



UNIVERSIDAD POLITÉCNICA DE MADRID

ESCUELA TÉCNICA SUPERIOR DE INGENIEROS EN TOPOGRAFÍA, GEODESIA Y
CARTOGRAFÍA

TITULACIÓN DE GRADO EN INGENIERÍA DE LAS TECNOLOGÍAS Y LA
INFORMACIÓN GEOESPECIAL



TRABAJO FIN DE GRADO

**Generación de modelos de superresolución de
imágenes satelitales y ortoimágenes a partir de la
arquitectura VDSR**

Madrid, junio de 2024

Alumno: Ignacio Galo Vallejo Collados

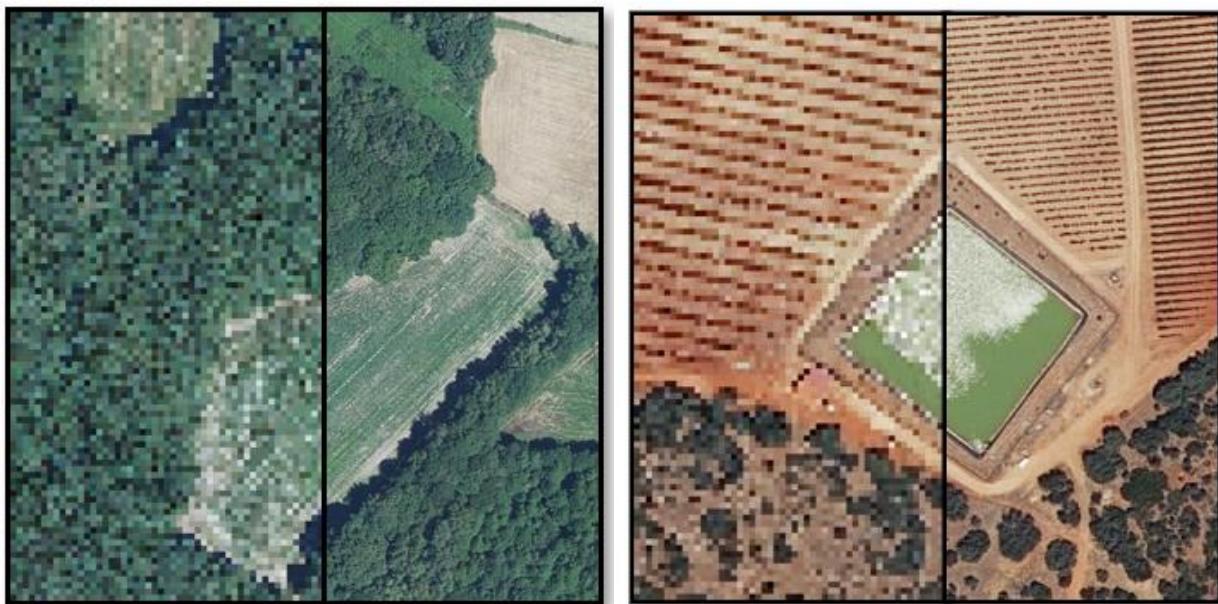
Tutores: Miguel Ángel Manso Callejo

Calimanut Ionut Cira

UNIVERSIDAD POLITÉCNICA DE MADRID
ESCUELA TÉCNICA SUPERIOR DE INGENIEROS EN TOPOGRAFÍA, GEODESIA Y
CARTOGRAFÍA
TITULACIÓN DE GRADO EN INGENIERÍA DE LAS TECNOLOGÍAS Y LA
INFORMACIÓN GEOESPECIAL

TRABAJO FIN DE GRADO

**Generación de modelos de superresolución de
ímagenes satelitales y ortoimágenes a partir de la
arquitectura VDSR**



Madrid, junio de 2024

Alumno: Ignacio Galo Vallejo Collados

Tutores: Miguel Ángel Manso Callejo

Calimanut Ionut Cira

RESUMEN

El crecimiento en los últimos años del volumen de imágenes procedentes de satélites y sensores aéreos ha provocado que, para ciertas aplicaciones, sea necesario el aumento de su resolución espacial, evitando la adquisición de nuevos datos. La superresolución, conocida como el proceso de reconstruir una imagen de alta resolución a partir de una de baja resolución, se ha convertido en una técnica clave para el procesamiento de imágenes.

Este trabajo de fin de grado presenta un estudio de generación de modelos de superresolución dirigido a imágenes satelitales y ortoimágenes a partir de la arquitectura VDSR (del inglés: "*very deep super-resolution*"). El proyecto ha abordado dos enfoques diferenciados. En primer lugar, se han realizado entrenamientos a partir de imágenes degradadas por interpolación y, en segundo lugar, se han utilizado pares de imágenes nativas de distinto origen (GEOSAT2 y PNOA), simulando condiciones reales.

Para la generación del dataset, se desarrollaron procesos de recorte, normalización estadística y emparejamiento espacial entre los pares de imágenes. Posteriormente, se adaptó el código base de la red VDSR y se entrenaron múltiples modelos bajo distintos hiperparámetros. La evaluación de los modelos se realizó mediante métricas cuantitativas (PSNR y SSIM) junto con valoraciones perceptuales de las imágenes resultantes.

Los resultados mostraron que, en condiciones donde se utiliza un buen conjunto de datos y se realizan entrenamientos supervisados, la arquitectura VDSR supera de forma significativa a los métodos clásicos de interpolación, tanto en fidelidad estructural como en detalle visual. En el escenario con datos de distinta procedencia, si bien no se presentaron mejoras en las métricas cuantitativas, se apreciaron ganancias perceptuales en la representación de color, facilitando el entendimiento de imágenes con cromatismos deficientes.

En conjunto, el proyecto valida el potencial de las redes profundas como herramienta para aumentar la utilidad de imágenes geoespaciales de baja resolución y posiciona a VDSR como una de las mejores arquitecturas en términos de coste computacional–eficiencia.

El código y los modelos resultantes se puede encontrar en el repositorio GitHub:

https://github.com/ignacioVallejo03/TFG_Ignacio_Vallejo

ABSTRACT

The growth in recent years of the volume of images from satellites and aerial sensors has led to the need, for certain applications, to increase their spatial resolution, avoiding the acquisition of new data. Super-resolution, known as the process of reconstructing a high-resolution image from a low-resolution one, has become a key technique for image processing.

This final degree project presents a study on the generation of super-resolution models aimed at satellite images and orthophotos based on the VDSR (very deep super-resolution) architecture. The project has addressed two different approaches. First, training was carried out using images degraded by interpolation, and second, pairs of native images from different sources (GEOSAT2 and PNOA) were used, simulating real conditions.

For the generation of the dataset, cropping processes, statistical normalization, and spatial matching between image pairs were developed. Subsequently, the base code of the VDSR network was adapted and multiple models were trained under different hyperparameters. The evaluation of the models was carried out using quantitative metrics (PSNR and SSIM) along with perceptual assessments of the resulting images.

The results showed that, in conditions where a good dataset is used and supervised training is performed, the VDSR architecture significantly outperforms classical interpolation methods, both in structural fidelity and visual detail. In the scenario with data from different sources, although no improvements were observed in the quantitative metrics, perceptual gains were noted in color representation, facilitating the understanding of images with deficient chromatic content.

Overall, the project validates the potential of deep networks as a tool to increase the usefulness of low-resolution geospatial images and positions VDSR as one of the best architectures in terms of computational cost-efficiency.

The resulting code and models can be found on the GitHub repository:

https://github.com/ignacioVallejo03/TFG_Ignacio_Vallejo

ÍNDICE DE CONTENIDOS

RESUMEN.....	3
ABSTRACT	4
1. INTRODUCCIÓN.....	13
1.1. Objetivos	13
1.1.a Pares de imágenes con misma fuente	14
1.1.b Pares de imágenes de distinta fuente	14
2. MARCO TEÓRICO	15
2.1. La Imagen	15
2.1.1. Métricas de evaluación	15
2.2. Aprendizaje profundo	17
2.2.1. Redes neuronales artificiales	18
2.2.2. Redes neuronales convolucionales.....	19
2.3. La superresolución	21
2.3.1. Métodos tradicionales de superresolución (Nasrollahi & Moeslund, 2014).....	21
2.3.2. Superresolución basada en redes neuronales	21
3. ESTADO DEL ARTE	23
3.1. arquitecturas relevantes	23
3.1.1. SRCNN.....	23
3.1.2. VDSR.....	23
3.1.3. SRResNet.....	24
3.1.4. EDSR.....	24
3.1.5. RDN.....	24

3.1.6. RCAN	25
3.1.7. Arquitecturas actuales.....	25
3.2. Arquitectura seleccionada	26
3.2.1. Estructura de VDSR (J. Kim et al., 2016b)	27
3.3. Trabajos previos	29
4. MATERIALES	30
4.1. Hardware.....	30
4.2. Software.....	30
5. METODOLOGÍA.....	32
5.1. Generación del dataset	33
5.1.1. división y preparación del dataset	37
5.2. Generación del modelo	39
5.2.1 Modificaciones realizadas	39
5.3. Entrenamiento.....	41
5.4. Inferencia y evaluación de resultados.....	42
6. RESULTADOS	44
6.1. Resultados del Objetivo_1.....	44
6.1.1. Evaluación de modelos para imágenes del GEOSAT2.....	44
6.1.2 Evaluación de modelos para ortoimágenes del PNOA.....	61
6.2. Resultados del objetivo_2	77
6.2.1 GEOSAT2 redimensionado a PNOA.....	77
6.2.1 PNOA redimensionado a GEOSAT2	79
7. CONCLUSIONES	82
8. PRESUPUESTO.....	84
8.1 Costes del hardware	84
8.2 Costes de software.....	84

8.3 Costes de adquisición de los datos	85
8.4 Coste de personal o de trabajo.....	86
9. PLANIFICACIÓN	87
9.1. Revisión del estado del arte	88
9.2. Preparación y generación del dataset	88
9.3. Adaptación e implementación del modelo.....	88
9.4. Entrenamiento, validación y ajuste de parámetros	89
9.5. Elaboración de la memoria	89
REFERENCIAS	90

ACRONIMOS Y ABREVIATURAS

PNOA- Plan nacional de ortofotografía aérea

GEOSAT2- Geo-Stationary Satellite (Satélite geoestacionario)

VDSR – Very Deep super-resolution

IA – Inteligencia artificial

RNA- Redes neuronales artificiales

CNN- Convolutional neuronal network (Red neuronal convolucional)

MSE- Medium square error (Error medio cuadrático)

SGD- Stochastic gradient descent

PNSR- Peak signal-to-noise ratio

SSIM-Structural Similarity Index

SRCNN-Super-Resolution Convolutional Neural Network

SRResNet -Super-Resolution Residual Network

EDSR - Enhanced Deep Super-Resolution Network

DRRN- Deep Recursive Residual Network

RDN - Residual Dense Network

RCAN - Residual Channel Attention Network

LR-low resolution (Baja resolución)

HR-high resolution (Alta resolución)

ReLU- Rectified lineal unit

CPU-Central Processing Unit

RAM- Random Access Memory

GPU- Graphics Processing Unit

BBOX- bounding box

GAN- Generative adversarial network

LISTA DE TABLAS

Tabla 1. Comparación de PSNR y SSIM para el Set14 y escalado x2. Modificado de (X. Wang et al., 2022)	26
Tabla 2. Numero de parches generados para las fuentes del GEOSAT2 y PNOA	37
Tabla 3. Hiperparámetros constantes en los modelos finales.....	44
Tabla 4. Métricas de los modelos para imágenes de GEOSAT2 con factor de escalado x2	45
Tabla 5. Métricas de los modelos para imágenes de GEOSAT2 con factor de escalado x3	47
Tabla 6. Métricas de los modelos para imágenes de GEOSAT2 con factor de escalado x4	50
Tabla 7. Métricas de los modelos para imágenes de GEOSAT2 con factor de escalado x2, x3 y x4	52
Tabla 8. Comparativa de métricas entre el modelo 3 e interpolaciones clásicas con factor de escalado x2	55
Tabla 9. Comparativa de métricas entre el modelo 6 e interpolaciones clásicas con factor de escalado x3.	57
Tabla 10. Comparativa de métricas entre el modelo 9 e interpolaciones clásicas con factor de escalado x4	59
Tabla 11. Métricas de los modelos para imágenes del PNOA con factor de escalado x2....	61
Tabla 12. Métricas de los modelos para imágenes del PNOA con factor de escalado x3 ...	64
Tabla 13. Métricas de los modelos para imágenes del PNOA con factor de escalado x4 ...	67
Tabla 14. Comparativa de métricas entre el modelo 14 e interpolaciones clásicas con factor de escalado x2	70
Tabla 15. Comparativa de métricas entre el modelo 17 e interpolaciones clásicas con factor de escalado x3	72
Tabla 16. Comparativa de métricas entre el modelo 21 e interpolaciones clásicas con factor de escalado x4	74

Tabla 17. Métricas de los modelos entrenados con imágenes de distinta fuente (GEOSAT2 redimensionado a PNOA).....	77
Tabla 18. Métricas de los modelos entrenados con imágenes de distinta fuente (PNOA redimensionado a GEOSAT2)	79
Tabla 19. Costes de proyecto en Hardware.....	84
Tabla 20. Costes del proyecto en Software	85
Tabla 21. Costes de adquisición de datos	86
Tabla 22. Costes de personal del proyecto.....	86
Tabla 23. Presupuesto bruto Total.	87
Tabla 24. Presupuesto bruto con contrato.	87

LISTA DE FIGURAS

Ilustración 1. Ejemplo de pares de imágenes del GEOSAT2 y PNOA.....	14
Ilustración 2. Par de imágenes conforme al segundo objetivo.....	14
Ilustración 3. Aprendizaje profundo, automático e inteligencia artificial.	17
Ilustración 4. Estructura básica de una neurona artificial. Modificado de (Incio-Flores et al., 2023)	18
Ilustración 5. Estructura base de una CNN (LauAI, 2024).....	19
Ilustración 8. Arquitectura VDSR. Tomado de (J. Kim et al., 2016).....	28
Ilustración 9. Esquema de la Metodología aplicada en el proyecto.	33
Ilustración 10. Ejemplo de coberturas para el año 2023.....	34
Ilustración 11. Imágenes de alta y baja resolución del GEOSAT2 con escalado x2	46
Ilustración 12. Imágenes reconstruidas por los modelos 1,2 y 3	46
Ilustración 13. Imágenes de alta y baja resolución del GEOSAT2 con escalado x3	48
Ilustración 14. Imágenes reconstruidas por los modelos 4,5 y 6	49
Ilustración 15. Imágenes de alta y baja resolución del GEOSAT2 con escalado x4.....	51
Ilustración 16. Imágenes reconstruidas por los modelos 7,8 y 9.	51
Ilustración 17. Imágenes del modelo multiescala	54
Ilustración 18. Imágenes comparativas de modelos vs métodos clásicos, del GEOSAT2 y factor de escalado x2.....	56
Ilustración 19. Imágenes comparativas de modelos vs métodos clásicos, del GEOSAT2 y factor de escalado x3.....	58
Ilustración 20. Imágenes comparativas de modelos vs métodos clásicos, del GEOSAT2 y factor de escalado x4.....	60
Ilustración 21. Imágenes de alta y baja resolución del PNOA con escalado x2.....	62

Ilustración 22. Imágenes reconstruidas por los modelos 11,12,13 y 14.....	63
Ilustración 23. Imágenes de alta y baja resolución del PNOA con escalado x3.....	65
Ilustración 24. Imágenes reconstruidas por los modelos 15,16 y 17.	66
Ilustración 25. Imágenes de alta y baja resolución del PNOA con escalado x2.....	68
Ilustración 26. Imágenes reconstruidas por los modelos 18,19,20,21.	69
Ilustración 27. Imágenes comparativas de modelos vs métodos clásicos, del PNOA con factor de escalado x2.....	71
Ilustración 28. Imágenes comparativas de modelos vs métodos clásicos, del PNOA con factor de escalado x3.....	73
Ilustración 29. Imágenes comparativas de modelos vs métodos clásicos, del PNOA con factor de escalado x4.....	75
Ilustración 30. Imágenes comparativas entre los modelos 1, 2 y 3 de entrenamiento con fuentes dispares.	78
Ilustración 31. Imágenes comparativas entre los modelos 4, 5 y 6 de entrenamiento con fuentes dispares.	80
Ilustración 32. Evaluación perceptual del modelo número 6.....	81

1. INTRODUCCIÓN

En la actualidad, el uso de las imágenes satelitales y ortoimágenes de alta calidad ha resultado ser un recurso de gran valor para un amplio número de disciplinas como la teledetección, la gestión de territorios, la agricultura, la planificación urbana (Wellmann et al., 2020) e incluso la respuesta a desastres naturales (Kaku, 2019) de una forma más eficiente.

Sin embargo, en muchos casos, estas imágenes no presentan la resolución espacial idónea o necesaria para su utilización. Esta problemática es debida a restricciones económicas, tecnológicas o físicas de los instrumentos utilizados para la adquisición, lo que en ocasiones provoca un uso limitado de las imágenes, reduciendo significativamente el beneficio que podemos obtener de ellas.

En este contexto es donde las técnicas de superresolución pueden marcar la diferencia. La superresolución de imágenes es una técnica avanzada utilizada en el ámbito del procesamiento digital de imágenes que permite mejorar la resolución de imágenes, generando salidas de mayor calidad a partir de imágenes de menor resolución. Esta técnica se basa en el uso de algoritmos de aprendizaje profundo que aprenden a reconstruir detalles finos y texturas que pueden perderse en imágenes de baja resolución («Superresolución de Imágenes», s. f.). Así, aplicando estas técnicas, conseguiremos mejorar la precisión de los posibles análisis futuros sin necesidad de adquirir nuevas imágenes de mayor resolución que, en la mayoría de los casos, podrían tener un coste elevado o simplemente no estar disponibles.

En este trabajo de fin de grado se detallarán los procedimientos llevados a cabo y los resultados obtenidos de la aplicación de técnicas de superresolución sobre imágenes satelitales y ortoimágenes. Estas técnicas se llevarán a cabo a través del uso de redes convolucionales (CNNs), enfocándose principalmente en la arquitectura *Very Deep Super-Resolution* (VDSR), probada ser una de las mejores arquitecturas para la realización de tareas de superresolución y que se presentará con detalle más adelante.

1.1. Objetivos

Este proyecto ha sido realizado dentro de la Cátedra IGN-CNIG-UPM sobre tecnologías aplicadas a la información geoespacial y en el que se han propuesto dos objetivos principales. El primero de ellos se enfoca en la obtención o generación de modelos de superresolución que se entrenen a partir de pares de imágenes de la misma fuente. Y en segundo lugar se ha planteado el objetivo de generar modelos que se entrenen a partir de imágenes de distinta fuente.

1.1.a Pares de imágenes con misma fuente

En este primer objetivo del proyecto, los pares de imágenes de alta y baja resolución con los que se entrenará la red neuronal se generarán a partir de imágenes satelitales del GEOSAT2 y ortoimágenes del PNOA de forma individual, es decir, cada fuente de datos tendrá sus propias imágenes de alta y baja resolución (HR, LR respectivamente).

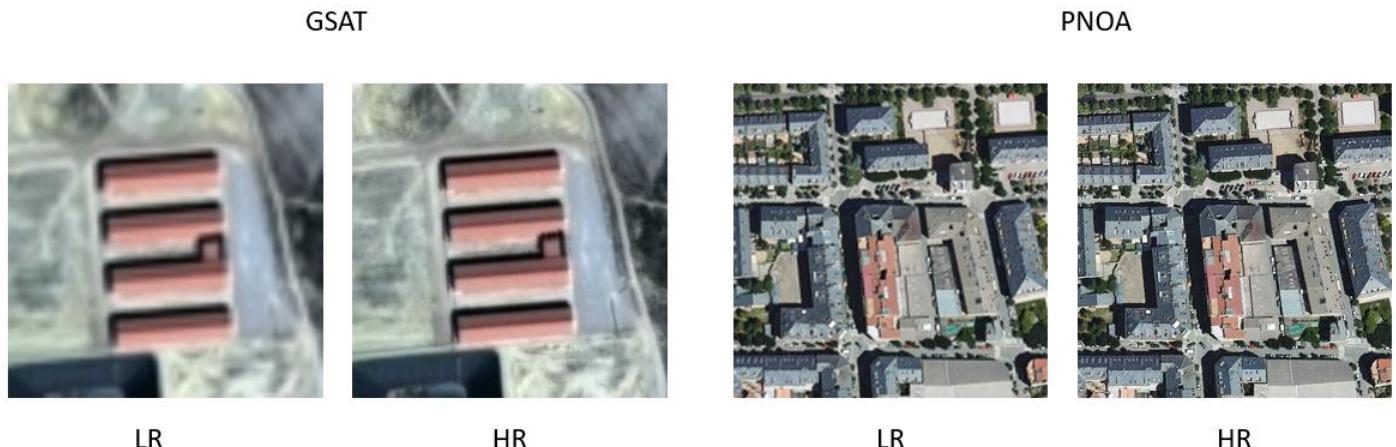


Ilustración 1. Ejemplo de pares de imágenes del GEOSAT2 y PNOA.

1.1.b Pares de imágenes de distinta fuente

En este segundo objetivo, los pares de imágenes se generarán a partir de las imágenes satelitales y ortoimágenes conjuntamente, es decir, las imágenes de alta resolución serán las ortoimágenes provenientes del PNOA, mientras que sus pares de baja resolución serán las imágenes satelitales obtenidas del GEOSAT2, consiguiendo así pares de imágenes de resoluciones nativas.

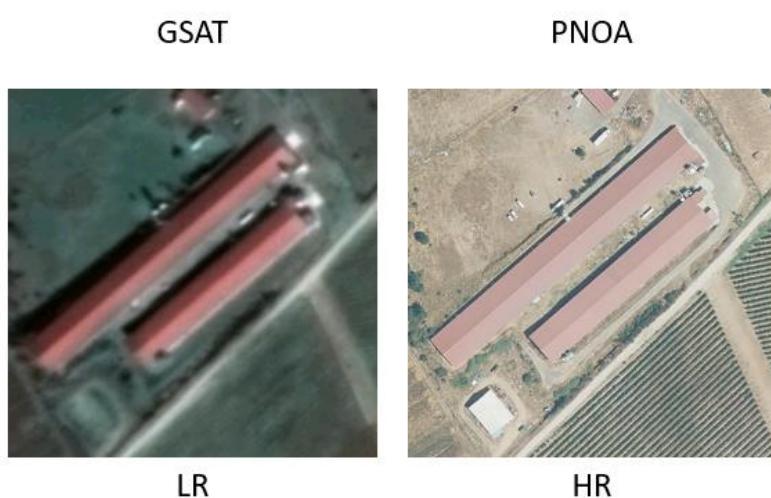


Ilustración 2. Par de imágenes conforme al segundo objetivo.

