se obtuvieron resultados aceptables con DIV2K, aunque se identificaron áreas para mejorar. Posteriormente, utilizando el dataset AID y entrenando durante 900,000 iteraciones, se lograron mejoras sustanciales en métricas como PSNR, SSIM y MSE.

Se concluyó que el aumento de iteraciones y el incremento en el número de imágenes procesadas llevaron a resultados positivos mejorados. Los análisis de casos específicos resaltaron la necesidad de enfoques diferenciados según el tipo de imagen.

Para futuros proyectos, se planea ajustar hiperparámetros para optimizar el rendimiento, ampliar el dataset para mejorar la generalización del modelo, explorar entrenamientos con imágenes de mayor resolución y utilizar recursos computacionales mas avanzados para acelerar el proceso.

En resumen, el estudio destacó el potencial del modelo Real-ESRGAN para generar imágenes de alta calidad, subrayando la importancia de la optimización continua y la exploración de nuevas técnicas en investigaciones futuras.

AGRADECIMIENTOS

Deseo expresar mi sincero agradecimiento al profesor Robert Benavente Vidal por su guía y apoyo durante el desarrollo de este proyecto. Su disposición y orientación constante fueron esenciales para alcanzar los resultados obtenidos.

También agradezco al equipo de desarrolladores del modelo Real-ESRGAN, Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao y Chen Change Loy, por su excepcional trabajo en la estructuración y documentación de este modelo, fundamental para la implementación exitosa del proyecto.

REFERENCIAS

[1] Research, E. (2023, 18 abril). The value of super resolution — real world use case. Medium. https://medium.com/sentinel-hub/the-value-of-super-resolution-real-world-use-case-2ba811f4cd7f

[2] Geoawesoemenss. (2024, 27 marzo). Enhancing Satellite Imagery Readability with Super-resolution Machine Learning Models - Geoawesomeness. Geoawesomeness. https://geoawesomeness.com/eo-hub/enhancing-satellite-imagery-readability-with-super-resolution-machine-learning-models/

[3] Ci, O. S. (2023, 21 abril). Image Super Resolution: A Comparison between Interpolation & Deep Learning-based Techniques to Improve Clarity of Low-Resolution Images. Medium. https://medium.com/htx-s-s-coe/image-super-resolution-a-comparison-between-interpolation-deep-learning-based-techniques-to-25e7531ab207

[4] What Is a Convolutional Neural Network? — 3 things you need to know. (s. f.). MATLAB & Simulink. https://www.mathworks.com/discovery/convolutional-neural-network.html

[5] Dong, C., Loy, C. C., & Tang, X. (2016). Accelerating the Super-Resolution Convolutional Neural Network. arXiv (Cornell University). https://doi.org/10.48550/arxiv.1608.00367

[6] Kim, J., Lee, J. K., & Lee, K. M. (2015). Accurate image Super-Resolution using very deep convolutional networks. arXiv (Cornell University). https://doi.org/10.48550/arxiv.1511.04587

[7] Lim, B., Son, S., Kim, H., Nah, S., & Lee, K. M. (2017). Enhanced Deep Residual Networks for Single Image Super-Resolution. arXiv (Cornell University). https://doi.org/10.48550/arxiv.1707.02921

[8] Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial networks. arXiv (Cornell University). https://doi.org/10.48550/arxiv.1406.2661

[9] Ledig, C., Theis, L., Huszar, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., & Shi, W. (2016). Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. arXiv (Cornell University).

[10] Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Loy, C. C., Qiao, Y., & Tang, X. (2018). ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks. arXiv (Cornell University). https://doi.org/10.48550/arxiv.1809.00219

[11] Karras, T., Aila, T., Laine, S., & Lehtinen, J. (2017). Progressive Growing of GANs for Improved Quality, Stability, and Variation. arXiv (Cornell University). https://doi.org/10.48550/arxiv.1710.10196

[12] Merritt, R. (2022, 16 septiembre). What Is a Transformer Model? — NVIDIA Blogs. NVIDIA Blog. https://blogs.nvidia.com/blog/what-is-a-transformer-model/

[13] Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., & Timofte, R. (2021). SWiNIR: Image restoration using SWIN Transformer. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2108.10257

[14] Touvron, H., Cord, M., Douze, M., Massa, F., Sablayrolles, A., & Jégou, H. (2020). Training data-efficient image transformers & distillation

[15] Cochard, D. (2024, 11 enero). Real ESRGAN: Super-Resolution Model Enhanced for Denoising. Medium. https://medium.com/axinc-ai/real-esrgan-super-resolution-model-enhanced-for-denoising-dd581b2702a8

[16] Stewart, K. (2024, 24 mayo). Mean squared error (MSE) — Definition, Formula, Interpretation, & Facts. Encyclopedia Britannica. https://www.britannica.com/science/mean-squared-error

[17] Hu, M., Luo, X., Chen, J., Lee, Y. C., Zhou, Y., & Wu, D. (2021). Virtual reality: A survey of enabling technologies and its applications in IoT. Journal Of Network And Computer Applications, 178, 102970. https://doi.org/10.1016/j.jnca.2020.102970

[18] Larkin, K. G. (2015). Structural Similarity Index SSIMplified: Is there really a simpler concept at the heart of image quality measurement? arXiv (Cornell University). https://doi.org/10.48550/arxiv.1503.06680

[19] DIV2K Dataset. (s. f.). https://data.vision.ee.ethz.ch/cvl/DIV2K/

[20] AID: A Benchmark Dataset for Performance Evaluation of Aerial Scene Classification. (s. f.). https://captain-whu.github.io/AID/