2. MARCO TEÓRICO

2.1. La Imagen

Una imagen digital es una representación bidimensional de una escena que se almacena en una matriz de píxeles. Cada uno de esos píxeles contiene valores que definen la intensidad de la luz en las imágenes representadas en escala de grises o los componentes de color en las imágenes multiespectrales. (Balter, 1993)

Una imagen se compone de distintas propiedades en las que encontramos la **resolución espacial**, que indica la cantidad de detalle visual a través de la densidad de píxeles; la **resolución radiométrica**, que representa la cantidad de niveles de intensidad que un píxel es capaz de representar; el **contraste**, que se define como la diferencia entre una región más clara y otra más oscura de la propia imagen. La **nitidez** que se asocia con el grado de claridad de los detalles de la imagen y la **relación señal-ruido** que mide la información útil y el ruido presente en la imagen (un valor alto de SNR indica una imagen de mayor calidad). (Rubel et al., 2022)

Estas características no solo indican la calidad con la que percibimos, sino que también influyen de manera directa al aplicar diferentes procesamientos sobre ella, como podrían ser los que se producen en la realización de un proceso de superresolución, ya que la capacidad que un modelo de superresolución tendrá para mejorar la calidad sin la introducción de ruido o artefactos dependerá en gran medida de cómo se interpreten estas cualidades de las imágenes.

2.1.1. Métricas de evaluación (*Image Quality Metrics - MATLAB & Simulink, s. f.*)

Cuando una imagen ha sido procesada, como en el caso de la superresolución, y queremos comparar el resultado con una imagen de referencia para evaluar los resultados del proceso, las métricas de uso más frecuentes son el **PSNR** (*Peak signal-to-noise ratio*) y el **SSIM** (*Structural Similarity Index*).

El PSNR mide la relación entre la señal original y el ruido introducido en el procesamiento. Esta métrica, se basa en el MSE (error cuadrático medio) entre la imagen procesada y la de referencia.

$$MSE = \frac{1}{M \times N \times O} \sum_{x=1}^M \sum_{y=1}^N \sum_{z=1}^O \left[(I_{x,y,z} - I'_{x,y,z})^2 \right] \quad (1)$$

Donde I es la imagen de referencia, I' es la imagen procesada, y M , N y O son las dimensiones de la imagen.

A partir del MSE, podemos calcular el PSNR a partir de la siguiente formula.

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

Donde \max es el valor máximo posible de un píxel.

Cuanto mayor sea el PSNR, menor es el error y, por tanto, mejor es la calidad de la imagen. Generalmente, los valores mayores a 40 dB indican una reconstrucción excelente, siendo difícil distinguir visualmente entre la imagen reconstruida y la original. Entre los 25 y 40 dB, encontramos resultados buenos o aceptables. Por debajo de los 25 dB encontramos resultados más pobres, aunque dependiendo del contexto pueden ser resultados más que útiles para diferentes contextos.

El SSIM evalúa la similitud estructural entre dos imágenes considerando factores como la luminancia, el contraste y la estructura. Se calcula mediante la siguiente fórmula:

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

El SSIM toma valores entre 0 y 1. Un valor de 1 indica una similitud estructural perfecta. Por lo general, un SSIM mayor a 0.9 se asocia con una imagen de muy buena calidad. Entre 0.7 y 0.9 encontramos una similitud moderada. Por último, valores de SSIM menores a 0.7 suelen representar reconstrucciones más pobres, aunque, al igual que con el PSNR, dependerá mucho del contexto.

2.2. Aprendizaje profundo

En la última década, el aprendizaje profundo o *deep learning* ha supuesto una revolución para diversos campos dentro de la computación y uno de ellos es precisamente la superresolución de imágenes. El aprendizaje profundo es una forma de aprendizaje automático que permite a las computadoras aprender de la experiencia y comprender el mundo en términos de una jerarquía de conceptos. Esta jerarquía de conceptos permite a la computadora aprender ideas complejas construyéndolas a partir de otras más simples (K. G. Kim, 2016).

Si se quiere comprender cómo el aprendizaje profundo se aplica al ámbito de la superresolución, es necesario contextualizar este aprendizaje dentro del marco de la inteligencia artificial. La IA se concibió para tratar de realizar tareas humanas, como podría ser el razonamiento, a partir del diseño de sistemas basados en reglas fijas.

Esta metodología tenía ciertas limitaciones y por ello surgió el aprendizaje automático o *machine learning*, que surgió como una sub-disciplina de la IA en la cual desaparece la necesidad de crear reglas explícitas, sino que las máquinas aprenden distintos patrones a partir de los datos seleccionados, ya que, como apunta una de las definiciones más aceptadas, un programa de computadora aprende de la experiencia E con respecto a una clase de tareas T y una medida de desempeño P, si su rendimiento en las tareas de T, medido según P, mejora con la experiencia E (*Machine Learning textbook*, s. f.). Dentro del sub-campo del ML encontramos distintos tipos de aprendizaje, como son el supervisado, no supervisado y por refuerzo.

En este contexto es donde aparece el DL mencionado anteriormente, el cual es una evolución directa del ML y que se diferencia de este último en la inclusión de arquitecturas llamadas redes neuronales profundas. (*Aprendizaje profundo | Aula En Abierto*, s. f.)

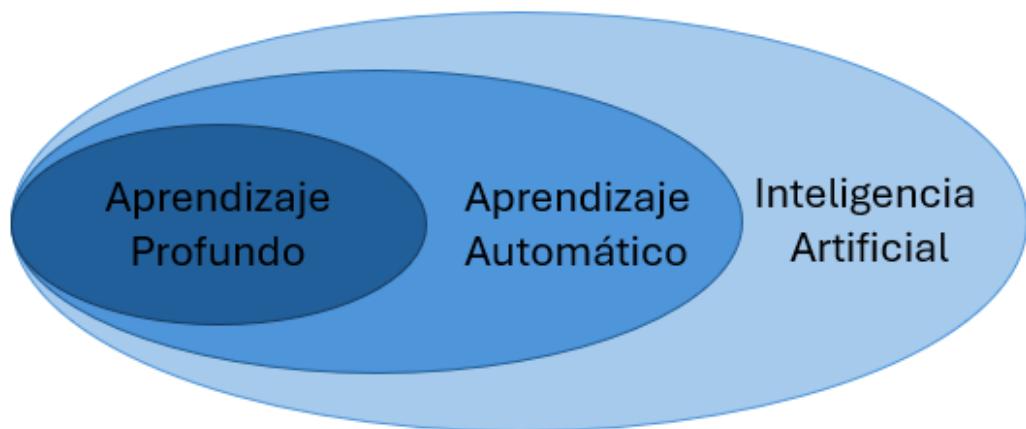


Ilustración 3. Aprendizaje profundo, automático e inteligencia artificial.

2.2.1. Redes neuronales artificiales

Como acabamos de mencionar, las redes neuronales artificiales conforman el esqueleto principal del aprendizaje profundo. Una red neuronal artificial (RNA) se puede definir como un modelo computacional conformado por múltiples capas de nodos conectados, el cual se diseña con el fin de reconocer patrones complejos y realizar tareas de clasificación o generación de datos, tratando de imitar el funcionamiento de las neuronas biológicas (Nielsen, s. f.). Estas RNA están compuestas por un gran número de neuronas interconectadas; cada una de ellas recibe señales de entrada, las procesa y devuelve una señal de salida.

De esta forma, estas redes están formadas por un conjunto de **sinapsis ponderadas** (pesos), un **sumador** que acumula los datos de entrada multiplicados por sus pesos correspondientes y una **función de activación** que limita la amplitud de la salida de la neurona (Z. Zhang, 2018). (Incio-Flores et al., 2023)

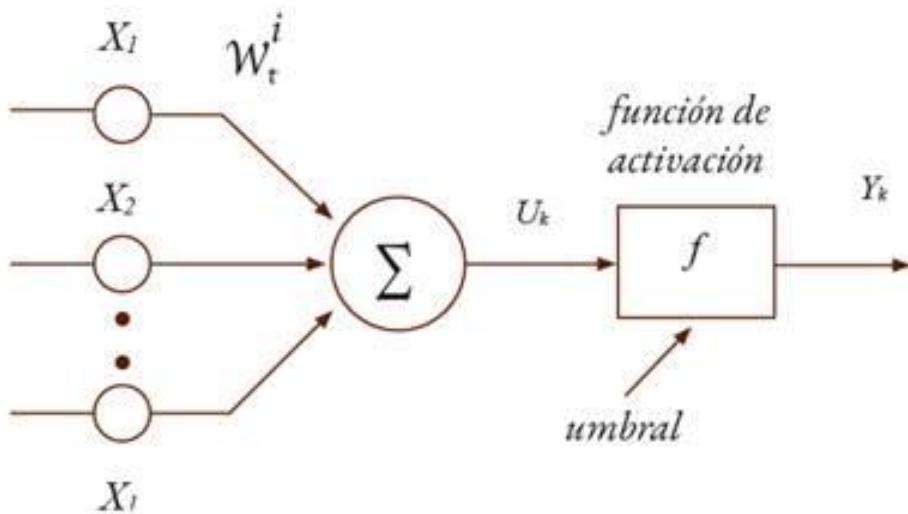


Ilustración 4. Estructura básica de una neurona artificial. Modificado de (Incio-Flores et al., 2023)

Si nos enfocamos en el campo de la superresolución de imágenes, estas RNA han demostrado desde el comienzo y las primeras aplicaciones que tenían una gran capacidad para aprender la relación no lineal que se da entre los pares de imágenes de alta y baja resolución.

La capacidad que tienen estas redes para capturar patrones complejos dejó a la vista el potencial que tenían estas frente a los métodos tradicionales basados en interpolación. A partir de estas primeras redes, se desarrollaron arquitecturas más potentes como las redes neuronales convolucionales o las redes generativas adversarias que han perdurado hasta nuestros días, enfocándose en diversas tareas de múltiples ámbitos.

2.2.2. Redes neuronales convolucionales

Como hemos visto, las redes neuronales convolucionales son un tipo especializado de RNA. Una red neuronal convolucional (CNN) es una clase de red neuronal artificial diseñada para procesar datos con estructura en forma de cuadrícula, como las imágenes. Utiliza capas convolucionales que aplican filtros para extraer características jerárquicas y espaciales de los datos de entrada, permitiendo aprender patrones complejos mediante entrenamiento supervisado (Heaton, 2018). Enfocándonos en la superresolución, estas redes son capaces de extraer automáticamente características de imágenes, como pueden ser los bordes, texturas y formas, lo que las hace especialmente eficaces en el ámbito de este proyecto.

Estas redes tienen una estructura diferenciada de las RNA y están comúnmente formadas por cuatro tipos de capas. En primer lugar, encontramos las **capas convolucionales** que dan nombre a estas redes y que son las que aplican distintos filtros o *kernels* que se deslizan por la imagen para extraer las características locales de esta.

En segundo lugar, se encuentran las **capas de activación**, que son aquellas que introducen no linealidades, permitiendo así el aprendizaje de relaciones complejas. En tercer lugar, las **capas de agrupamiento** o *pooling layers* reducen la dimensionalidad y conservan la información más relevante. Finalmente encontramos las **capas totalmente conectadas**, las cuales procesan las características extraídas y generan una salida final. (*Stanford University CS231n: Deep Learning for Computer Vision*, s. f.). («(PDF) Inteligencia Artificial En La Enfermedad de Parkinson y Otros Trastornos Del Movimiento», s. f.)

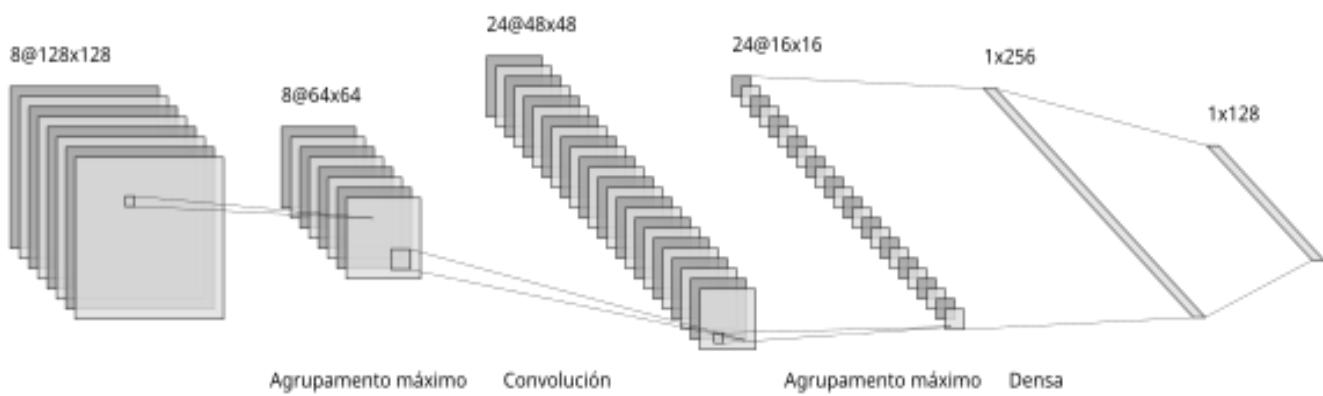


Ilustración 5. Estructura base de una CNN. (Autor: (LauAI, 2024), material distribuido bajo la licencia CC_BY 4.0).

En el proceso de entrenamiento de estas redes convolucionales, las funciones de pérdida y los algoritmos de optimización son imprescindibles para conseguir que los distintos modelos converjan y se obtenga un buen resultado final. Las funciones de pérdida se utilizan para cuantificar la diferencia entre las salidas predichas y las etiquetas reales, guiando el proceso de optimización para minimizar errores (*Loss Functions in Deep Learning: A Comprehensive Review*, s. f.).

A su vez, los algoritmos de optimización son una herramienta utilizada en el aprendizaje profundo para actualizar los parámetros del modelo y minimizar la función de pérdida definida, con el objetivo de mejorar el rendimiento del modelo al minimizar el valor de la función objetivo (*Optimization Algorithm - an overview* | ScienceDirect Topics, s. f.).

En el contexto específico de la superresolución de imágenes, la función de pérdida más comúnmente empleada es el error medio cuadrático (RMS), el cual realiza una evaluación de la diferencia promedio al cuadrado entre los píxeles de la imagen generada y la imagen original. En cuanto a la optimización, entre los métodos más ampliamente utilizados se encuentra el descenso por gradiente estocástico (SGD), que ajusta los pesos de forma iterativa a partir de estimaciones parciales del gradiente, permitiendo una convergencia eficiente incluso en grandes conjuntos de datos (Ruder, 2017).

Para la obtención de un buen entrenamiento de una CNN, es indispensable el ajuste o configuración de diversos parámetros llamados hiper-parámetros, que son los encargados de controlar el proceso de aprendizaje. Algunos de los hiper-parámetros más relevantes y comunes en este tipo de redes son el **tamaño del kernel**, que define el área local en la que se detectarán patrones y el **stride**, que determina cuánto se desplazará ese filtro sobre la imagen. También encontramos el **padding** que agregará ceros en los bordes de la imagen si es necesario para conservar las dimensiones de la imagen.

El **número de filtros** afectará a cómo el modelo podrá aprender distintos patrones. El **número de capas convolucionales** dictaminará la profundidad de la red y su capacidad para aprender representaciones complejas. A su vez encontramos el tamaño del **batch**, el cual ayuda a la convergencia del entrenamiento. Y por último, la **tasa de aprendizaje** que controla cómo se realizan las actualizaciones de los pesos durante el entrenamiento (Dong et al., 2016).

2.3. La superresolución

Como se ha mencionado en la introducción, la superresolución es una tarea dentro del procesamiento de imágenes que tiene como objetivo aumentar la resolución de una imagen, es decir, generar una imagen de mayor definición a partir de una de baja resolución. Esta tarea es de gran importancia en aplicaciones donde la nitidez de los detalles es crucial.

2.3.1. Métodos tradicionales de superresolución (Nasrollahi & Moeslund, 2014)

Previo al auge del aprendizaje profundo, la superresolución se abordaba a través de métodos algorítmicos clásicos como los siguientes.

- **Interpolación por vecino más cercano:** este método asigna al nuevo píxel el valor del más próximo, produciendo resultados pixelados.
- **Interpolación bilineal y bicúbica:** calculan valores intermedios usando el entorno de píxeles vecinos. Aunque generan imágenes más suaves, tienden a perder detalles y no pueden recuperar información de alta frecuencia.
- **Métodos basados en reconstrucción:** es un método que utiliza múltiples imágenes donde se combinan varias tomas ligeramente desplazadas para obtener una imagen más nítida.
- **Modelos basados en aprendizaje estadístico:** como las representaciones dispersas, que trataban de aprender relaciones entre parches LR y HR, pero con limitaciones de generalización y escalabilidad.

2.3.2. Superresolución basada en redes neuronales

Con la aparición del aprendizaje profundo, la superresolución ha cambiado de forma significativa. Se han desarrollado redes neuronales que aprenden a generar imágenes de alta resolución utilizando un gran volumen de datos.

Estos modelos presentan una diferencia fundamental frente a los métodos clásicos y es que, en vez de aplicar reglas fijas, estos aprenden relaciones no lineales de los patrones espaciales existentes en las imágenes. Las CNN vistas anteriormente son especialmente adecuadas para esta tarea, puesto que tienen una gran capacidad para capturar características locales y globales. Esta capacidad del aprendizaje profundo ha provocado el diseño de redes convolucionales específicas para las tareas de superresolución, incluyendo características diferenciadoras frente a las redes CNN utilizadas en tareas de clasificación o detección de objetos.

En primer lugar, las redes de superresolución no buscan generar una etiqueta de salida, sino que su objetivo es **reconstruir una imagen de mayor resolución** a partir de una degradada. Por ello, estas redes, a diferencia de otras, mantienen la estructura completa de la imagen a lo largo de toda la red, sin utilizar operaciones de reducción, con el fin de no eliminar información crítica. Otra diferencia es que estas redes integran **mecanismos de reconstrucción** de alta resolución, como convoluciones transpuestas, con el fin de generar detalles más definidos.

También encontramos diferencias en el entrenamiento de los modelos, ya que en las redes de superresolución se utilizan **métricas perceptuales** como el SSIM que ayudan a preservar la calidad percibida por el usuario.

Otra característica diferenciadora es la tendencia a utilizar **redes más profundas y con conexiones residuales**, intentando así capturar el contexto completo pero manteniendo la precisión de los detalles locales (Dong et al., 2016b; Z. Wang et al., 2020).

Por último, encontramos que, en los últimos años, además de los enfoques tradicionales, ha surgido un creciente interés por el uso de técnicas de aprendizaje generativo, especialmente mediante redes generativas adversarias (GANs) en el contexto de la superresolución. Estas redes, las cuales se basan en un esquema de dos modelos (generador y discriminador) permiten generar imágenes lo más perceptualmente realistas posibles. Un ejemplo puede ser la arquitectura SRGAN (Ledig et al., 2016), la cual introdujo una combinación del uso de la perdida adversaria y la perdida perceptual, consiguiendo reconstruir las texturas y los detalles de una forma muy satisfactoria (Goodfellow et al., 2014).

3. ESTADO DEL ARTE

3.1. Arquitecturas relevantes

Las CNN han demostrado ser una de las herramientas más eficaces a la hora de abordar el problema de la superresolución. A lo largo de la última década han surgido numerosas arquitecturas basadas en las redes convolucionales, intentando con ello mejorar la calidad perceptual y la estructura de las imágenes reconstruidas.

A continuación, se presentarán las arquitecturas más relevantes para trabajos de superresolución con imágenes satelitales y ortoimágenes.

3.1.1. SRCNN

Una de las primeras arquitecturas más relevantes fue la *Super-Resolution Convolutional Neural Network* (SRCNN) (Dong et al., 2016c), que se basa en el aprendizaje supervisado y la cual introdujo una red neuronal profunda que era capaz de aprender a comparar los pares de imágenes de alta y baja resolución y la cual estaba formada por tres capas principales en las que se extraían parches de las imágenes LR; a continuación, se aplicaban transformaciones no lineales y, por último, se producía la imagen de alta resolución. Se utilizaba la función de activación ReLU (función unidad rectificada lineal) y se empleaba el MSE como función de pérdida.

Esta arquitectura fue rápidamente utilizada y adaptada para múltiples aplicaciones, incluyendo el procesamiento de imágenes satelitales y ortoimágenes. Por ejemplo, en el trabajo de (Pouliot et al., 2018) se utilizó una versión adaptada de esta red para mejorar la resolución de imágenes provenientes del satélite Landsat-8, logrando mejorar la nitidez de los detalles de áreas de carácter urbano y agrícola.

3.1.2. VDSR

Posteriormente, se propuso la arquitectura VDSR (J. Kim et al., 2016a), de aprendizaje supervisado, y que introduce una estructura más profunda compuesta por 20 capas convolucionales con activaciones ReLU y MSE como función de pérdida. Esta arquitectura se basa en el aprendizaje residual, en el que la red no intenta reconstruir la imagen HR completa, sino los detalles ausentes de la imagen interpolada, lo que provoca una mejora significativa en la precisión de reconstrucción de las imágenes.

VDSR ha sido aplicada en diversos estudios; en concreto si nos centramos en el ámbito de la superresolución, encontramos trabajos como el de (Panagiotopoulou et al., 2020), donde los

investigadores reentrenaron la red VDSR a partir de imágenes de Sentinel-2 y ortoimágenes obtenidas por drones con el fin de mejorar la resolución de este tipo de imágenes para aplicaciones en teledetección.

3.1.3. SRResNet

Unos meses más tarde se desarrollaría la arquitectura *Super-Resolution Residual Network* (SRResNet) (Ledig et al., 2017) que, como su nombre indica, estaba inspirada en la red ResNet. Esta red de aprendizaje supervisado está formada por bloques residuales formados por múltiples capas convolucionales con normalización por lotes y activaciones ReLU, introduciendo mejoras con respecto a la arquitectura base. Esta arquitectura no solo fue optimizada para mejorar métricas como el PSNR, sino que se buscó construir los detalles visuales de una forma más realista.

En cuanto a los trabajos relacionados con imágenes satelitales, encontramos estudios como el de (Kramer et al., 2024), donde se utilizaron diversas arquitecturas como SRResNet para la superresolución de imágenes del Sentinel-2, obteniendo buenos resultados en las métricas de PSNR y SSIM.

3.1.4. EDSR

Basándose en SRResNet y por lo tanto de aprendizaje supervisado, surgió la arquitectura *Enhanced Deep Super-Resolution Network* (EDSR) (Lim et al., 2017), la cual eliminaba las capas de normalización por lotes para poder incrementar así la flexibilidad del modelo. Junto con ello, EDSR incorporó una escala residual en la que se multiplicaban las salidas de los bloques residuales por un factor pequeño antes de sumarlas a la señal principal. Esta técnica ayudó a estabilizar el entrenamiento de redes profundas, permitiendo la construcción de modelos con mayor cantidad de capas sin comprometer la convergencia. Estas modificaciones permitieron a EDSR superar a SRResNet en varios conjuntos de datos.

Debido a sus buenos resultados, fue rápidamente utilizada para labores de superresolución para labores de teledetección. Por ejemplo, en el estudio (White et al., 2021), se evaluó la eficacia y desempeño de EDSR en la mejora de imágenes satelitales degradadas, obteniendo resultados satisfactorios a la hora de recuperar detalles espaciales que se habían perdido durante la degradación.

3.1.5. RDN

Otro enfoque importante fue la creación de la arquitectura *Residual Dense Network* (RDN) (Y. Zhang, Tian, et al., 2018), también de carácter supervisado, la cual combinaba los bloques densos con una estructura residual. Cada capa dentro de un bloque estaba conectada con las capas anteriores,

facilitando así el flujo de la información y evitando posibles redundancias. Tras múltiples bloques de estas características, la salida pasaba por un módulo en el que se reestructuraban los mapas de características de la imagen para generar la imagen de alta resolución.

RDN también ha sido adaptada y aplicada para la mejora de resolución espacial de imágenes satelitales. En el estudio de (Jayanarayan et al., 2020), los investigadores implementaron esta arquitectura, consiguiendo mejorar de forma exitosa la resolución de las imágenes y obteniendo métricas satisfactorias. Cabe destacar que en este estudio se utilizó el aprendizaje por transferencia debido a que se tenía una limitada disponibilidad de muestras de entrenamiento.

3.1.6. RCAN

Por último, y siguiendo la misma línea que la red RDN, surgió la red *Residual Channel Attention Network* (RCAN) (Y. Zhang, Li, et al., 2018), que incorporaba mecanismos de atención al canal, permitiendo a la red enfocarse en los canales o bandas que contenían la información más relevante o crítica. RCAN utiliza el aprendizaje supervisado y se basa en bloques residuales profundos organizados en un esquema de Residual-in-Residual (RIR), que facilita el entrenamiento de redes muy profundas al permitir el flujo eficiente de información y gradientes.

Además, el mecanismo de atención al canal (CA) recalibra las características de cada canal considerando las interdependencias entre ellos. Así, RCAN demostró ser eficaz en la reconstrucción de detalles finos incluso en estructuras de imágenes complejas.

Los estudios en los que RCAN ha sido utilizada son muy variados. En cuanto a su aplicación con imágenes satelitales y ortoimágenes, encontramos estudios como el de (Zhu et al., 2021), donde RCAN fue adaptada para superresolucionar imágenes satelitales y ortoimágenes. La mejora de resolución que obtuvieron las imágenes fue crucial para la clasificación de coberturas del suelo.

3.1.7. Arquitecturas actuales

En conjunto, las arquitecturas de las que se acaba de hablar han establecido las bases estructurales y teóricas del aprendizaje profundo en el ámbito de la superresolución de imágenes. Cada una de ellas ha introducido innovaciones como la profundidad, el uso de bloques residuales, mecanismos de atención, etc. A partir de estas arquitecturas, han surgido en los últimos años algunas variantes que incorporan mejoras específicas, pero manteniendo estructuras muy parecidas a las originales.

A continuación, se mostrará una tabla modificada del estudio de Wang et al. con las arquitecturas más relevantes, con sus respectivas métricas y el número de parámetros generados, y en la que nos apoyaremos más adelante para sustentar la elección de la arquitectura utilizada en la realización del proyecto.

Los resultados de la tabla han sido obtenidos utilizando el conjunto de datos **Set14**, uno de los conjuntos más comunes en la evaluación de arquitecturas de superresolución. Set14 está compuesto por 14 imágenes de diversa temática, por lo que nos permite analizar el rendimiento general de los modelos. Es importante señalar que las métricas mostradas corresponden al factor de escalado $\times 2$, aunque los autores también muestran los resultados obtenidos con escalas mayores ($\times 3$ y $\times 4$), y estos, siguen una tendencia muy similar en cuanto al rendimiento entre arquitecturas.

Tabla 1. Comparación de PSNR y SSIM para el Set14 y escalado x2. Modificado de (X. Wang et al., 2022)

Modelos	Año de publicación	PSNR/SSIM	Parámetros
SRCNN	2014	32.45/0.9067	57K
VDSR	2016	33.03/0.9124	665K
DRCN	2016	33.04/0.9118	1.8M
DRRN	2017	33.23/0.9136	297K
CARN	2018	33.52/0.9166	1.6M
EDSR	2017	33.92/0.9195	43M
ELAN	2022	33.94/0.9207	8.3M
MSRN	2018	33.74/0.9170	6.5M
RCAN	2018	34.12/0.9216	16M
HAN	2021	34.16/0.9217	16.1M
RDN	2018	34.10/0.9218	22.6M
NLSN	2021	34.08/0.9231	16.1M
RFANet	2021	34.16/0.9220	11M
SAN	2019	34.07/0.9213	15.7M
SMSR	2021	33.64/0.9179	1M
TDPN	2021	34.16/0.9225	12.8M
SwinIR	2021	34.46/0.9250	12M

3.2. Arquitectura seleccionada

A pesar de la amplia variedad de arquitecturas avanzadas que han surgido en los últimos años en el contexto de la superresolución de imágenes, en este proyecto se ha optado por emplear una arquitectura VDSR como modelo principal. Esta decisión se ha tomado teniendo en cuenta distintos factores, buscando alinear al máximo la arquitectura escogida con los requerimientos del proyecto. En primer lugar, VDSR es una arquitectura que tiene una profundidad y robustez suficientes para conseguir los resultados que buscamos, como se puede deducir de su desempeño frente a los *datasets* más desafiantes (X. Wang et al., 2022) (J. Kim et al., 2016a).

Por otra parte, aunque se ve una mejora muy notoria frente a arquitecturas más simples, VDSR sigue manteniendo una estructura muy eficiente desde el punto de vista computacional con respecto a otros modelos. Esta gran eficiencia se puede visualizar en la generación de parámetros de la Tabla

1, donde se visualiza perfectamente como VDSR tiene un gasto computacional mucho menor que otros modelos.

Si comparamos a VDSR con el modelo que mejores métricas cuantitativas presenta en la Tabla 1, (NLSN), VDSR solo es superada por un 0.95% de rendimiento, mientras que NLSN genera 24,21 veces más parámetros que VDSR, evidenciando un gasto computacional mucho mayor. La cualidad de la eficiencia ha sido definitiva para la elección de VDSR, puesto que en este proyecto se pretendía procesar y gestionar un gran volumen de imágenes.

Otra de las ventajas que han decantado su elección frente a otras arquitecturas es la capacidad que VDSR tiene para ser entrenada al mismo tiempo con factores de escalado distintos, pudiendo así utilizar un único modelo para superresolver imágenes a dos, tres y cuatro veces su resolución inicial, dependiendo de lo que el usuario escoja. Esta cualidad implica un ahorro de tiempo y recursos que, en proyectos de este tipo, puede ser una cualidad muy favorable.

La última cualidad que decidió la elección de la arquitectura fue la gran adaptabilidad que tiene VDSR para realizar contextos personalizados, ya que uno de los objetivos del proyecto era el uso de imágenes nativas de distintas fuentes para crear los pares de imágenes de baja y alta resolución y, a diferencia de otras arquitecturas en las que solo se partía de imágenes de HR y a partir de ellas se creaban sus pares LR durante el entrenamiento, VDSR parte de imágenes HR y LR creadas previamente, lo que facilita en gran medida la ejecución de este objetivo.

3.2.1. Estructura de VDSR (J. Kim et al., 2016b)

Como ya se ha comentado anteriormente, VDSR fue concebida para abordar limitaciones que tenían arquitecturas anteriores a través de la incorporación de capas más profundas y la incorporación del aprendizaje residual. A continuación, se hará una explicación más precisa del conjunto de la arquitectura.

En primer lugar, VDSR consiste en un total de 20 capas convolucionales; todas ellas contienen filtros de tamaño 3x3 con un total de 64 canales, excepto la primera y la última capa. La primera realiza una convolución sobre la imagen de entrada de baja resolución; desde la segunda capa a la decimonovena encontramos bloques de convolución de tamaño 3x3 apoyados de la función de activación ReLU.

Finalmente encontramos la última capa, en la cual se realiza una convolución formada por un único filtro para generar la imagen final. Antes de la realización del entrenamiento, se realizan recortes o

parches de 42×42 píxeles tanto de las imágenes LR como de las HR con el fin de permitir un entrenamiento más manejable y eficiente desde el punto de vista computacional.

Para mejorar la capacidad de generalización de la red, se lleva a cabo una metodología de aumento de datos (*data augmentation*), por la cual se aplican rotaciones de 90, 180 y 270 grados a las imágenes y, junto con ellas, también se realizan giros horizontales y verticales.

En redes profundas de este estilo se debe prevenir el estancamiento de gradiente o la posible explosión de este; por ello se utilizan tasas de aprendizaje altas, de entorno al 0.1 en la mayoría de los casos, permitiendo una convergencia rápida. A su vez, se utiliza un recorte de gradiente, que permite mantener la estabilidad del aprendizaje. Y por último tenemos una inicialización He, que está optimizada para redes de activación ReLU.

En cuanto al entrenamiento, su principal labor es minimizar el error medio cuadrático entre la imagen HR y la reconstruida por la red. La optimización se realiza a través del SGD conformado por un *momentum* de 0.9 y una penalización L2 de 0.0001.

Finalmente, la salida producida por la red tendrá el mismo tamaño que la entrada, ya que no se utiliza *padding*, lo que permite predecir incluso los bordes de la imagen.

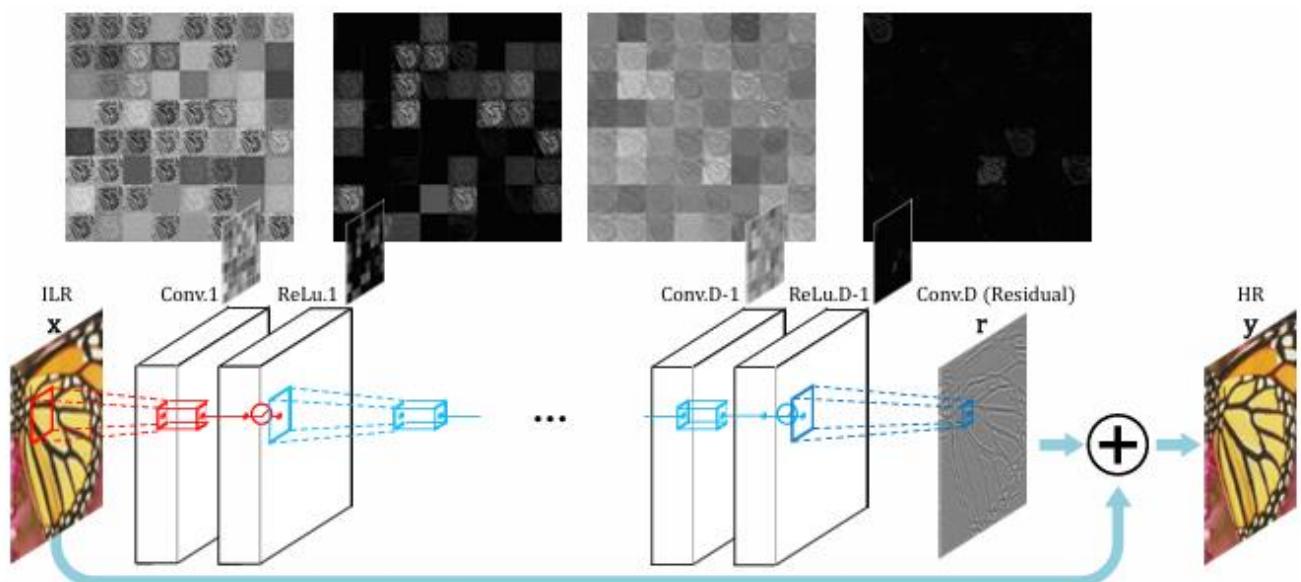


Ilustración 6. Arquitectura VDSR. Tomado de (J. Kim et al., 2016), material distribuido bajo la licencia CC_BY

3.3. Trabajos previos

Desde su publicación, VDSR ha sido una arquitectura muy utilizada en el campo de la superresolución. Numerosos trabajos han implementado VDSR en contextos muy variados.

En el ámbito médico se ha empleado esta arquitectura para la mejora de imágenes de resonancias magnéticas (J. H. Hwang et al., 2024), tomografías computarizadas (J. J. Hwang et al., 2020) y tomosíntesis digitales (Gomi et al., 2020). VDSR también ha sido utilizada en numerosas ocasiones para labores de restauración de videos y compresión de imágenes.

En el ámbito de la superresolución, existen trabajos que utilizan la arquitectura para superresolver imágenes satelitales y ortoimágenes, sin embargo, estos son muy limitados y no se ha podido encontrar ninguno en el que se utilicen pares de imágenes de distintas fuentes.

Este hecho también ha sido definitivo para la elección de VDSR, ya que supone una oportunidad para explorar si la arquitectura tiene potencial para el ámbito de la teledetección y análisis geoespacial.

4. MATERIALES

En este capítulo se realizará una descripción del hardware y software utilizado durante el proyecto.

4.1. Hardware

Para poder llevar a cabo el conjunto de entrenamientos se han utilizado los siguientes componentes hardware.

Componente	Especificación
CPU (<i>central processing unit</i>)	Intel(R) Xeon(R) W-2245 CPU @ 3.90GHz
GPU (<i>graphic processing unit</i>)	NVIDIA GeForce RTX 3050 10GB VRAM
RAM (<i>random access memory</i>)	256 GB

4.2. Software

El software utilizado durante el proyecto ha sido clave para la realización del mismo. A continuación, se detallará cada uno de los softwares empleados.

- **Python:** Es un lenguaje de programación interpretado, el cual es utilizado frecuentemente por su sencilla sintaxis y por la gran disponibilidad de bibliotecas para distintos propósitos, como análisis de datos o inteligencia artificial (*Welcome to Python.Org*, 2025).
- **PyTorch:** Es una biblioteca de código abierto principalmente utilizada para el aprendizaje profundo. Está diseñada con el fin de ofrecer una interfaz sencilla para la construcción y entrenamiento de redes neuronales, sobre todo en ámbitos de investigación y producción («PyTorch», s. f.).
- **Anaconda:** Es una distribución de Python y R que está principalmente enfocada en la ciencia de datos. Está conformada por herramientas como Conda y Anaconda Navigator, las cuales simplifican la configuración de entornos de trabajo como los utilizados para machine learning (*Anaconda Distribution*, s. f.).
- **CUDA:** Es una plataforma creada por la empresa NVIDIA que permite el uso de GPUs para acelerar aplicaciones que sean intensivas computacionalmente hablando. Facilita la

ejecución de algoritmos en entornos como el del aprendizaje profundo (*An Even Easier Introduction to CUDA (Updated)*, 2025).

- **GDAL:** Es una biblioteca de software libre que permite la lectura y escritura de múltiples formatos de datos geoespaciales. Se utiliza de manera recurrente en trabajos que implican el procesamiento de imágenes ráster y datos vectoriales (Rouault et al., 2025).
- **QGIS:** Es un software de código abierto para visualizar, editar y analizar información geoespacial. Se ha convertido en una de las herramientas más utilizadas dentro del campo de los SIG (*Spatial without Compromise · QGIS Web Site*, s. f.).
- **Zotero:** Es un gestor de referencias bibliográficas gratuito y de código abierto que permite recopilar, organizar, citar y compartir fuentes académicas (*Zotero | About*, s. f.).
- **Notepad++:** Es un editor de texto y código fuente para Windows compatible con numerosos lenguajes de programación. Ofrece una amplia variedad de funcionalidades para facilitar el manejo de código (*Notepad++, s. f.*).
- **FileZilla:** Es un cliente FTP gratuito que permite la transferencia de archivos entre servidores remotos y equipos locales, soportando una gran variedad de protocolos y ofreciendo una interfaz sencilla y eficiente (*FileZilla - The free FTP solution*, s. f.).

5. METODOLOGÍA

La metodología seleccionada para el desarrollo del proyecto está conformada por 5 bloques principales que a su vez se subdividen en diferentes tareas.

El primer bloque corresponde a la recopilación previa de información y la planificación previa para la realización del proyecto, incluyendo dentro de esta la selección de la arquitectura utilizada.

El segundo bloque comprende la creación del *dataset*, el cual estará subdividido en datos de entrenamiento, validación y test, y con los cuales se realizarán los posteriores recortes para su manejabilidad dentro de la red.

Como tercer bloque encontramos la selección de código correspondiente a la arquitectura seleccionada y las modificaciones necesarias realizadas en este para ajustarse a los requerimientos del proyecto.

El cuarto bloque está compuesto por la realización de los entrenamientos, durante los cuales se realizarán modificaciones de hiper-parámetros para conseguir los mejores resultados posibles.

Finalmente, encontramos la inferencia y evaluación de resultados, en la cual se probarán los modelos generados para evaluar su eficacia y desempeño.

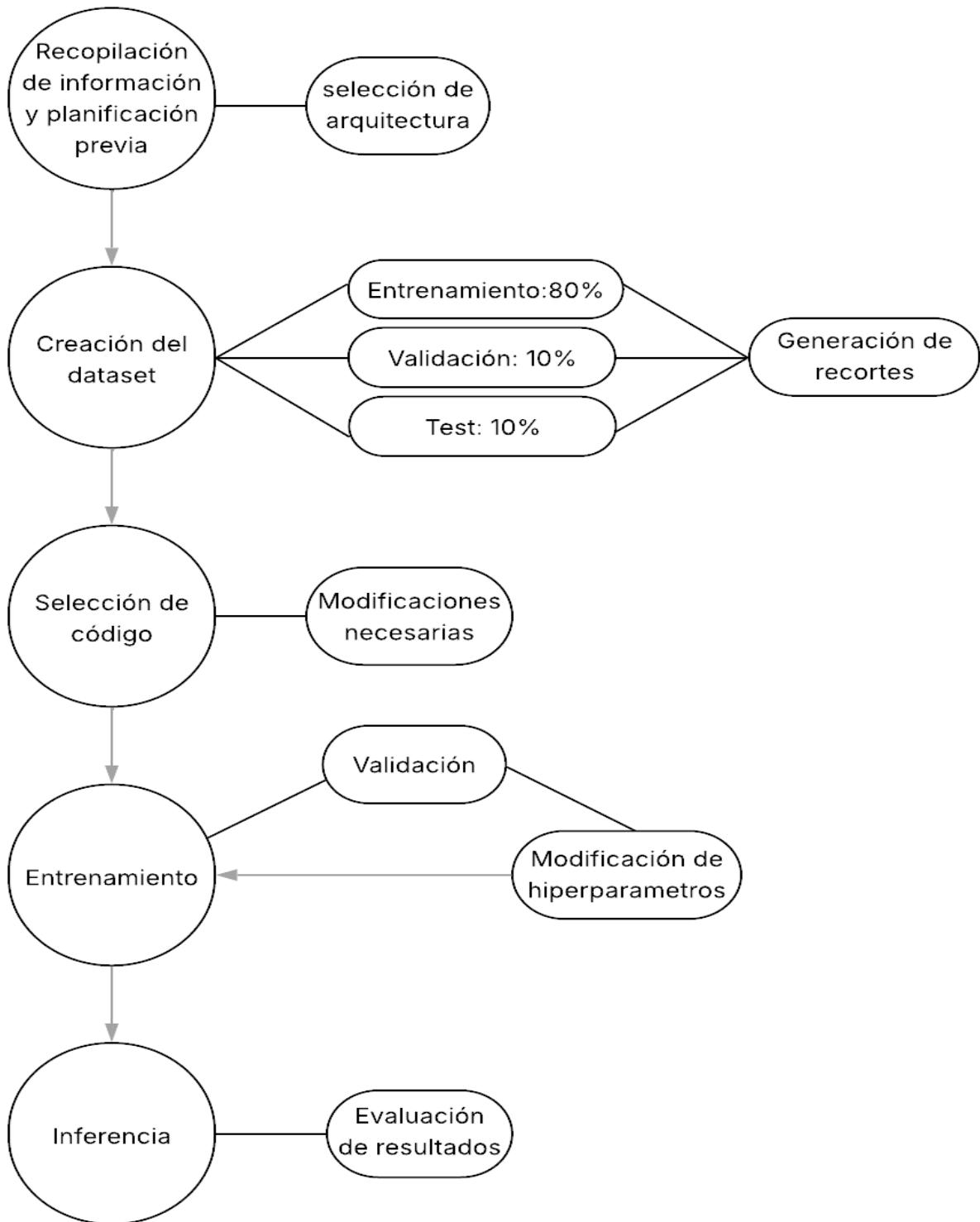


Ilustración 7. Esquema de la Metodología aplicada en el proyecto.

5.1. Generación del dataset

La generación del conjunto de datos comenzó con la recopilación de las coberturas del satélite GEOSAT2 que correspondían a los años 2021, 2022 y 2023, disponibles las tres en formato *geojson*. Estas coberturas representan los límites espaciales (BBOX) de aquellas zonas que fueron captadas

por el sensor del satélite durante esos años y se obtuvieron con el fin de seleccionar primeramente las principales áreas de las que se querían obtener estas imágenes.

Estas coberturas se visualizaron mediante la herramienta QGIS y mediante ella se pudo hacer una selección de aquellas coberturas que podrían representar las zonas de mayor interés, teniendo siempre en cuenta la distribución geográfica. Esta selección se realizó con el objetivo de hacer un primer filtrado y asegurando desde el principio un conjunto de datos lo más generalizado posible, con diferentes paisajes y posibles distinciones radiométricas, intentado así mejorar la capacidad futura del modelo a la hora de generalizar.

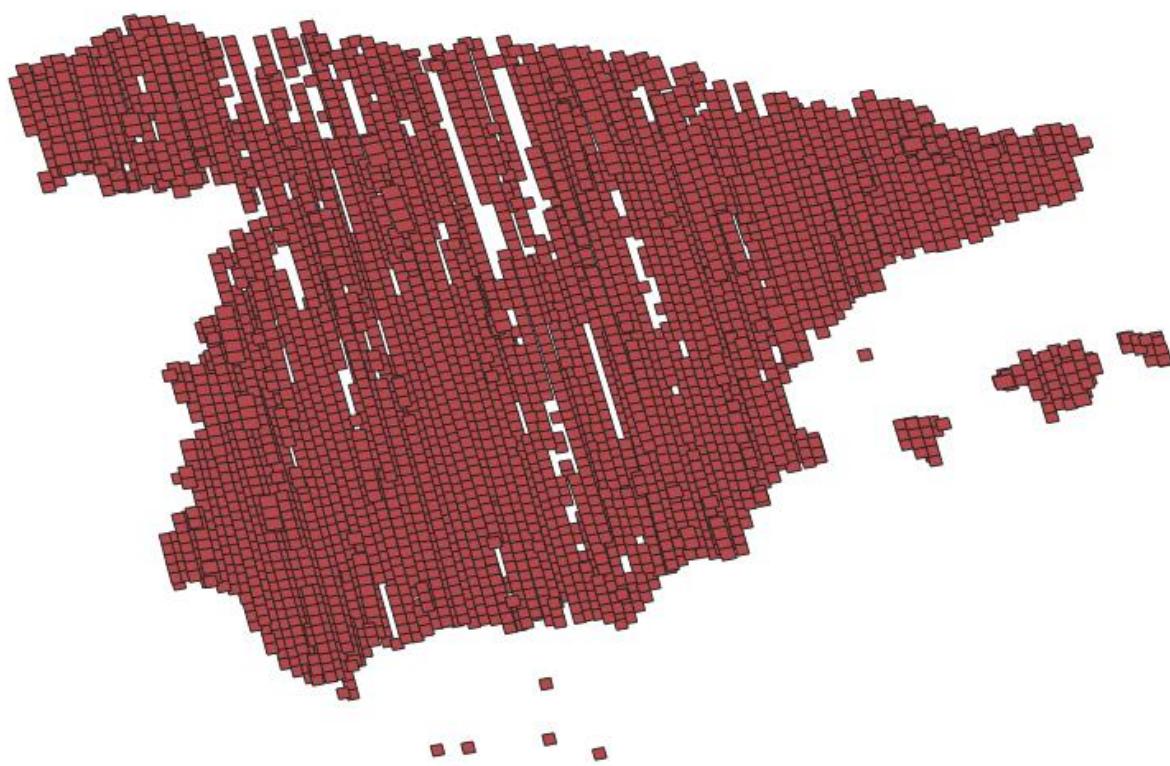


Ilustración 8. Ejemplo de coberturas para el año 2023.

Una vez seleccionadas las primeras coberturas, se tuvo en cuenta el segundo objetivo propuesto en el proyecto que, como ya se ha mencionado anteriormente, consistía en la obtención de pares de imágenes nativas procedentes del satélite GEOSAT2 y ortoimágenes del PNOA.

Por este motivo, se procedió a elegir el año óptimo para cada cobertura seleccionada, teniendo en cuenta los años de realización de las ortoimágenes del PNOA de máxima actualidad. Este criterio de selección permitiría minimizar las diferencias estacionales o de transformaciones del terreno entre los pares de imágenes, asegurando la mayor similitud entre los pares de baja y alta resolución y ayudando así al futuro aprendizaje de la red.

Con las coberturas seleccionadas y junto con sus identificadores correspondientes, se pudo proceder a la descarga de las imágenes asociadas a dichas zonas. Esta descarga se realizó a través del software de FileZilla, con el que se pudo establecer una conexión con el servidor remoto del Instituto Geográfico Nacional (IGN). Gracias a este acceso FTP fue posible la navegación por la estructura de carpetas del servidor para con ello localizar y traspasar los archivos correspondientes que contenían las imágenes deseadas.

Una vez realizada la transferencia de archivos, se procedió a su descompresión, puesto que estos archivos se encontraban en formato *zip*, conteniendo tanto las imágenes georreferenciadas como sus archivos asociados en formato *XML*, los cuales contenían los metadatos correspondientes.

Para el uso posterior de las imágenes, era indispensable la descompresión de todos los archivos descargados y, al tratarse de un conjunto grande de archivos con un tiempo de descompresión bastante elevado, se hizo indispensable la necesidad de realizar una descompresión automática y continua de todos los archivos comprimidos. Para realizar esta descompresión se realizaron los siguientes pasos. Primeramente, se generó un listado de los archivos a través de comandos dentro de la consola del sistema. A continuación, se generó un archivo de texto con el listado y en el que se incluyó para cada línea el comando correspondiente a la descompresión de los archivos, según indica la herramienta de descompresión 7-Zip. Por último, se generó un archivo por lotes (*bat*) a partir del archivo de texto, permitiendo su ejecución desde la consola de Windows y consiguiendo así una descompresión sin intervención manual.

Una vez completado el proceso de la descompresión de todos los archivos, se procedió a la verificación del contenido de las imágenes. Estas se visualizaron mediante la herramienta de QGIS, mediante la cual se confirmó que el tamaño del píxel de estas era de 0,75m con una profundidad radiométrica de 16 bits por canal. También se pudo comprobar cómo la imagen estaba compuesta por 4 bandas espectrales, donde las tres primeras correspondían a las bandas RGB, mientras que la cuarta hacía referencia a la banda pancromática.

Tras la verificación de la estructura de las imágenes, se pasó a definir las zonas de interés finales, que servirían como referencia espacial para recortar las imágenes. Esta tarea se realizó a través de la visualización de cada una de las imágenes seleccionadas mediante la herramienta de QGIS, donde se identificaron de manera manual las áreas que presentaban un mayor interés para el entrenamiento del modelo, cerciorándonos en todo momento de no escoger zonas con objetos en movimiento, los cuales dificultarían el entrenamiento en el caso de los pares de distintas fuentes.

Este proceso de definición consistió en la generación de puntos de nuevo mediante QGIS, que a través de su simbología se representaron como cuadrados para obtener la referencia de la zona que iban a cubrir los futuros recortes. Se generaron un total de 4500 puntos distribuidos por toda España

y se exportaron en formato de texto separado por comas (csv) manteniendo la componente espacial para aplicar los recortes conforme a esas ubicaciones.

Para posibilitar el proceso de recorte, fue necesario unificar todas las imágenes en un mosaico virtual. Por ello se creó un conjunto de datos virtual (Vrt), el cual actúa como un índice, referenciando a todas las imágenes y simulando una imagen única y unificada de todas las escenas. La generación del mosaico se realizó mediante las herramientas de la biblioteca GDAL y dio paso a la creación del *script* para el recorte de las imágenes.

El *script* creado extrae un recorte de 256x256 píxeles a partir del archivo de puntos extraído anteriormente. Para garantizar la correcta visualización de las imágenes, se tuvo que poner especial atención en las propiedades de las imágenes, puesto que, como se ha visto anteriormente, estas tenían un total de 4 bandas y valores de 16 bit por lo que se tuvo que hacer una reducción al rango estándar de 8 bits. Esta reducción provoca que los valores de intensidad no se distribuyan de forma óptima y, para evitar una representación oscura o con bajo contraste, fue necesaria una aplicación de normalización por auto-contraste.

A pesar de estas aplicaciones, las imágenes resultantes seguían mostrando problemas visuales debido a la distribución irregular de los valores radiométricos. Para poder mejorar esta situación se decidió aplicar una normalización específica para cada recorte, basada en los estadísticos de cada una de las bandas de las imágenes. Para que el *script* pudiese aplicar correctamente estos estadísticos a cada recorte, era necesario que cada punto con el que se realizaban los recortes contuviese el identificador de la imagen en la que estaba contenido. Para ello, se realizó una intersección espacial entre estos puntos y las coberturas que se seleccionaron al comienzo. Sin embargo, al existir solapes entre estas coberturas, algunos de los puntos caían en más de una imagen, generando así duplicados en los resultados. Tras analizar la función utilizada para crear el mosaico virtual, se comprobó que para la representación del mosaico en las zonas de solape se utilizaba la primera imagen leída por la función. Por tanto y teniendo en cuenta esto, se realizó un *script* para la eliminación de los puntos duplicados que no influían en la representación del mosaico, conservando únicamente los identificadores correctos para cada punto.

Finalmente, con los puntos asignados de forma correcta a sus respectivas imágenes, se modificó el *script* para la aplicación de la normalización atendiendo a los estadísticos de cada imagen y, por consiguiente, de los recortes realizados, permitiendo mejorar notablemente el contraste y, en consecuencia, la visualización de los recortes.

De forma simultánea y mediante el mismo *script*, se realizó el recorte de las imágenes del PNOA, pero en este caso el tamaño de los recortes debía ser de 768x768, ya que estas imágenes tienen 3 veces más resolución que las satelitales, por lo que si queremos la representación de la misma área

en ambas, necesitamos que los recortes del PNOA sean 3 veces más grandes que los satelitales. Los recortes del PNOA no supusieron ningún problema adicional.

Una vez generados todos los pares de recortes GEOSAT2-PNOA, se continuó con una fase final de limpieza en la que se eliminaron aquellos pares que presentaban discrepancias significativas, como la presencia de objetos en movimiento o por errores derivados de recortes ubicados en los bordes de las imágenes, que provocaron que en algunos casos los recortes quedaran incompletos.

5.1.1. División y preparación del dataset

Una vez completada la depuración del conjunto de los pares de imágenes, se procedió a su división en subconjuntos de entrenamiento, validación y prueba. Esta división siguió una distribución muy utilizada en la que el entrenamiento supone el 80% de los datos, la validación un 10% y las pruebas el 10% restante.

Esta partición se realizó mediante un script que garantizaba una separación completamente aleatoria. Junto con ello, se prepararon los recortes de entrada al tamaño requerido por la arquitectura VDSR, la cual fue creada para funcionar sobre recortes de 42x42 píxeles. Por este motivo, cada recorte previo fue dividido en múltiples parches de dicho tamaño, consiguiendo de esta forma un número elevado de muestras para el aprendizaje de la red.

Tabla 2. Número de parches generados para las fuentes del GEOSAT2 y PNOA

Fuente	Tamaño parches	Tamaño recorte	Número de imágenes	Número de parches
GEOSAT2	42x42	256x256	4500	162.000
PNOA	42x42	768x768	500/ 1500/ 2500	165.000/ 495.000/ 825.000

5.1.1.a. Preparación del primer objetivo

Tanto para las imágenes satelitales como para las del PNOA y atendiendo al primer objetivo del proyecto, se crearon los pares de imágenes mediante degradación de la resolución de las imágenes de alta resolución para escalados de x2, x3 y x4.

Para ello, se partía de los parches realizados de las imágenes originales y se generaban sus versiones re-escaladas por interpolación bicúbica. Para el caso de los parches satelitales se utilizaron los 4500 recortes. Sin embargo, en el caso de las imágenes del PNOA, se utilizaron solamente 500,

ya que, como se visualiza en la tabla1, una imagen del PNOA generaba muchos más parches que una imagen satelital debido a la diferencia de píxeles entre ambas.

Pese a ello, se realizaron aumentos del número de recortes utilizados. Inicialmente se realizaron pruebas con 1500 recortes, pero se finalizó con pruebas de hasta los 2500 recortes con el fin de conseguir los mejores resultados, pese a que el tiempo de aprendizaje se viera aumentado.

5.1.1.b. Preparación del segundo objetivo

En relación con el segundo objetivo, centrado en el uso de pares nativos de distintas fuentes, se realizaron dos enfoques para llevarlo a cabo, siempre ajustándose a los requisitos de la arquitectura VDSR.

El **primer enfoque** se basó en la generación de parches de 42x42 píxeles a partir de las imágenes GEOSAT2 y parches de 128x128 píxeles para las imágenes del PNOA, preservando la relación espacial de las imágenes. Posteriormente, los recortes del PNOA se redujeron a 42x42 píxeles mediante distintos tipos de interpolación. Esta transformación fue necesaria ya que VDSR trabaja con parches del mismo tamaño para poder realizar comparaciones. A su vez, para la realización de pruebas y obtención de resultados, se tuvo que reducir el tamaño de las imágenes de prueba del PNOA desde los 768x768 píxeles a los 256x256 píxeles para que coincidiesen con el tamaño de las imágenes del GEOSAT2. Esta reducción de tamaño se hizo mediante los mismos métodos de reducción que se habían empleado para los parches. Cabe destacar que se realizó una selección de los mejores métodos de interpolación con el fin de preservar lo mejor posible la resolución de los parches previos. Para este enfoque se utilizó el *dataset* completo de 4500 imágenes.

En el **segundo enfoque** se propuso el método inverso. Por una parte, se generaron parches de 14x14 píxeles desde las imágenes GEOSAT2 y parches de 42x42 píxeles desde las imágenes del PNOA. Luego, los parches satelitales fueron escalados a 42x42 píxeles mediante distintas técnicas de interpolación, consiguiendo de nuevo que los pares de parches tuvieran el mismo tamaño. Y, en consecuencia, y de la misma forma que en el objetivo anterior, se realizó un aumento artificial de las imágenes de prueba del GEOSAT2 desde los 256x256 píxeles hasta los 768x768 píxeles que conformaban las imágenes del PNOA. En este segundo enfoque se obtendrían muchos más parches, por lo que solo se utilizaron 500 imágenes del *dataset* inicial.

Para ambos enfoques se utilizó la misma partición de datos que para el primer objetivo: 80% del *dataset* para el entrenamiento, 10% para la validación y 10% para las pruebas.

5.2. Generación del modelo

Para el desarrollo del modelo de superresolución empleado en el proyecto, se partió del código original publicado por los autores del artículo *VDSR Accurate Image Super-Resolution Using Very Deep Convolutional Networks* de Kim et al. (Lorna, 2020/2025). Este código implementa la arquitectura propuesta y será la base sobre la que se realizarán las adaptaciones necesarias.

El código está desarrollado en lenguaje Python y está estructurado en distintos módulos que se explicarán a continuación de forma resumida.

- El módulo **train.py** es el script principal que ejecuta el entrenamiento. Carga el conjunto de datos, inicializa el modelo y aplica la configuración de los entrenamientos. Dentro de este propio entrenamiento encontramos una ejecución de validación por época.
- El módulo **model.py** contiene la definición de la arquitectura VDSR. Implementando la red convolucional de 20 capas y sus activaciones ReLU.
- El módulo **validate.py** se utiliza para la evaluación del modelo. Carga los pesos guardados y calcula el PNSR sobre el conjunto de prueba, para conseguir un resultado objetivo del modelo.
- El módulo **setup.py** se encarga de la configuración inicial del entrenamiento. Se definen las dependencias del proyecto y permite la instalación del propio mediante comandos dentro del entorno de Python.
- EL módulo **config.py** define los parámetros de configuración del proyecto, incluyendo los hiper-parámetros de entrenamiento y las opciones de escalado. Es un script de gran utilidad, puesto que nos permite modificar el comportamiento del entrenamiento sin tener que editar el código fuente.
- Por último, encontramos el módulo **imgproc.py**, el cual contiene funciones auxiliares para el procesamiento de las imágenes y actúa como soporte de los scripts principales.

5.2.1 Modificaciones realizadas

5.2.1.a. Modificaciones conformes al primer objetivo

En primer lugar, se van a definir los cambios realizados para llevar a cabo el primer objetivo del proyecto. Estos cambios comenzaron por la carga personalizada de datos, ya que la versión original del código estaba diseñada para la carga de *datasets* genéricos. Por ello se desarrolló un sistema de carga de datos adaptado a la estructura específica del conjunto de recortes generado.

En cuanto al apartado de la validación, se añadieron distintas métricas como el SSIM, que aunque sea una métrica cuantitativa, permite una valoración que se asemeja en la mayoría de los casos a la valoración perceptual que puede tener un usuario.

A su vez, se realizaron pruebas con otras funciones de optimización, cambiando el SGD que utilizaba inicialmente el código por el optimizador Adam. Sin embargo, después de la realización de los entrenamientos, se reversionó este cambio, puesto que no suponía ningún beneficio frente a la función inicial.

También se realizaron cambios en el ajuste del *learning rate* por épocas, aplicando un decrecimiento de este algo distinto al que se planteaba inicialmente, definiendo de forma distinta el número de épocas que debía pasar sin mejorar el modelo para realizar una disminución del valor de aprendizaje y permitir así que el entrenamiento mejorase en las fases finales.

Para aumentar la variedad de muestras y mejorar la generalización del modelo, se añadieron técnicas de *data augmentation* distintas a las que ya contenía el código inicial.

5.2.1.b. Modificaciones conformes al segundo objetivo

Para el segundo objetivo, en el que se iban a utilizar las imágenes del GEOSAT2 como imágenes de baja resolución y las del PNOA como imágenes de alta resolución, fueron necesarias modificaciones más notorias sobre el código base, ya que la arquitectura no estaba diseñada para manejar la información de esta manera. Para este objetivo también se utilizaron las modificaciones previas realizadas para el primero de los objetivos. VDSR está diseñada para que, durante la prueba del modelo, se generen imágenes de baja resolución a través de la interpolación de las imágenes de alta resolución, para seguidamente aplicar el modelo sobre la imagen interpolada, obteniendo un resultado que finalmente se comparaba con la imagen original de alta resolución, obteniendo así las distintas métricas de evaluación.

Sin embargo, en este proyecto, la imagen de baja resolución ya se encontraba de forma nativa, por lo que no debía ser degradada ni modificada de forma artificial. En lugar de ello, el modelo debía procesar la imagen GEOSAT2 original y, una vez aplicado el modelo sobre esta imagen, se debía comparar con su par de alta resolución del PNOA. Este nuevo enfoque obligó a la modificación de la lógica del código de validate.py junto con los métodos que utilizaban la misma lógica dentro del entrenamiento.

Así, se adaptó el código para que el sistema de lectura trabajase directamente con los pares predefinidos. Se eliminaron las operaciones de interpolación interna durante las pruebas del modelo y a su vez se ajustaron las funciones de cálculo de métricas para que estas compararan correctamente la salida producida por el modelo y la imagen de alta resolución a la que hacía referencia.

Estas modificaciones fueron imprescindibles para poder cumplir con el segundo objetivo del proyecto y poder mostrar a su vez una evaluación precisa de los resultados para deducir objetivamente si los modelos generados eran de utilidad.

5.3. Entrenamiento

Una vez adaptado el código y preparado el conjunto de datos para ambos objetivos, se procedió a la fase de entrenamiento del modelo. Este proceso consistió en la búsqueda de los mejores modelos para cada escenario. Para conseguir mejoras progresivas, se necesitó la realización de ajustes en los pesos de la red mediante aprendizaje supervisado, y a su vez, para mejorar estos pesos, fue necesaria la realización de pruebas utilizando diferentes combinaciones de hiper-parámetros. Cabe destacar que los entrenamientos dirigidos para cada objetivo fueron realizados por separado, pero en ambos se utilizaron valores de hiper-parámetros muy similares, por lo que se hablará de ellos en relación con ambos objetivos.

A continuación, se repasarán los principales hiper-parámetros utilizados en la búsqueda de los mejores modelos.

- **Número de épocas:** Se realizaron pruebas con 80, 200, 300 e incluso 400 épocas, observando por lo general mejoras a medida que aumentábamos el número.
- **Tamaño del lote o *batch size*:** El tamaño del lote se fue aumentando con cada entrenamiento hasta que se llegó al máximo que soportaba la tarjeta gráfica, llegando hasta un tamaño de 30.
- **Tasa de aprendizaje:** En los primeros entrenamientos se partió de una tasa de 0.1, aunque se realizaron diversas pruebas con valores menores como 0.05 o 0.01. Además de la disminución de este valor, también se realizaron pruebas con distintos valores de reducción progresiva de esta tasa, como por ejemplo la división entre 5 de la tasa de aprendizaje cada 5 épocas en las que no se había producido una mejora en el entrenamiento.
- **Optimización:** Como ya se ha comentado anteriormente, el optimizador principal fue el SGD con un *momentum* de 0.9. No obstante, se realizaron pruebas con el optimizador Adam sin obtención de mejoras aparentes con respecto al anterior.
- **Función de pérdida:** Se utilizó el MSE como función de pérdida, tal y como presenta de inicio la arquitectura VDSR, por lo que en este aspecto no se realizó ninguna modificación.
- **Recorte de gradiente:** Se utilizaron los valores de 0.1 y 0.15 para este hiper-parámetro, buscando evitar en todo momento la explosión del gradiente y una mayor estabilidad de este.

- **Número de datos:** Aunque este número no sea un hiper-parámetro como tal, sí que repercute directamente en el aprendizaje de la red, por lo que se debe mencionar que se han utilizado diferentes tamaños de *dataset*, partiendo desde las 500 imágenes y llegando hasta las 4500.

Todo el proceso de entrenamiento se llevó a cabo en numerosas sesiones en las que se reutilizaron pesos guardados y se exploraron distintas combinaciones de los parámetros ya mencionados. Cabe destacar que la guía principal para la obtención de mejoras de rendimiento a través del cambio de hiper-parámetros fueron las métricas de PSNR y SSIM, cuyo aumento o disminución cuantitativa dirigía el cambio de los hiper-parámetros mencionados.

Así, con este enfoque iterativo de entrenamiento, validación y cambio de parámetros, fue posible mejorar progresivamente el comportamiento del modelo.

5.4. Inferencia y evaluación de resultados

Una vez finalizado el proceso de entrenamiento, se procedió a la fase de inferencia y valoración de resultados, cuyo objetivo es comprobar la capacidad que tiene el modelo para superresolucionar imágenes a partir de imágenes de baja resolución y, por lo tanto, evaluar de forma objetiva y perceptual la calidad de estas reconstrucciones.

En primer lugar, debemos recordar que, a pesar de que la red entrena con parches de 42x42 píxeles, las imágenes a las que se aplica el modelo pueden ser del tamaño que se desee, aunque, como se ha explicado anteriormente, para poder obtener métricas precisas necesitamos que la imagen reconstruida y la de referencia sean del mismo tamaño.

A continuación, se describirá el proceso seguido para la evaluación de resultados, el cual es el mismo para ambos objetivos del proyecto, aunque parte de la lógica del código de prueba no sea la misma.

Este proceso comienza con la aplicación de los modelos sobre el conjunto de las imágenes de prueba. Una vez obtenidas las salidas, estas se compararán con sus correspondientes imágenes de alta resolución que, en el caso del primer objetivo, serán imágenes del GEOSAT2 y PNOA respectivamente y, en el caso del segundo objetivo, serán solo las imágenes del PNOA.

A continuación, se realiza la evaluación cuantitativa para una valoración más objetiva de las reconstrucciones. Para ello se utilizarán las dos métricas mencionadas en múltiples ocasiones, el PSNR y el SSIM.

Además del análisis numérico, se realizará una revisión visual de los resultados. Esta valoración permitirá identificar errores perceptuales como artefactos o diferencias visuales relevantes que se hayan reflejado en las métricas cuantitativas. Finalmente, se realizarán comparaciones entre los

distintos modelos generados y a su vez con resultados producidos por métodos de interpolación clásicos, lo que nos permitirá valorar la mejora real obtenida mediante el uso del aprendizaje profundo frente a las técnicas clásicas.

La combinación de las evaluaciones objetivas y cuantitativas nos permitirá medir el rendimiento del modelo desde un punto de vista técnico, a la vez que podremos valorar su utilidad en escenarios reales, como se apuntaba en la introducción del proyecto.

6. RESULTADOS

En este apartado se van a presentar los resultados obtenidos tras la ejecución de la metodología explicada anteriormente. Estos resultados van a estar divididos según el objetivo al que atendían en el proyecto, comenzando por el objetivo número 1 y finalizando con el segundo. Los resultados del primer objetivo se dividirán a su vez en los modelos generados para las imágenes GEOSAT2 y los generados para las ortoimágenes del PNOA. En ambos casos, los resultados se van a presentar en forma de tabla, mostrando los hiper-parámetros utilizados junto con las métricas cuantitativas y, en el caso del primer objetivo, haciendo distinción entre los distintos factores de escalado.

Algunos hiper-parámetros han sido constantes durante los entrenamientos finales, por lo que se mostrarán a continuación en una tabla, con el fin de no mostrarlos repetidamente durante la evaluación de los modelos. Como añadido a las distintas tablas de resultados, se mostrarán las imágenes generadas por los modelos y se realizarán comparaciones para poder realizar evaluaciones perceptuales además de las cuantitativas.

Todos los modelos resultantes junto con el código utilizado se encuentran en

https://github.com/ignacioVallejo03/TFG_Ignacio_Vallejo

Tabla 3. Hiper-parámetros constantes en los modelos finales.

Tamaño del lote	Número de hebras en paralelo	Momentum del modelo	Tasa de decaimiento	Número imágenes GEOSAT2
30	10	0.9	0.0001	4500

6.1. Resultados del Objetivo_1

6.1.1. Evaluación de modelos para imágenes del GEOSAT2

Como ya se ha mencionado, los modelos han sido entrenados para tres escalas de superresolución, intentando así escalar imágenes al doble, triple y cuádruple de su resolución original. En las siguientes tablas se mostrarán los resultados obtenidos para los mejores modelos encontrados, haciendo distinción entre los distintos factores de escalado. Sin embargo, como ya sabemos, VDSR es capaz de entrenarse para varios escalados al mismo tiempo, generando un solo modelo, por lo que también se mostrarán las métricas obtenidas de este modelo concreto.

6.1.1.a. Factor de escalado = 2

Tabla 4. Métricas de los modelos para imágenes de GEOSAT2 con factor de escalado x2

ID modelo	Tasa de aprendizaje	Recorte de gradiente	Épocas	Factor de escalado	PNSR (dB)	SSIM
1	0.01	0.15	200	2	45.65	0.9894
2	0.01	0.1	200	2	46.14	0.9900
3	0.1	0.1	80	2	47.04	0.9913

La tabla muestra los resultados de los tres mejores modelos obtenidos para el factor de escalado x2. En el modelo 1 se utilizó una tasa de aprendizaje de 0.01 con un valor de recorte de gradiente de 0.15 y 200 épocas debido al tamaño reducido de la tasa de aprendizaje. Como se obtuvieron unos buenos resultados, se buscó obtener información sobre qué tasa de recorte de gradiente era la más óptima, por lo que en el modelo 2 solo se modificó esta tasa, obteniendo mejores resultados con una tasa algo menor que en el anterior ejemplo. Por último, para el tercer modelo se buscó variar la tasa de aprendizaje y con ello el número de épocas, debido a que con una tasa mayor la convergencia sería más rápida. Se mantuvo el recorte de gradiente a 0.1 por su rendimiento anterior. Como se puede observar, este tercer modelo fue el que mejores métricas obtuvo, sugiriendo que una tasa de aprendizaje más agresiva puede ser más efectiva si es combinada con un recorte de gradiente suficiente.

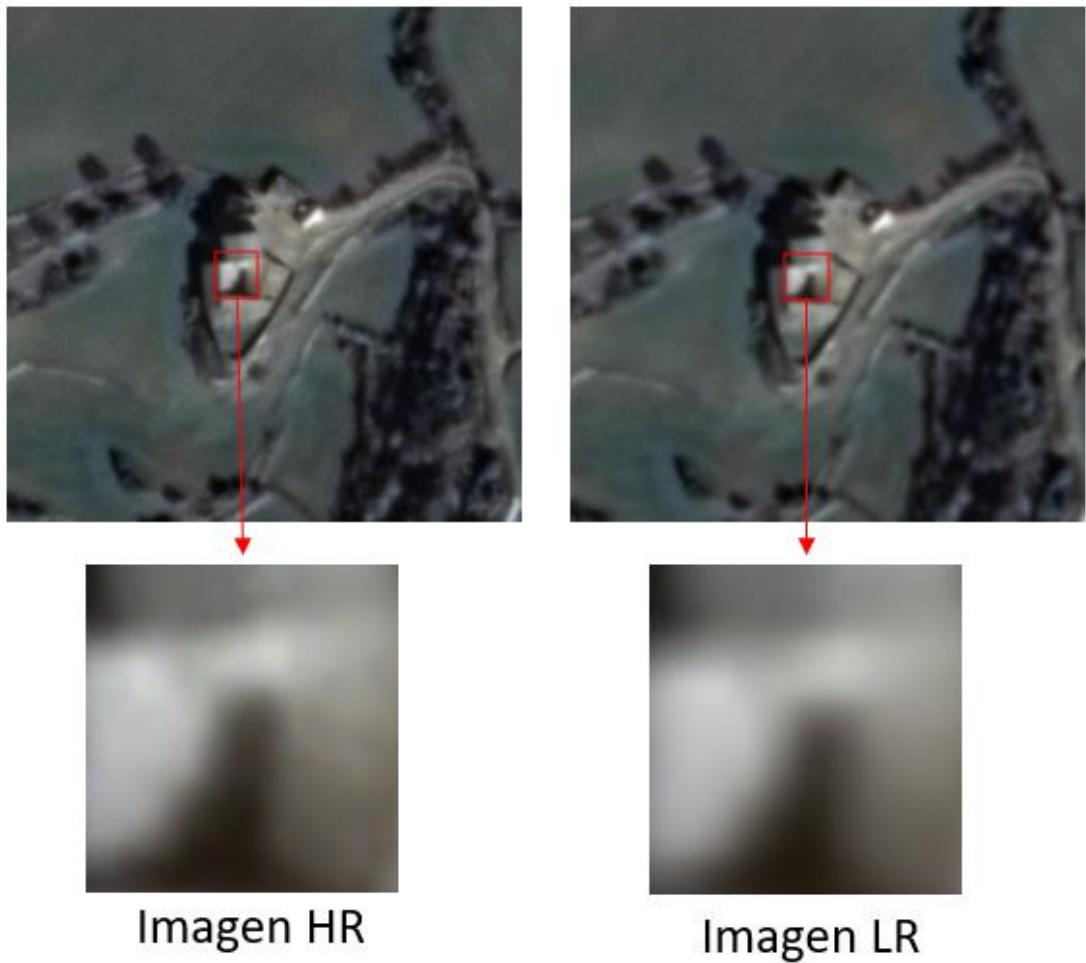


Ilustración 9. Imágenes de alta y baja resolución del GEOSAT2 con escalado x2



Ilustración 10. Imágenes reconstruidas por los modelos 1,2 y 3

Las figuras representadas muestran en primer lugar una comparación entre la imagen de alta resolución y la imagen de baja resolución, la cual será superresolucionada por los modelos. En segundo lugar, encontramos las imágenes superresolucionadas por los modelos mencionados con anterioridad. Además de las propias imágenes, también se muestran recortes de una cierta región para poder realizar comparaciones perceptuales más precisas.

Desde el punto de vista cualitativo, encontramos diferencias sutiles en la reconstrucción de detalles finos, ya que, como se ha visto, los resultados cuantitativos son bastante parejos. El modelo 1 parece que es el que más se aleja en la reconstrucción de bordes con respecto a la imagen de alta resolución. El segundo modelo parece que capta de una forma más precisa las formas y reconstruye con una menor difusión y algo más de detalle en los bordes. Por último, nos encontramos con el tercer modelo que, a la vista y corroborado por las métricas, es el que más se asemeja a la imagen original. Se visualiza como en los detalles y bordes; la reconstrucción del tercer modelo es la más semejante a la imagen de alta resolución, evidenciando que en este caso las métricas sí se enlazan con la percepción visual.

En cuanto a los resultados globales, podemos ver unos resultados muy buenos, consiguiendo en el mejor modelo una similitud con respecto a la imagen de alta resolución de más del 99%.

6.1.1.b Factor de escalado = 3

Tabla 5. Métricas de los modelos para imágenes de GEOSAT2 con factor de escalado x3

ID modelo	Tasa de aprendizaje	Recorte de gradiente	Épocas	Factor de escalado	PNSR (dB)	SSIM
4	0.1	0.1	80	3	40.43	0.9707
5	0.1	0.15	80	3	40.57	0.9711
6	0.01	0.15	200	3	41.87	0.9751

La tabla muestra los resultados de los tres mejores modelos obtenidos para el factor de escalado x3. En el cuarto modelo se utilizó una tasa de aprendizaje de 0.1 junto con un recorte de gradiente de 0.1 y este fue entrenado durante 80 épocas, ya que era la configuración que mejores resultados había dado en el entrenamiento con escalado x2.

Para el entrenamiento del quinto modelo se realizó un aumento del recorte de gradiente que en este caso se tradujo en una leve mejora del modelo. Aunque los resultados fueron buenos, se planteó la posibilidad de que en este caso la red necesitase aprender más lentamente. Por ello, en el sexto

modelo se rebajó la tasa de aprendizaje a la vez que se aumentó el número de épocas para compensar la ralentización del proceso de aprendizaje. Los resultados fueron notoriamente mejores, lo que sugirió que para factores más grandes la red necesitaba aprender con una menor velocidad.

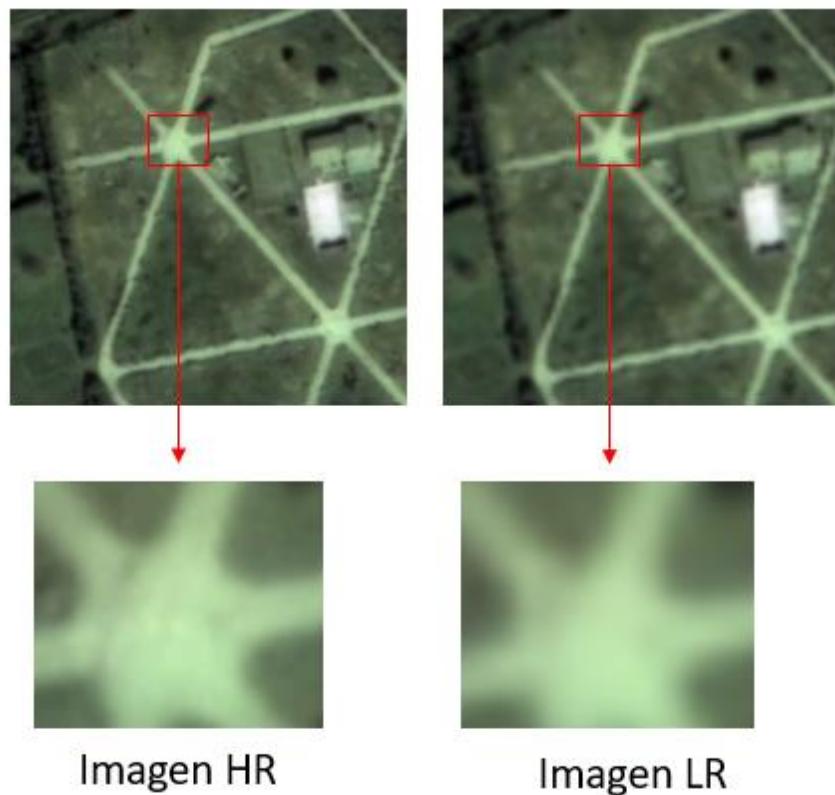


Ilustración 11. Imágenes de alta y baja resolución del GEOSAT2 con escalado x3

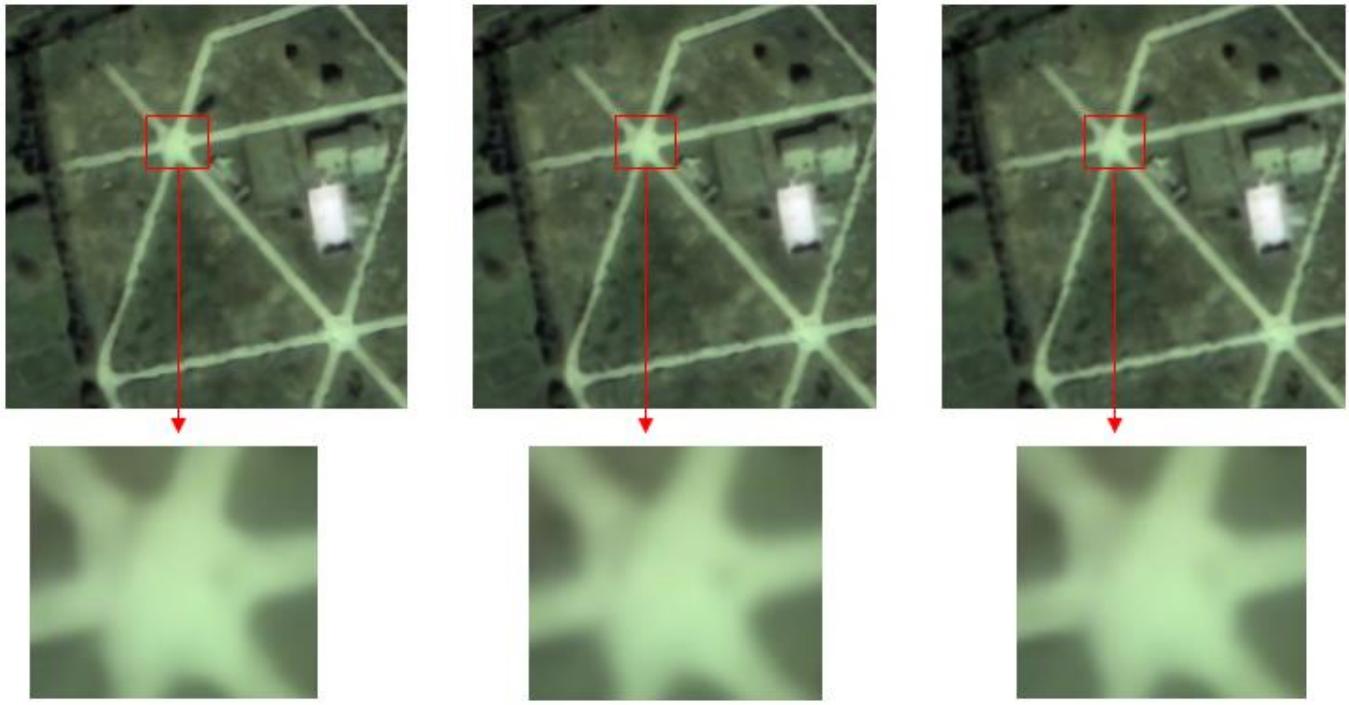


Ilustración 12. Imágenes reconstruidas por los modelos 4,5 y 6

De forma análoga a los modelos presentados para el escalado x2, se han incluido las imágenes generadas por los modelos 4, 5 y 6 incluyendo de la misma manera un recorte ampliado de la zona marcada para la realización de una buena comparación perceptual.

Desde el punto de vista cualitativo y al igual que pasaba anteriormente, se aprecian diferencias sutiles pero suficientes para decidir que, modelos están haciendo una mejor reconstrucción de la imagen original. En este caso se puede observar cómo el cuarto y quinto modelo presentan una reconstrucción muy parecida y algo más difusa que la del sexto modelo, con bordes más suavizados y una forma que parece ser menos definida que la del último modelo. Por ello y como apuntaban las métricas, se observa una mejor recuperación en el caso del sexto modelo, asemejándose algo más que las otras dos en las formas principales de la imagen original.

En cuanto al resultado global de los tres modelos, es bastante bueno, consiguiendo en los tres casos una similitud de más del 97 % con respecto a la imagen de referencia y mejorando en gran medida la imagen de baja resolución en la que ya se apreciaba una pérdida de detalles bastante elevada.

6.1.1.c Factor de escalado = 4

Tabla 6. Métricas de los modelos para imágenes de GEOSAT2 con factor de escalado x4

ID modelo	Tasa de aprendizaje	Recorte de gradiente	Épocas	Factor de escalado	PNSR (dB)	SSIM
7	0.1	0.15	80	4	36.62	0.9344
8	0.01	0.1	200	4	37.70	0.9414
9	0.01	0.15	200	4	37.72	0.9418

La tabla recoge los resultados de los tres mejores modelos para el factor de escalado x4, el más exigente desde el punto de vista de la reconstrucción. Aunque se había comprobado con el modelo de escalado x3 que las tasas de aprendizaje bajas son mejores para escalados mayores, se realizó una primera prueba con una tasa de aprendizaje de 0.1 junto con el recorte de gradiente que mejores resultados había proporcionado anteriormente.

Pese a conseguir buenos resultados, se esperaba una leve mejora con la utilización de una tasa de aprendizaje menor junto con un número de épocas mayor. Se cumplieron las predicciones, y para los modelo octavo y noveno se produjeron mejoras cuantitativas. Se hizo un leve cambio en el recorte de gradiente entre los últimos modelos para definir que recorte obtenía mejores resultados y como sucedió anteriormente, fue el de valor 0.15 el que obtuvo mejores resultados.

En términos generales los resultados fueron bastante buenos para un factor de escalado elevado como es el caso del x4, llegando a más de 94% de similitud con la imagen original en el mejor de los modelos.

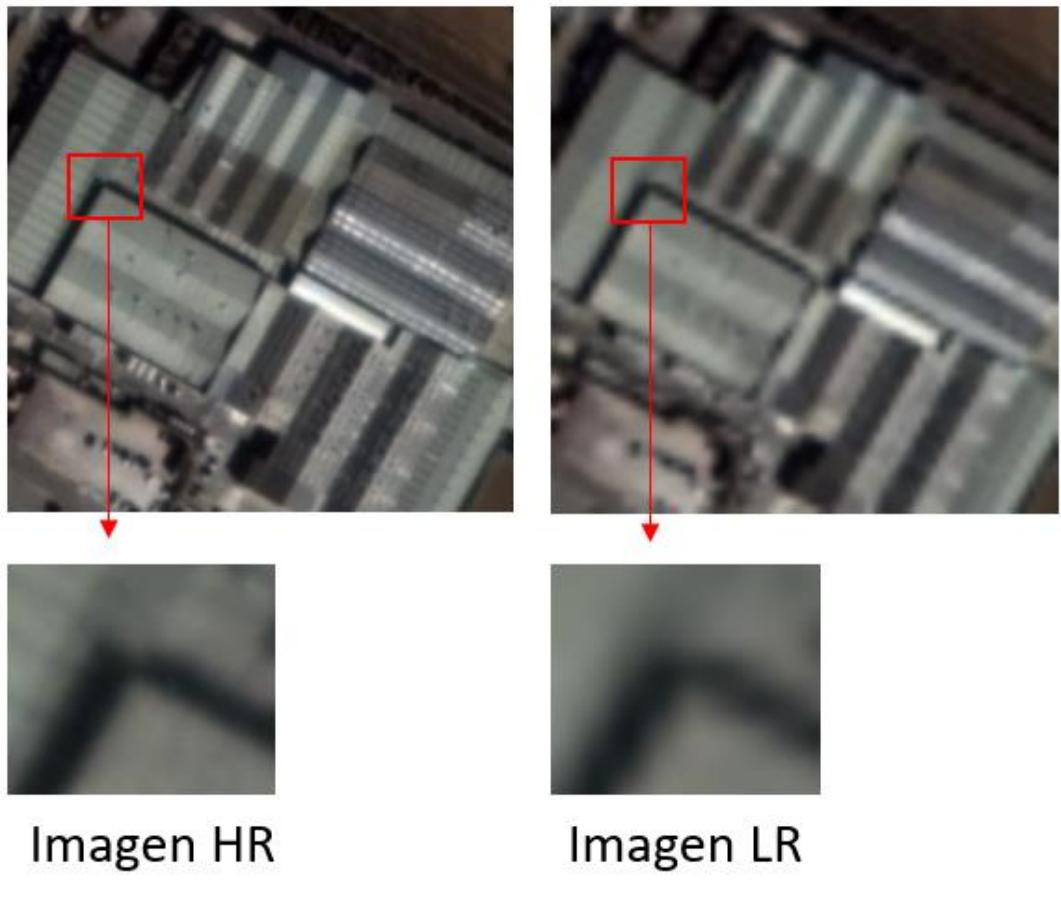


Ilustración 13. Imágenes de alta y baja resolución del GEOSAT2 con escalado x4.

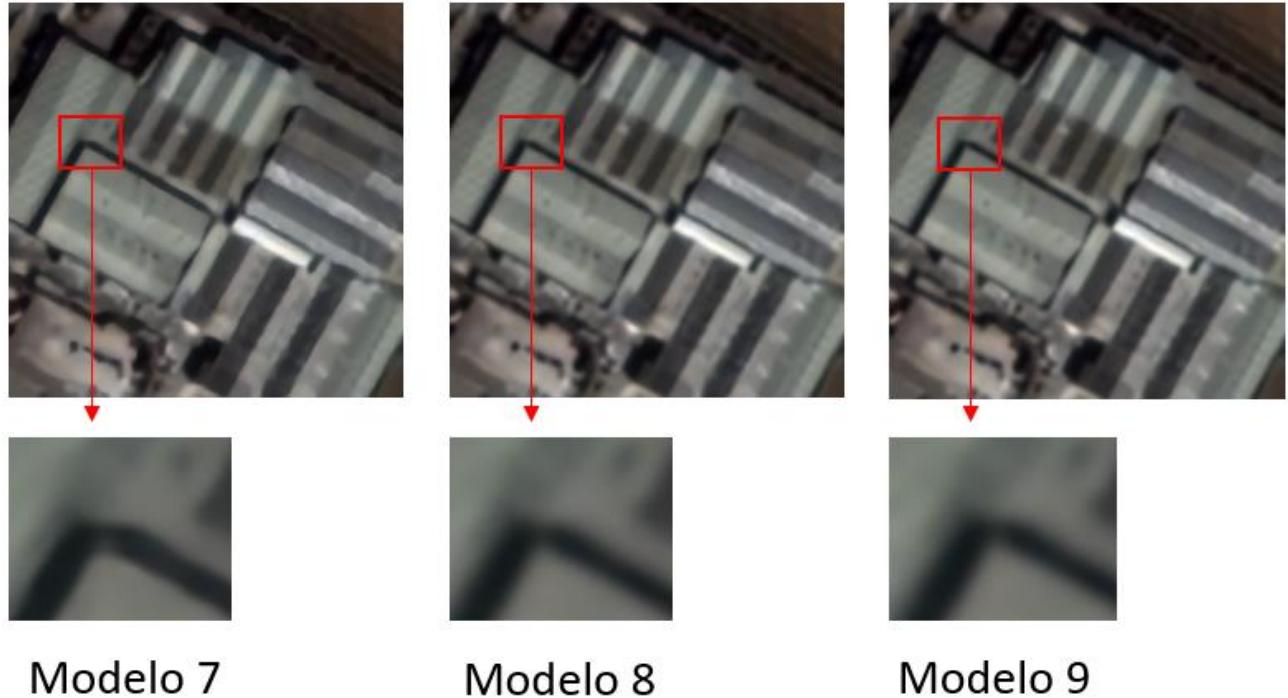


Ilustración 14. Imágenes reconstruidas por los modelos 7,8 y 9.

A nivel perceptual se observan diferencias notables entre las reconstrucciones de los modelos. El séptimo modelo destaca por una aparente mayor nitidez, ya que, como se puede observar, las líneas parecen más rectas y definidas, dando la impresión de que es una reconstrucción más precisa.

Sin embargo, esta aparente mejora visual no es realista, puesto que, si miramos las diferencias frente a la imagen original, se puede ver como el modelo está introduciendo detalles que no están presentes en la imagen de referencia, lo que indica un grado de invención estructural.

Por otra parte, tenemos los modelos octavo y noveno, los cuales ofrecen resultados más suaves, pero se ajustan de un modo más realista a la imagen original. Vemos como en estos dos últimos se ha respetado la estructura algo difusa de la imagen de referencia y entre ambas no se ven diferencias perceptuales claras, por lo que ambos modelos podrían ser escogidos para los trabajos de factor de escalado x4.

Esta evaluación perceptual ha evidenciado que en ocasiones la opinión subjetiva no es tan relevante, puesto que el modelo puede generar artefactos que demos como buenos cuando no lo son.

6.1.1.d Factor de escalado 2,3 y4

Tabla 7. Métricas de los modelos para imágenes de GEOSAT2 con factor de escalado x2, x3 y x4

ID modelo	Tasa de aprendizaje	Recorte de gradiente	Épocas	Factor de escalado	PNSR (dB)	SSIM
10	0.01	0.15	200	2	46.91	0.9909
10	0.01	0.15	200	3	41.80	0.9748
10	0.01	0.15	200	4	37.60	0.9407

Como ya vimos anteriormente, el modelo 10 fue entrenado utilizando la opción de VDSR de aprendizaje multi-escalado, la cual nos permite entrenar con múltiples factores de escalado a la vez. Para este entrenamiento se utilizó una configuración de hiper-parámetros con una tasa de aprendizaje de 0.1, un recorte de gradiente de 0.15 y un total de 200 épocas, los tres seleccionados en base a su buen rendimiento en entrenamientos anteriores.

Los resultados obtenidos muestran que el modelo ha sido capaz de adaptarse correctamente a todos los factores de escalado, obteniendo resultados muy cercanos en cada uno de ellos a los obtenidos en los mejores modelos entrenados con un solo factor.

En conjunto, estos resultados nos demuestran que el entrenamiento multi-escalado no solo reduce el coste computacional y el almacenamiento, sino que también obtiene resultados robustos en los distintos escalados sin necesidad de ajustar la configuración o los hiper-parámetros para cada escenario.

A continuación, se realizará la valoración perceptual con base en la figura mostrada a continuación. En ella podemos ver, para cada uno de los escalados, la imagen de alta resolución, la imagen de baja resolución y la reconstrucción realizada por el décimo modelo. En el caso del factor de escalado x2, las diferencias son muy sutiles; se puede visualizar como la imagen de baja resolución tiene bordes algo más difusos que en el caso de la imagen reconstruida, la cual se acerca en mayor medida a la representación de la imagen original. En el segundo caso del factor de escalado x3 ya se pueden visualizar diferencias muy notorias, donde encontramos una reconstrucción de imagen muy parecida a la imagen original, mientras que la imagen de la que partía había perdido en gran medida las formas originales. Finalmente, podemos visualizar los resultados para el factor de escalado x4, en el que sucede algo similar al segundo caso; el modelo es capaz de detallar los bordes y eliminar en gran medida la difusión de la imagen de la que partía, acercándose en gran medida a la representación original.

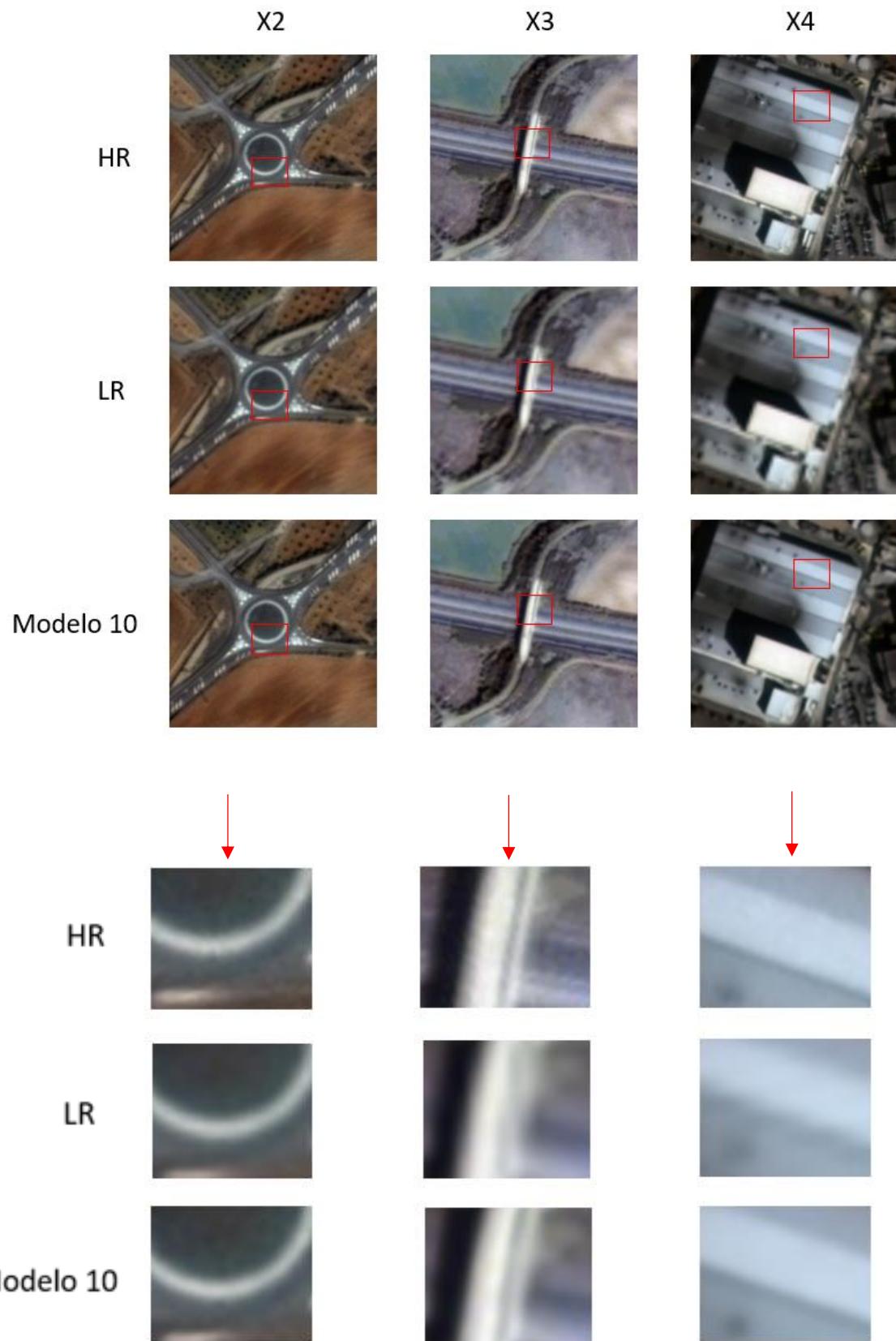


Ilustración 15. Imágenes del modelo multiescala

6.1.1.e Modelos GEOSAT2 vs métodos tradicionales

En este apartado se procederá a comparar los mejores modelos generados frente a los métodos clásicos de interpolación para cada uno de los factores de escala. Los métodos clásicos seleccionados serán el vecino más próximo, la interpolación bilineal y la interpolación bicúbica.

Tabla 8. Comparativa de métricas entre el modelo 3 e interpolaciones clásicas con factor de escalado x2

	Vecino más próximo	Bilineal	Bicúbica	Modelo 3
PNSR (dB)	34.82	40.35	42.62	47.04
SSIM	0.9486	0.9786	0.9865	0.9913

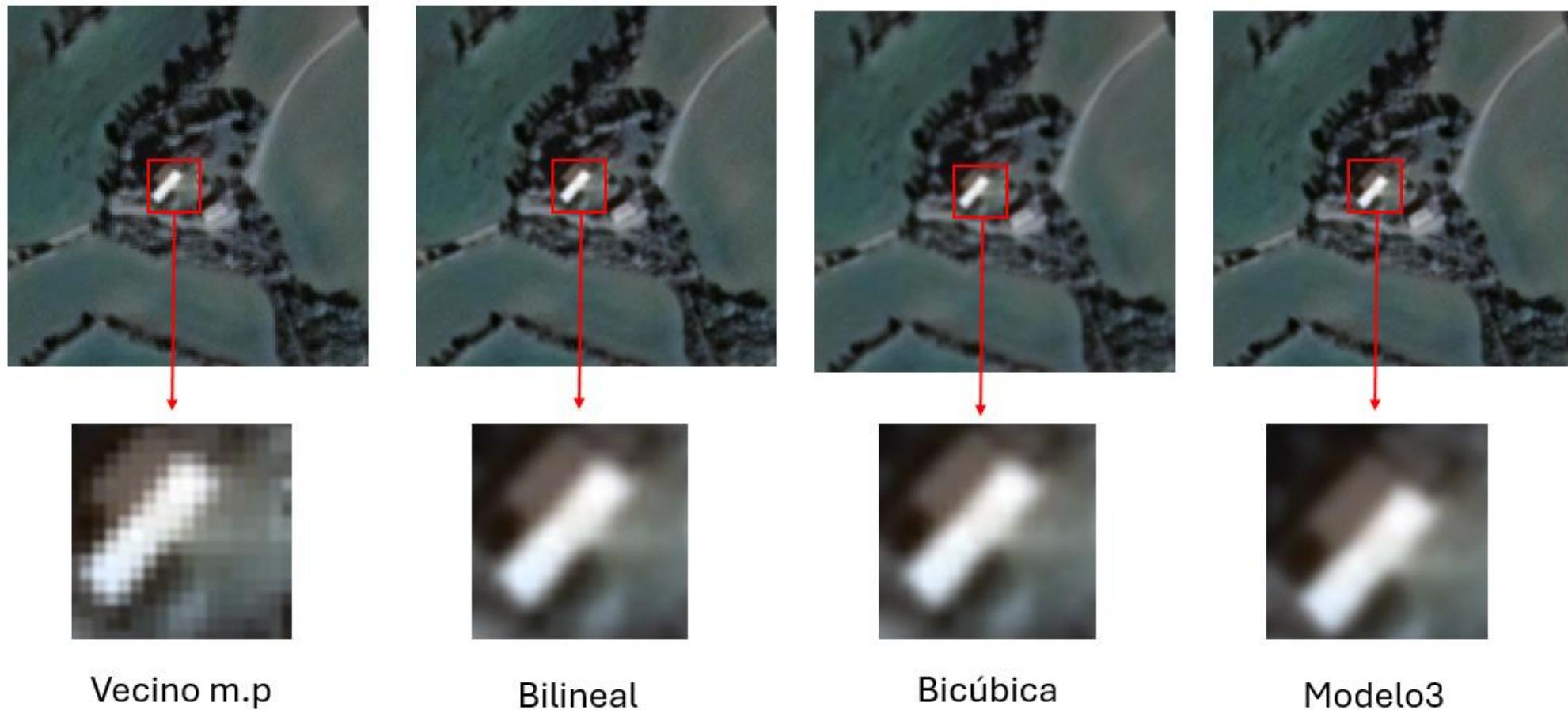
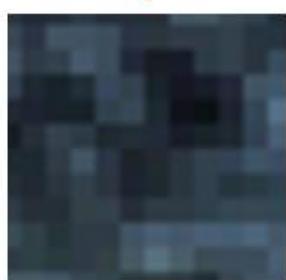


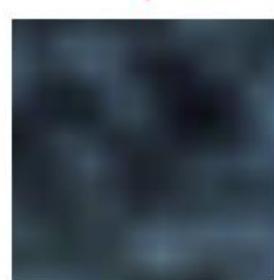
Ilustración 16. Imágenes comparativas de modelos vs métodos clásicos, del GEOSAT2 y factor de escalado x2.

Tabla 9. Comparativa de métricas entre el modelo 6 e interpolaciones clásicas con factor de escalado x3.

	Vecino más próximo	Bilineal	Bicúbica	Modelo 6
PNSR (dB)	30.97	35.58	37.68	41.87
SSIM	0.8807	0.9412	0.9577	0.9751



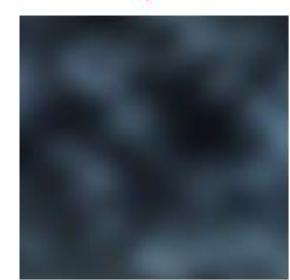
Vecino m.p



Bilineal



Bicúbica



Modelo6

Ilustración 17. Imágenes comparativas de modelos vs métodos clásicos, del GEOSAT2 y factor de escalado x3.

Tabla 10. Comparativa de métricas entre el modelo 9 e interpolaciones clásicas con factor de escalado x4

	Vecino más próximo	Bilineal	Bicúbica	Modelo 9
PNSR (dB)	28.54	32.36	33.93	37.72
SSIM	0.8109	0.8888	0.9128	0.9418

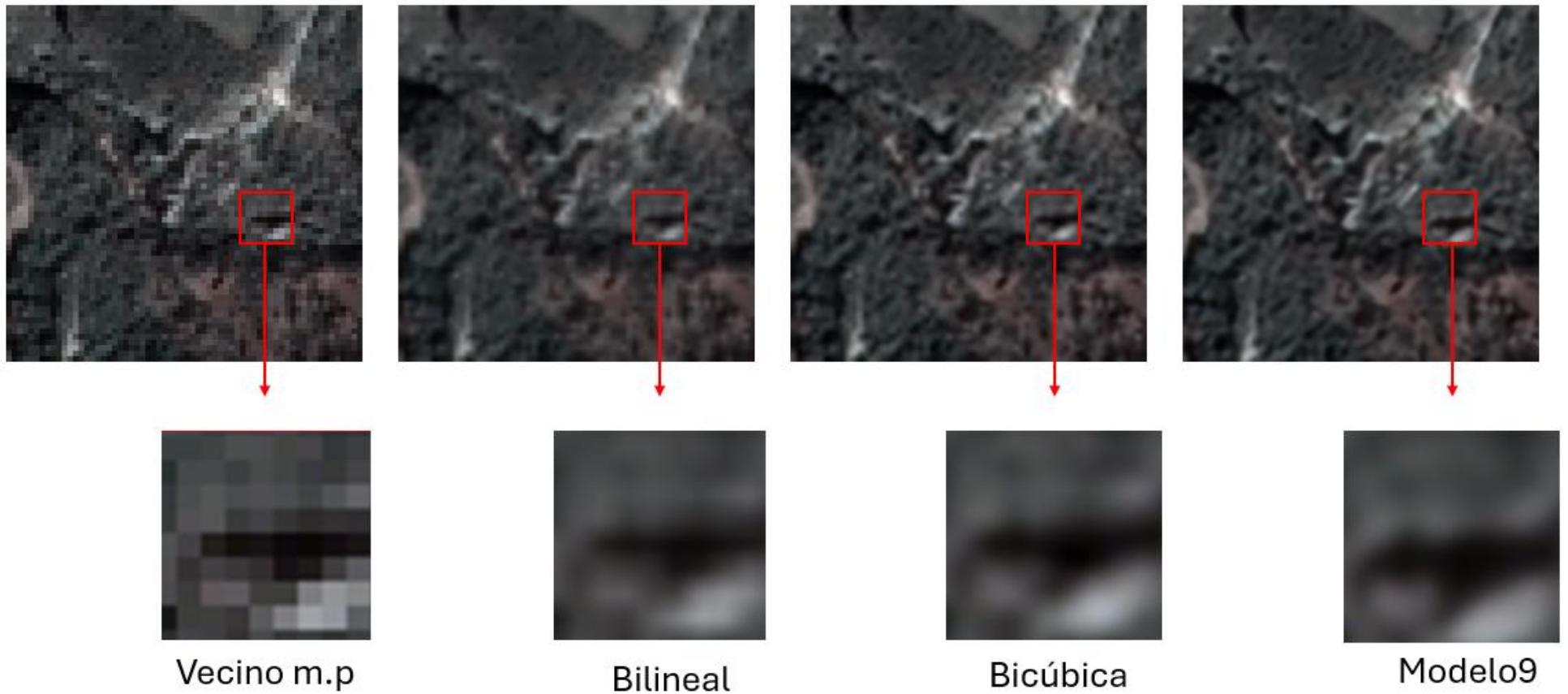


Ilustración 18. Imágenes comparativas de modelos vs métodos clásicos, del GEOSAT2 y factor de escalado x4.

Analizando las métricas obtenidas para los distintos factores de escalado, se observa una mejora progresiva en las métricas dentro de los métodos clásicos, donde la interpolación bicúbica muestra los mejores resultados entre los métodos de interpolación.

Sin embargo, podemos ver como los modelos generados superan ampliamente a todas las técnicas clásicas tanto en PSNR como SSIM. En el escalado x2 ya podemos visualizar una gran diferencia en cuanto a métricas y según agrandamos el factor de escalado estas diferencias se hacen más notorias. Este comportamiento era de esperar puesto que a mayor factor de escalado mayor complejidad de reconstrucción y un modelo entrenado podrá aplicar relaciones que un modelo de interpolación no podrá.

Desde el punto de vista perceptual también encontramos diferencias notorias. En términos generales, se puede visualizar cómo los modelos reconstruyen las imágenes de una forma más precisa, con bordes más definidos y un menor grado de difusión. Estas diferencias van incrementando conforme aumentamos el factor de escala, tal y como se preveía en la evaluación cualitativa.

Los resultados obtenidos muestran que, para imágenes satelitales como las del GEOSAT2, los modelos de superresolución basados en redes neuronales superan ampliamente a los métodos tradicionales de interpolación, tanto en términos cuantitativos como en cualitativos. Mientras que los modelos tradicionales tienen limitaciones a la hora de recuperar detalles, sobre todo en escalados altos, los modelos entrenados son capaces de reconstruir las imágenes de una forma más nítida y coherente con la imagen original.

6.1.2 Evaluación de modelos para ortoimágenes del PNOA

6.1.2.a Factor de escalado = 2

Tabla 11. Métricas de los modelos para imágenes del PNOA con factor de escalado x2

ID modelo	Tasa de aprendizaje	Recorte de gradiente	Épocas	Número de imágenes	Factor de escalado	PSNR (dB)	SSIM
11	0.01	0.15	200	500	2	27.83	0.7943
12	0.01	0.15	300	1500	2	28.94	0.8337
13	0.1	0.15	200	2500	2	29.00	0.8342
14	0.01	0.15	400	2500	2	29.43	0.8422

Estos cuatro modelos corresponden a los entrenamientos realizados con las ortoimágenes del PNOA utilizando un factor de escala de x2. Como diferencia principal a las evaluaciones de las imágenes del GEOSAT2, se ha incorporado la variable del número de imágenes, ya que, como se explicó anteriormente, se comenzó por un número de imágenes que equivalía en número de recortes a las imágenes del GEOSAT2 debido a las diferencias de tamaño, como se mencionó anteriormente. Sin embargo, los resultados no eran del todo óptimos, por lo que se decidió sacrificar tiempo de entrenamiento para obtener mejores resultados.

Para los modelos número 11 y 12 se utilizaron los hiper-parámetros con mejores resultados en entrenamientos anteriores, pero con la distinción del número de imágenes utilizadas. Estos dos primeros resultados mostraron que el aumento del *dataset* conseguía mejorar los resultados finales de forma evidente. Por ello se decidió hacer un segundo aumento y probar con un entrenamiento de aprendizaje más rápido y otro más lento con el uso de una tasa de aprendizaje de 0.1 en el primer caso y 0.01 en el segundo. Finalmente, mediante los modelos número 13 y 14, se corroboró que, para la reconstrucción óptima de imágenes de alta resolución, se necesitaban entrenamientos más lentos.

En conjunto, los resultados pueden parecer malos comparados con los obtenidos en las imágenes satelitales, pero hay que tener en cuenta que las ortoimágenes tienen 3 veces mayor resolución que las imágenes del GEOSAT2, por lo que es mucho más compleja la reconstrucción de las primeras. Pese a la mayor complejidad, se ha conseguido más de un 84% de similitud con la imagen original en el mejor modelo generado, por lo que, si tenemos en cuenta la resolución de la imagen, podemos afirmar que es un resultado más que aceptable.

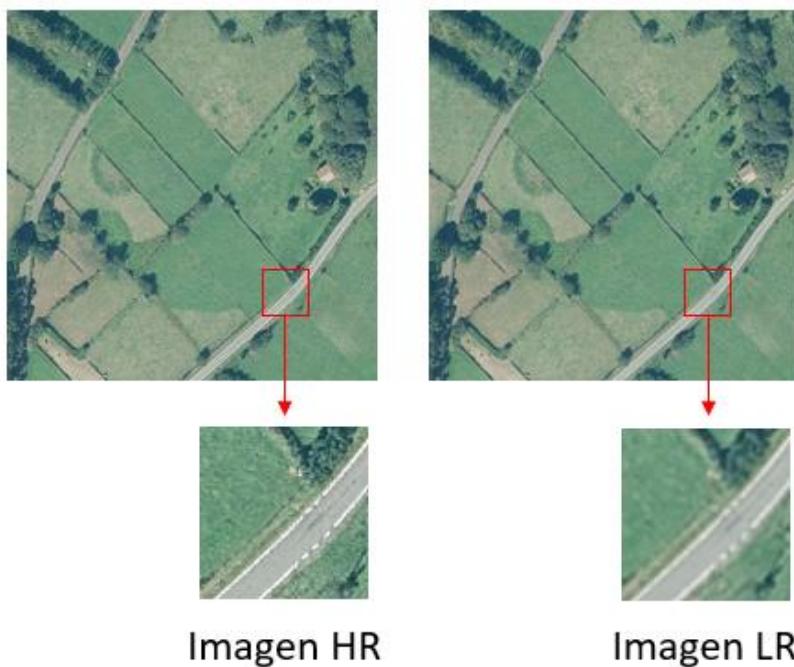


Ilustración 19. Imágenes de alta y baja resolución del PNOA con escalado x2



Ilustración 20. Imágenes reconstruidas por los modelos 11,12,13 y 14.

En cuanto a la valoración perceptual, el undécimo modelo no presenta una mejora aparente en los detalles mientras que los modelo número 12 y 13 parecen presentar una mayor definición en los bordes y texturas asemejándose en mayor medida a la imagen original.

Esta mejora frente al modelo número 11 se puede ver claramente en los recortes presentados, donde las líneas de carretera ganan algo de detalle quitando parte de la difusión que presentaba el primer modelo.

Finalmente encontramos el decimocuarto modelo que mejora en gran medida frente a los modelos anteriores asemejándose en gran medida a la representación original, obteniendo una definición en los detalles bastante similar a la imagen de alta resolución visible tanto en las líneas de carretera como en los arbustos cercanos a estas.

6.1.2.b Factor de escalado = 3

Tabla 12. Métricas de los modelos para imágenes del PNOA con factor de escalado x3

ID modelo	Tasa de aprendizaje	Recorte de gradiente	Épocas	Número de imágenes	Factor de escalado	PNSR (dB)	SSIM
15	0.01	0.15	200	500	3	26.40	0.6926
16	0.01	0.15	300	1500	3	26.48	0.6952
17	0.01	0.15	400	2500	3	26.63	0.7001

En la tabla anterior encontramos los modelos entrenados con un factor de escalado x3 para las imágenes del PNOA.

Los entrenamientos de estos tres modelos fueron realizados modificando únicamente las épocas y el número de imágenes utilizadas, dejando todos los hiper-parámetros restantes como constantes. Las épocas se incrementaron con el aumento del número de imágenes, con el fin de facilitar el aprendizaje del entrenamiento. El resto de hiper-parámetros se mantuvieron constantes y fueron seleccionados por su rendimiento en entrenamientos anteriores.

En cuanto a los resultados podemos visualizar como de forma esperada fueron mejorando conforme se aumentaba el número de imágenes de entrenamiento hasta llegar a un 70% de similitud entre la imagen re-escalada por el mejor modelo y la original.



Imagen HR

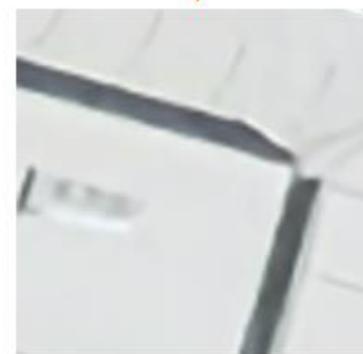


Imagen LR

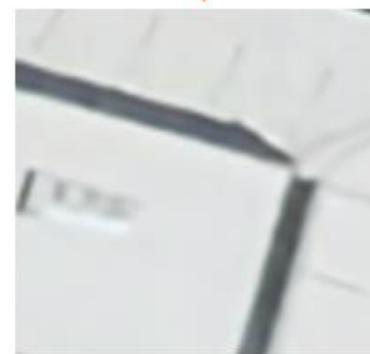
Ilustración 21. Imágenes de alta y baja resolución del PNOA con escalado x3.



Modelo 15



Modelo 16



Modelo 17

Ilustración 22. Imágenes reconstruidas por los modelos 15,16 y 17.

La comparación visual entre los resultados generados por los modelos número 15,16 y 17 muestra diferencias perceptuales aparentes.

El modelo decimoquinto presenta una reconstrucción con texturas más suaves y contornos poco definidos quedándose entre medias de la imagen original y la degradada.

El modelo decimosexto presenta una representación más cercana a la original con bordes más marcados y en consecuencia menos difusos. Vemos como los cables del exterior de la planta del edificio obtienen una mayor definición.

El modelo decimoséptimo presenta unos resultados muy similares a nivel perceptual con respecto al decimosexto modelo, haciendo complicada la visualización de diferencias aparentes en los detalles más finos. Estos dos últimos modelos consiguen una reconstrucción bastante parecida a la original consiguiendo mejorar en gran medida la difusión existente en la imagen degradada.

6.1.2.c Factor de escalado = 4

Tabla 13. Métricas de los modelos para imágenes del PNOA con factor de escalado x4

ID modelo	Tasa de aprendizaje	Recorte de gradiente	Épocas	Número de imágenes	Factor de escalado	PNSR (dB)	SSIM
18	0.1	0.15	80	500	4	24.56	0.5140
19	0.01	0.15	200	500	4	24.58	0.5148
20	0.01	0.15	300	1500	4	24.70	0.5755
21	0.01	0.15	400	2500	4	25.20	0.5841

En la tabla anterior encontramos los modelos del número 18 al 21, los cuales fueron entrenados con un factor de escalado x4. En este caso se realizaron dos entrenamientos con un número de imágenes bajo, con el fin de corroborar que al igual que en casos anteriores la tasa de aprendizaje baja resultaba en un modelo con mejores resultados.

Como se esperaba, la tasa de aprendizaje menor junto con un número mayor de épocas dio mejores resultados, por lo que se utilizaron estos hiper-parámetros para los entrenamientos siguientes. En el vigésimo modelo se realizó un aumento en el número de imágenes, resultando en una mejora sustancial de los resultados.

Finalmente se realizó un último entrenamiento diferenciado por un nuevo aumento del número de imágenes utilizadas. Este último entrenamiento obtuvo los mejores resultados cuantitativos con una diferencia sustancial frente al modelo anterior, evidenciando y corroborando que el tamaño del dataset es crucial para la obtención de buenos resultados en un modelo.

Los resultados cuantitativos parecen no ser suficientemente buenos para su utilización, pero al tener en cuenta de que el factor de escalado es tan elevado se deberán realizar las valoraciones perceptuales antes de decidir si los resultados son suficientemente buenos para el uso de los modelos.

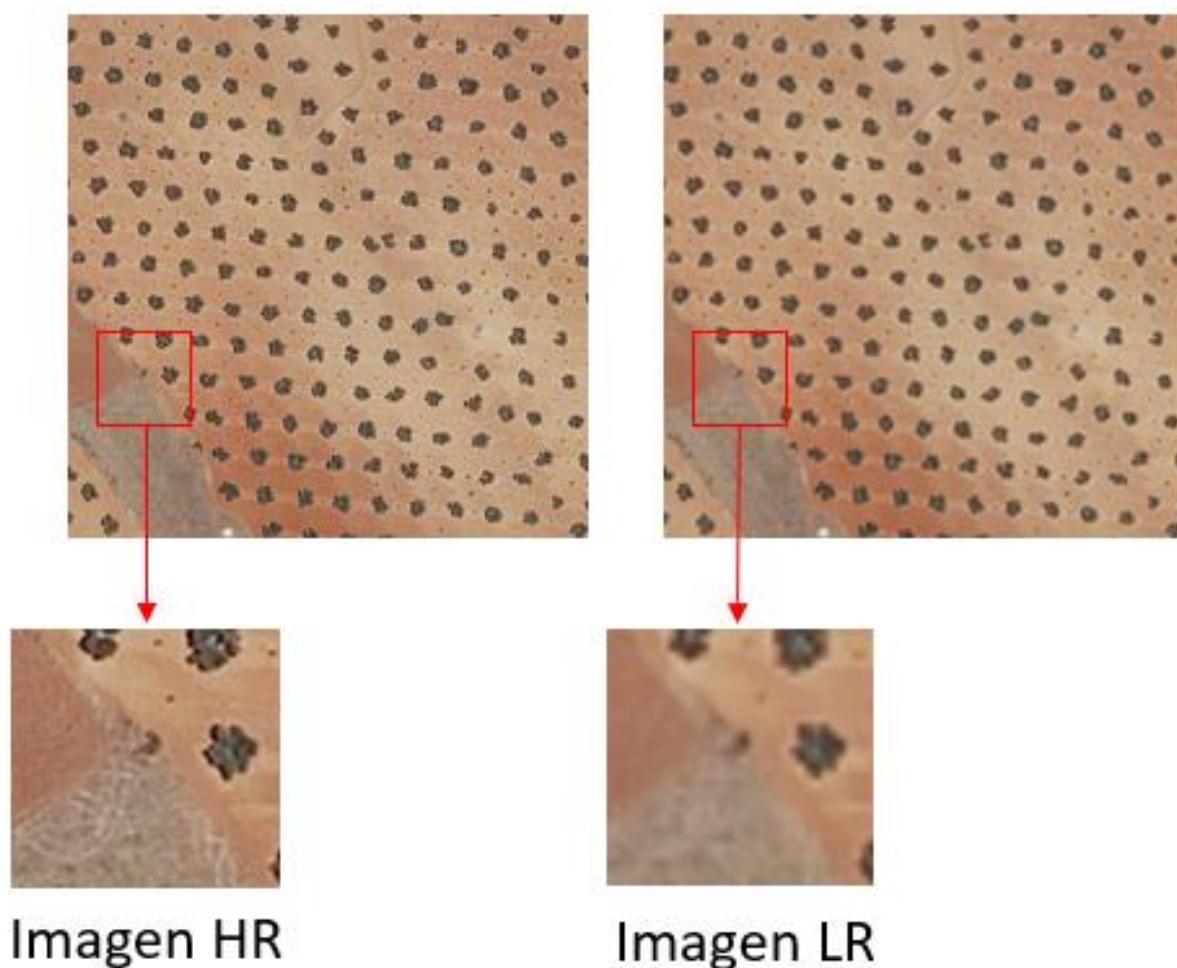
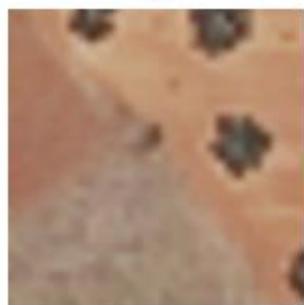
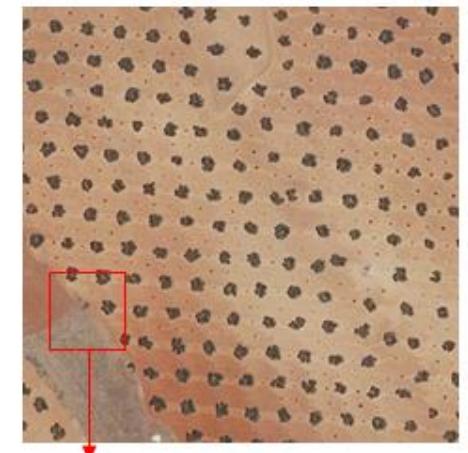
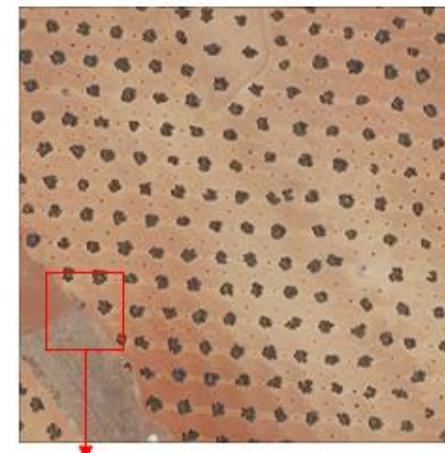
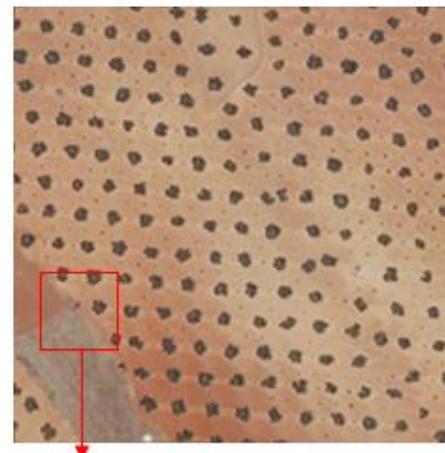
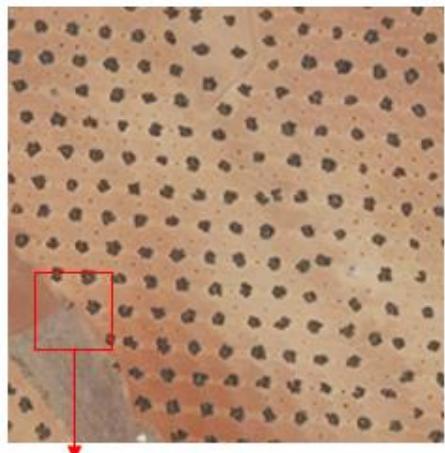
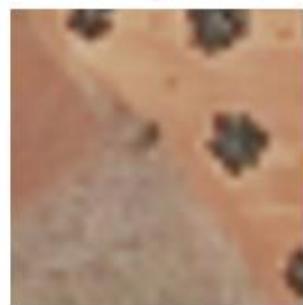


Ilustración 23. Imágenes de alta y baja resolución del PNOA con escalado x2.



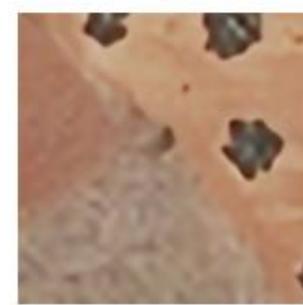
Modelo 18



Modelo 19



Modelo 20



Modelo 21

Ilustración 24. Imágenes reconstruidas por los modelos 18,19,20,21.

La comparación visual entre los resultados generados por los modelos refleja la dificultad de recuperar detalles finos en escalados altos como el utilizado. Sin embargo, en esta comparativa perceptual se pueden visualizar diferencias sustanciales entre los modelos.

El modelo número 18 muestra una imagen con bordes y detalles bastante difusos, muy parecidos a los de la imagen de la que partía, evidenciando que la mejora ha sido muy sutil en la reconstrucción de estos detalles. En cuanto al modelo número 19, nos encontramos con una situación muy parecida a la anterior, con una reconstrucción pobre y sin una mejora de detalle aparente, lo que nos podría hacer descartar estos dos últimos modelos para las tareas de superresolución.

Pese a estos dos primeros modelos, el modelo vigésimo tiene una mejora sustancial frente a los últimos, recuperando de una forma mucho más satisfactoria los bordes y detalles, reduciendo en gran medida las zonas difusas de la imagen degradada y consiguiendo así acercarse a la representación original.

Finalmente encontramos el modelo vigesimoprimer con una representación visual aún mejor que el modelo anterior, acercándose de forma sustancial a la imagen de alta resolución. Pese a que los resultados cuantitativos no eran buenos, la percepción visual indica que este último modelo si tienen utilidad en tareas de superresolución, mostrando una reconstrucción de la imagen que podría ser muy útil para ciertos usos en los que se necesite una ganancia de resolución bastante elevada.

Los resultados perceptuales han evidenciado que los resultados cuantitativos no siempre son determinantes o suficientes a la hora de valorar un modelo, necesitando la valoración perceptual para deliberar si un modelo es útil o, por el contrario, se debe descartar.

6.1.2.d Modelos PNOA vs métodos tradicionales

Tabla 14. Comparativa de métricas entre el modelo 14 e interpolaciones clásicas con factor de escalado x2

	Vecino más próximo	Bilineal	Bicúbica	Modelo 14
PNSR (dB)	25.25	26.19	26.61	29.43
SSIM	0.7683	0.7487	0.7935	0.8422



Vecino m.p

Bilineal

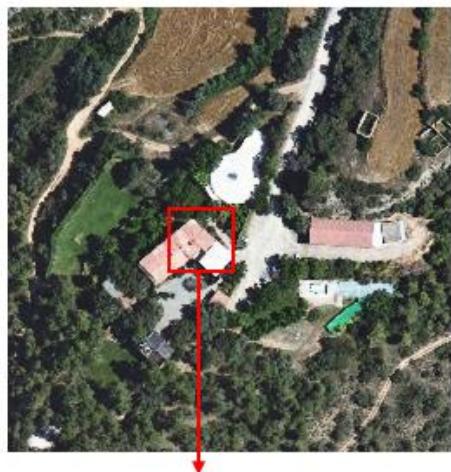
Bicúbica

Modelo 14

Ilustración 25. Imágenes comparativas de modelos vs métodos clásicos, del PNOA con factor de escalado x2.

Tabla 15. Comparativa de métricas entre el modelo 17 e interpolaciones clásicas con factor de escalado x3

	Vecino más próximo	Bilineal	Bicúbica	Modelo 17
PNSR (dB)	21.75	23.24	22.76	26.63
SSIM	0.5361	0.5730	0.5692	0.7001



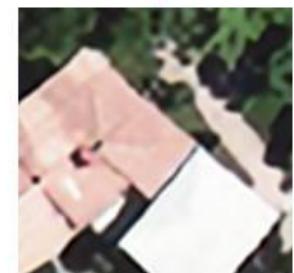
Vecino m.p



Bilineal



Bicúbica



Modelo 17

Ilustración 26. Imágenes comparativas de modelos vs métodos clásicos, del PNOA con factor de escalado x3.

Tabla 16. Comparativa de métricas entre el modelo 21 e interpolaciones clásicas con factor de escalado x4

	Vecino más próximo	Bilineal	Bicúbica	Modelo 21
PNSR (dB)	20.91	22.34	21.83	25.20
SSIM	0.4454	0.4711	0.4766	0.5841



Vecino m.p



Bilineal



Bicúbica



Modelo 21

Ilustración 27. Imágenes comparativas de modelos vs métodos clásicos, del PNOA con factor de escalado x4.

De una forma muy similar a lo que sucedía con las imágenes satelitales del GEOSAT2, encontramos que el rendimiento de los modelos entrenados es muy superior a los métodos tradicionales, con diferencias aún más significativas en las métricas SSIM, evidenciando que las imágenes reconstruidas por los modelos mantienen una similitud con imagen original mucho mayor que las imágenes reconstruidas con métodos de interpolación clásicos. Estas diferencias varían desde el 5% hasta el 14% en el mejor caso, evidenciando que para imágenes de resoluciones mayores como las ortoimágenes del PNOA se hace imprescindible el uso de estos modelos para trabajos de superresolución.

Desde el punto de vista perceptual, encontramos diferencias aún más notorias que en el caso de las imágenes del GEOSAT2. Se pueden visualizar las grandes diferencias en la reconstrucción de bordes y detalles, donde los modelos entrenados permiten un entendimiento de la imagen mucho mayor que las reconstrucciones por métodos clásicos, eliminando en gran medida las partes difusas que las interpolaciones clásicas no son capaces de minimizar.

Estos resultados muestran la necesidad de utilizar modelos de superresolución para tareas de este ámbito con imágenes de alta resolución como las del PNOA, ya que como se ha podido visualizar, los métodos clásicos de interpolación obtienen un peor rendimiento frente a estos modelos, incrementándose la diferencia de rendimiento conforme se aumenta la resolución y el factor de escalado.

6.2. Resultados del Objetivo_2

En este apartado se van a presentar los resultados correspondientes al segundo objetivo del proyecto. Como ya se ha mencionado en la preparación del *dataset*, este objetivo se ha llevado a cabo mediante dos estrategias y por lo tanto estas dos se evaluarán por separado.

En primer lugar, se evaluarán los modelos entrenados con recortes en los que aquellos del PNOA tuvieron que ser redimensionados al tamaño de los recortes del GEOSAT2 y en segundo lugar se evaluarán los modelos en los que los recortes del GEOSAT2 fueron redimensionados al tamaño de los del PNOA. En ambos casos se utilizaron los mismos métodos de interpolación para realizar el redimensionamiento, por lo que se podrá comparar entre las dos estrategias.

6.2.1 GEOSAT2 redimensionado a PNOA

Tabla 17. Métricas de los modelos entrenados con imágenes de distinta fuente (GEOSAT2 redimensionado a PNOA)

ID modelo	Tasa de aprendizaje	Recorte de gradiente	Épocas	Número de imágenes	Método de interpolación	PNSR (dB)	SSIM
1	0.01	0.15	200	500	bicúbica	15.54	0.2108
2	0.01	0.15	200	500	Vecino más cercano	13.73	0.2125
3	0.01	0.15	200	500	box	20.04	0.2459

La tabla presenta los resultados de los modelos entrenados bajo la estrategia basada en parches de 42 pixeles de imágenes del PNOA y 14 pixeles para las del GEOSAT2, las cuales se han re-escalado a 42 pixeles mediante los métodos de interpolación mostrados en las tablas.

Los hiper-parámetros han sido constantes en los tres modelos para poder realizar comparaciones objetivas y el número de imágenes utilizadas ha sido de 500 por el gran aumento de parches debido al tamaño reducido de estos.

Viendo los resultados obtenidos, las métricas resultantes son considerablemente bajas en todos los casos. Los valores de PSNR que oscilan entre 13.73 y 20.04dB junto con los valores de SSIM que no superan los 0.25 indican que las imágenes generadas por el modelo presentan una fidelidad frente a las imágenes de alta resolución demasiado limitada. Aunque el método box proporciona las métricas más altas, no parece ser suficiente.

Estos resultados sugieren que los modelos no son efectivos en términos cuantitativos, sin embargo, será necesaria la realización de la evaluación perceptual, ya que podrían existir mejoras visuales que no se reflejan en las métricas.

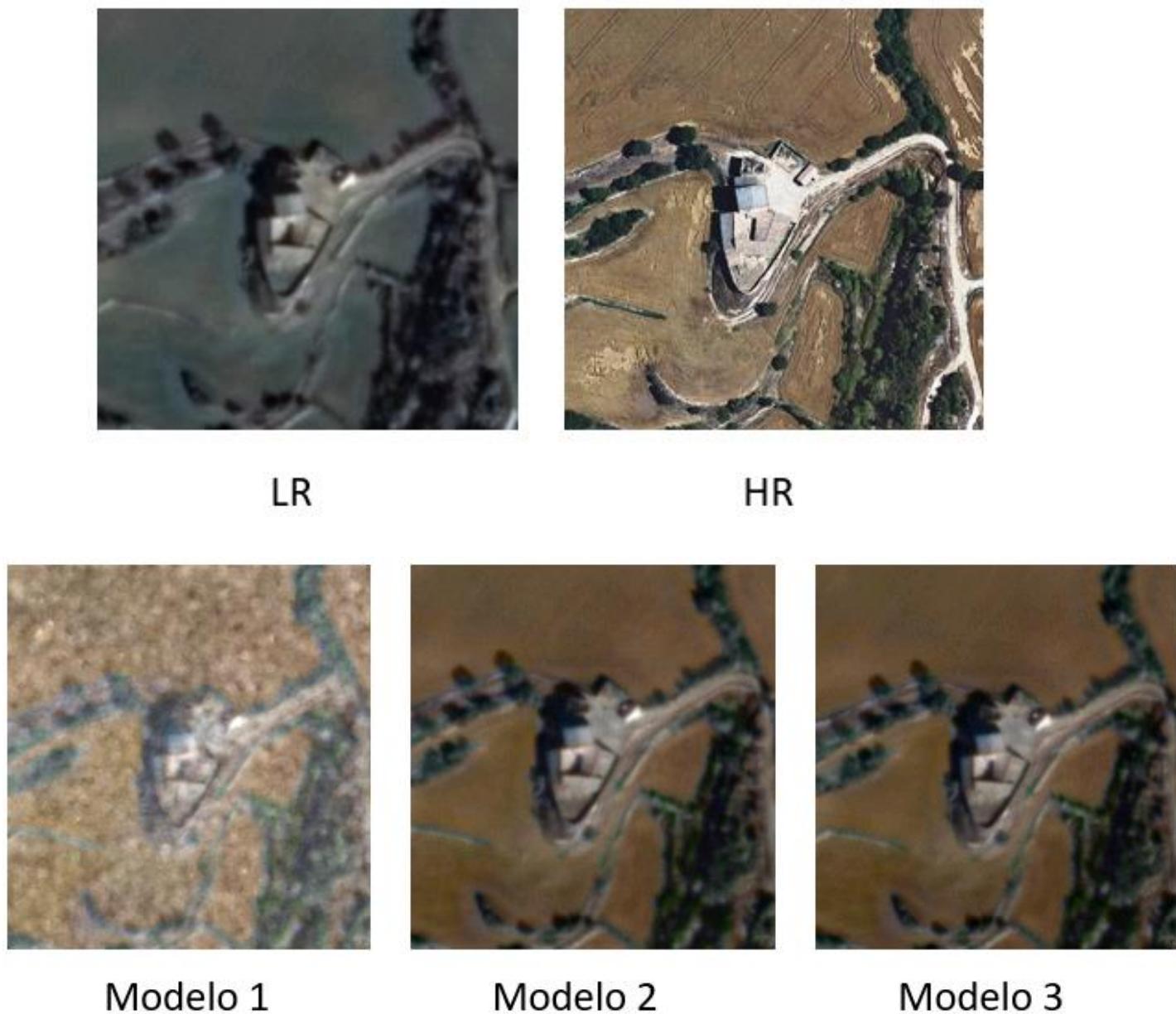


Ilustración 28. Imágenes comparativas entre los modelos 1, 2 y 3 de entrenamiento con fuentes dispares.

Desde el punto de vista perceptual, los resultados obtenidos por los modelos 1, 2 y 3 muestran comportamientos diferenciados. El modelo número uno basado en la interpolación bicúbica no es capaz de generar una imagen coherente ni visualmente útil, sugiriendo que el modelo no ha alcanzado una convergencia adecuada durante el entrenamiento.

La imagen generada presenta una textura con abundancia de artefactos y difusa. En contraste con los modelos número dos y tres, a pesar de no obtener una mejora visible en la resolución espacial, sí aportan información o valor en términos de cromatismo.

Las imágenes generadas presentan una coloración y equilibrio tonal bastante próximos a la imagen de alta resolución, lo que facilita la interpretación visual y resulta en una mejora de la percepción de la imagen. Esto es relevante, puesto que sabemos que las ortoimágenes del PNOA han sido capturadas en condiciones óptimas, ofreciendo una representación más precisa que las satelitales.

Por lo tanto, aunque no se observe una mejora efectiva de la resolución, estos modelos podrían ser útiles como herramienta de ajuste radiométrico o de mejora visual para aquellas imágenes satelitales cuya radiometría no se ajusta al terreno.

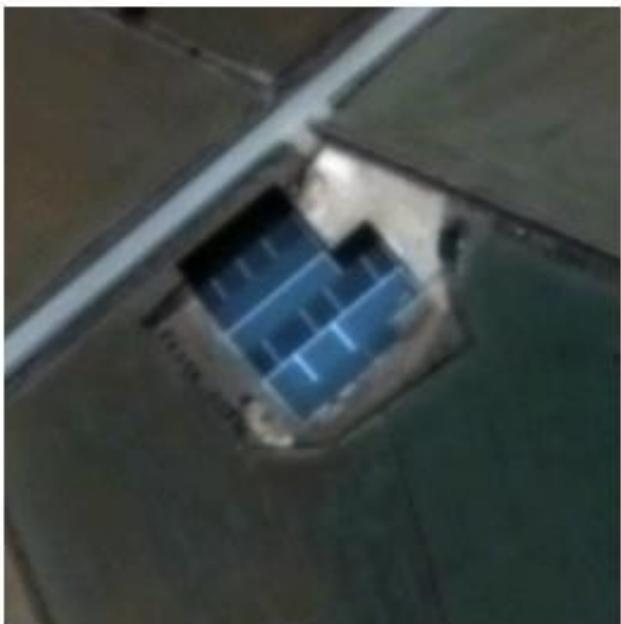
6.2.1 PNOA redimensionado a GEOSAT2

Tabla 18. Métricas de los modelos entrenados con imágenes de distinta fuente (PNOA redimensionado a GEOSAT2)

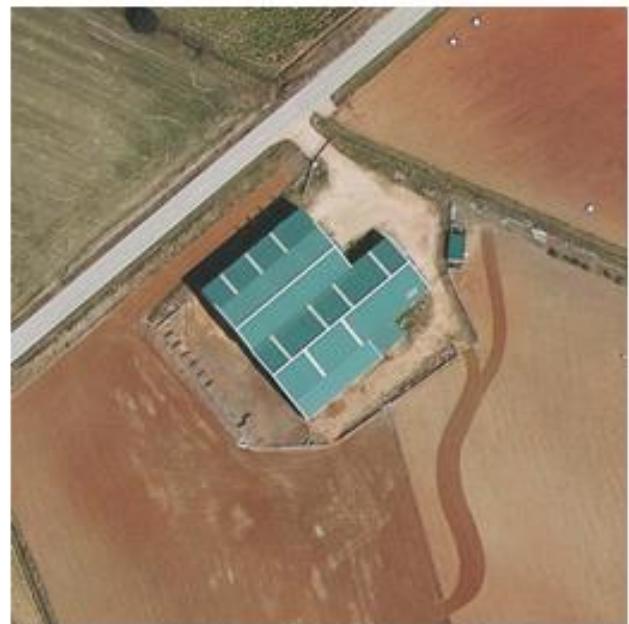
ID modelo	Tasa de aprendizaje	Recorte de gradiente	Épocas	Número de imágenes	Método de interpolación	PNSR (dB)	SSIM
4	0.01	0.15	200	4500	bicúbica	15.52	0.2553
5	0.01	0.15	200	4500	Vecino más cercano	15.18	0.2606
6	0.01	0.15	200	4500	box	18.98	0.2768

La tabla presenta los resultados de los modelos entrenados bajo la segunda estrategia e inversa de la primera, basada en parches de 42 píxeles de imágenes del GEOSAT2 y 128 píxeles para las del PNOA, las cuales se han re-escalado a 42 píxeles mediante los métodos de interpolación mostrados en las tablas. En este caso se pudo utilizar la totalidad de las imágenes, puesto que el tamaño de los parches lo permitía.

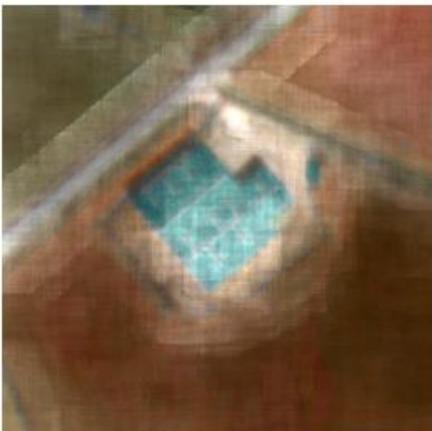
A nivel cuantitativo obtenemos resultados muy parecidos a los obtenidos con la primera estrategia. De nuevo los valores indican una baja similitud estructural con las imágenes de alta resolución y, como era de esperar, el método box vuelve a ser el que mejores resultados obtiene, aunque insuficientes en términos cuantitativos.



LR



HR



Modelo 4



Modelo 5



Modelo 6

Ilustración 29. Imágenes comparativas entre los modelos 4, 5 y 6 de entrenamiento con fuentes dispares.

En cuanto a la evaluación perceptual, sucede algo parecido a lo sucedido en la primera estrategia. Los modelos 4 y 5 muestran unas reconstrucciones malas con mayor degradación que la propia imagen de baja resolución, indicando que el modelo no ha sido capaz de aprender las relaciones entre las imágenes de alta y baja resolución.

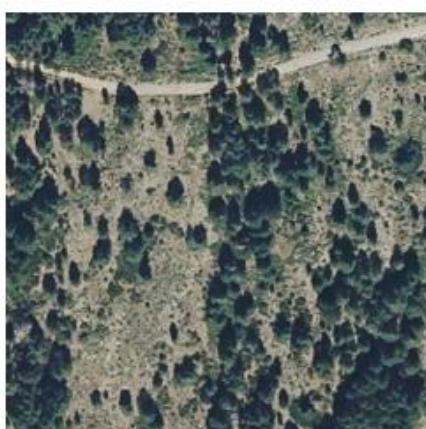
En segundo lugar, tenemos el modelo número 6 con el que sucede algo parecido que en los modelos 2 y 3, ya que, aunque no presente un aumento aparente en el nivel de detalle, si logra una reconstrucción coherente cromáticamente hablando, acercándose en gran medida al cromatismo de la imagen de alta resolución y permitiendo un mayor entendimiento de la imagen.

En este sentido, aunque el modelo no consiga superresolucionar la imagen si puede ser útil para conseguir mejoras visuales sobre imágenes satelitales.

A continuación, se mostrarán más ejemplos en los que la mejora cromática puede ayudar al entendimiento de la imagen satelital.



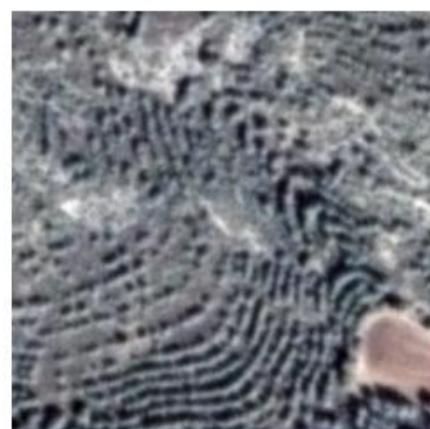
LR



HR



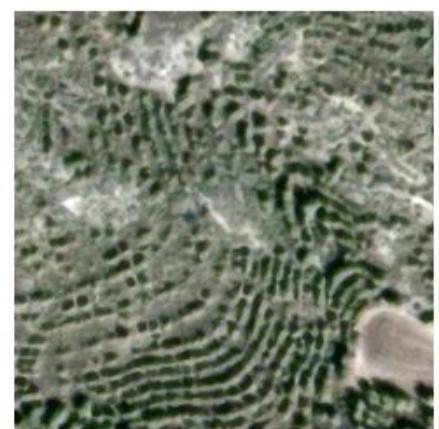
Modelo 6



LR



HR



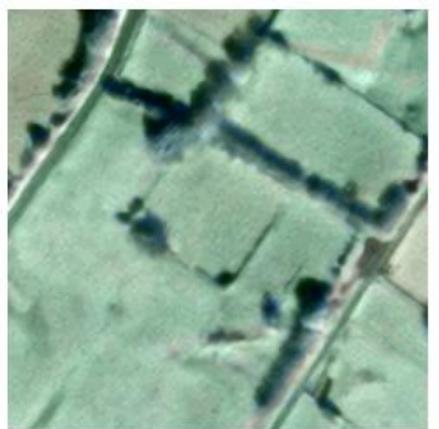
Modelo 6



LR



HR



Modelo 6

Ilustración 30. Evaluación perceptual del modelo número 6

7. CONCLUSIONES

El desarrollo de este TFG ha permitido demostrar la viabilidad y efectividad del uso de redes neuronales profundas para la tarea de la superresolución aplicada a imágenes satelitales y ortoimágenes a partir de la arquitectura VDSR. El proyecto ha abordado tanto la preparación del conjunto de datos, la modificación del código base, el entrenamiento, el análisis y la discusión de los resultados alcanzados por los modelos generados.

El hallazgo principal referente al primer objetivo ha sido la clara superioridad de esta arquitectura frente a los métodos clásicos de interpolación. Mientras que los métodos como la interpolación bilineal o bicúbica generan resultados suavizados en los que se pierden detalles estructurales, VDSR logra reconstruir detalles complejos y texturas de una forma más precisa, gracias al aprendizaje no lineal y su arquitectura profunda. Estos resultados se han reflejado tanto en las métricas cuantitativas (PNSR y SSIM) como en las evaluaciones perceptuales de las imágenes generadas.

En concreto, para la reconstrucción de imágenes satelitales, los modelos número 3, 6 y 9 han resultado ser los más propicios para los escalados de $\times 2$, $\times 3$ y $\times 4$ respectivamente, obteniendo reconstrucciones que presentaban diferencias de hasta un 3% frente al mejor método clásico, y alcanzando hasta un 13% de mejora con respecto a las reconstrucciones obtenidas por métodos clásicos más pobres.

De forma similar, para la reconstrucción de ortoimágenes, los modelos número 14, 17 y 21 han sido los que mejor rendimiento han obtenido para los factores de escalado $\times 2$, $\times 3$ y $\times 4$, mostrando diferencias de hasta un 13,1% frente al mejor método clásico y hasta un 17% de mejora respecto a los métodos tradicionales con reconstrucciones más pobres.

En cuanto al segundo objetivo del trabajo, enfocado en pares de imágenes de distinta fuente (GEOSAT2-PNOA), los resultados métricos no han sido los esperados. Sin embargo, se ha comprobado que ciertos modelos como el número 3 o el 6 han sido capaces de mejorar la coherencia cromática de las imágenes de baja resolución, lo que provoca que estos modelos sí tengan utilidad en contextos reales, ya que aunque no haya una mejora cuantitativa de resolución, su uso permite una mejor comprensión de la imagen, que es el objetivo buscado en las tareas de superresolución.

En conclusión, este trabajo demuestra que la superresolución basada en redes neuronales profundas no solo supera las limitaciones de los métodos clásicos, sino que es una metodología indispensable si nuestro objetivo es mejorar la resolución de imágenes para disciplinas como la teledetección.

De cara a futuras líneas de trabajo, se proponen dos formas por las que se podría mejorar el presente proyecto: En primer lugar y en relación con el primer objetivo, se propone el uso de alguna arquitectura más compleja para intentar obtener con ello una mejora de rendimiento en la reconstrucción de imágenes, teniendo en cuenta que esta posible mejora requerirá un consumo de recursos considerable, como se ha visto en el apartado del estado del arte.

En relación con el segundo objetivo, se propone el uso de pares de imágenes nativas de distinta resolución, pero con mayor similitud estructural, como podría ser el caso de diferentes sensores de un mismo satélite o de satélites distintos, pero con características ópticas similares. Esta aproximación permitiría seguir trabajando con imágenes reales de baja resolución, pero reduciendo la complejidad del aprendizaje al presentar una estructura similar. En consecuencia, los modelos podrían generalizar mejor y obtener reconstrucciones de mayor calidad perceptual y métrica.

8. PRESUPUESTO

En este capítulo se presentará una estimación económica del proyecto realizado. El presupuesto en cuestión ha sido dividido en cuatro apartados principales, en los cuales encontramos los costes del hardware, los costes del software, los costes de adquisición de datos y los costes de trabajo o personal. A continuación, se harán estimaciones de costes de cada uno de los apartados y finalmente se proporcionará un coste total estimado del proyecto.

8.1 Costes del hardware

El apartado de hardware se compone de un ordenador propio proporcionado por la cátedra del IGN y una máquina remota proporcionada por el tutor de este proyecto, localizada en la escuela de topografía de la Universidad Politécnica de Madrid. En cuanto al coste de la máquina remota, se ha realizado una estimación del coste en base al hardware utilizado durante la realización del proyecto.

Tabla 19. Costes de proyecto en Hardware.

Material	Coste bruto (€)	Cantidad	Tiempo de uso(meses)	Vida útil (meses)	Factor de amortización	Coste total (€)
LENOVO ThinkBook 16 G7 ARP	899	1	5	120	5/120	37,45
Maquina remota	2500	1	5	48	5/48	260,41

8.2 Costes de software

A continuación, se desglosará el conjunto de softwares utilizados y sus respectivos precios.

Tabla 20. Costes del proyecto en Software

Material	Coste bruto (€)	Cantidad	Tiempo de uso(meses)	Vida útil (meses)	Factor de amortización	Coste total (€)
Windows 11	141,99	1	5	48	5/48	14,79
Microsoft Office	70,00	1	5	12	5/12	29,16
Python	0,00	1	4	-	-	0,00
PyTorch	0,00	1	3	-	-	0,00
Anaconda	0,00	1	3	-	-	0,00
CUDA(nvidia)	0,00	1	3	-	-	0,00
GDAL	0,00	1	3	-	-	0,00
QGIS	0,00	1	2	-	-	0,00
Zotero	0,00	1	1	-	-	0,00
NotePad++	0,00	1	4	-	-	0,00
Filezilla	0,00	1	1	-	-	0,00
Total						43,95€

8.3 Costes de adquisición de los datos

En este apartado se ha tenido en cuenta el precio por la adquisición de los datos que, en el caso de las ortoimágenes, el coste es gratuito, puesto que provienen del centro nacional de descargas del IGN. En cuanto a las imágenes de satélite GEOSAT2, no se ha realizado una estimación, puesto que han sido una concesión gratuita de la cátedra del IGN.

Tabla 21. Costes de adquisición de datos

Material	Coste bruto (€)	Cantidad	Tiempo de uso(meses)	Vida útil (meses)	Factor de amortización	Coste total(€)
Recortes de ortoimágenes CNIG	0,00	4500	5	-	-	0,00
Recortes de Imágenes satélites GEOSAT2	0,00	4500	5	-	-	0,00

8.4 Coste de personal o de trabajo

En este apartado se tendrá en cuenta el coste del personal que ha ejecutado el proyecto. Este coste se calculará en base a las horas equivalentes a los créditos ECTS pertenecientes al proyecto. En este caso el TFG representa 12 ECTS, que multiplicado por el número de horas al que equivale cada uno de ellos (12×30), obtenemos un resultado de 360 horas. Se ha considerado que este proyecto se ha realizado por un ingeniero en tecnologías de la información geoespacial y que el salario medio anual referente a este puesto se sitúa en torno a los 26000€ (*Sueldo, s. f.*).

Las horas aproximadas que tendría que trabajar una persona con este puesto serían de unas 2080 teniendo en cuenta 40 horas semanales y 52 semanas de trabajo, lo que resultaría en un salario de 12,5 €/h que en nuestro caso multiplicado por 360 daría un total de 4500€.

Tabla 22. Costes de personal del proyecto

Empleo	ECTS	Horas totales	Salario anual (€)	Salario por hora	Coste total (€)
Ingeniero geoespacial	12	360	26.000	12,5	4.500

A continuación, se mostrará el presupuesto bruto total, el cual se realizará a través de la suma de los anteriores costes.

Tabla 23. Presupuesto bruto Total.

Tipo de coste	Coste total (€)
Hardware	297,86
Software	43,95
Personal	4.500
Total	4841,81

A este presupuesto bruto se le debe añadir el IVA, un 10% de beneficios y los consecuentes posibles gastos por imprevistos.

Tabla 24. Presupuesto bruto con contrato.

Tipo de coste	Coste total (€)
Presupuesto bruto total	4841,81
IVA (21%)	1016.78
Beneficios (10%)	484,18
Gastos imprevistos (13%)	629,43
Total	6.972,2

9. PLANIFICACIÓN

La planificación de este proyecto está conformada por varias fases y su objetivo ha sido el de organizar el desarrollo del mismo, así como facilitar su seguimiento durante los meses de su ejecución. A continuación, se describirán las distintas etapas en las que el proyecto se ha dividido junto con una descripción de las principales actividades llevadas a cabo. Finalmente se aportará un diagrama de Gantt para una comprensión temporal de esta planificación.

9.1. Revisión del estado del arte

Durante esta fase se revisó gran parte de la literatura científica relacionada con la superresolución de imágenes centrándose en las redes convolucionales y su aplicación a partir de imágenes satelitales y ortoimágenes. Además de la lectura de artículos científicos, se investigaron distintos repositorios de código junto con la visualización de un curso online de la universidad de Standford sobre redes convolucionales. Se debe puntualizar que esta fase no solo se limitó al inicio del proyecto, sino que se mantuvo a lo largo del desarrollo del mismo.

9.2. Preparación y generación del *dataset*

En esta fase se llevó a cabo la obtención y preparación del conjunto de datos necesarios para poder realizar el entrenamiento del modelo. Se recogieron imágenes satelitales proporcionadas por el IGN junto con las ortoimágenes correspondientes. Se utilizó QGIS para la selección de las zonas más convenientes y más tarde se realizaron los recortes para su posterior utilización.

9.3. Adaptación e implementación del modelo

En esta etapa se realizó la selección de la arquitectura que se iba a utilizar y se realizaron las modificaciones pertinentes para ajustarla a las necesidades del proyecto. También se realizaron distintas configuraciones de hiper-parámetros. Esta fase estuvo compuesta principalmente por pruebas iterativas para la depuración de errores que surgían a partir de la implementación o modificación del código existente.

9.4. Entrenamiento, validación y ajuste de parámetros

Una vez que el modelo ya se había adaptado al proyecto, se comenzó con el entrenamiento de los modelos a partir del conjunto de datos generado anteriormente. Durante los entrenamientos se realizaban validaciones periódicas para ir evaluando el rendimiento y evitar posibles errores por sobreajuste o explosión del gradiente. A medida que se iban obteniendo los resultados de los entrenamientos, se procedía a ajustar los parámetros del modelo para poder optimizar al máximo el desempeño de la red.

9.5. Elaboración de la memoria

Finalmente, se llevó a cabo la redacción de este documento en el que se ha tratado de recoger de forma detallada las distintas partes que han conformado el TFG.

REFERENCIAS

An Even Easier Introduction to CUDA (Updated). (2025, mayo 2). NVIDIA Technical Blog.

<https://developer.nvidia.com/blog/even-easier-introduction-cuda/>

Anaconda Distribution. (s. f.). Anaconda. Recuperado 21 de mayo de 2025, de

<https://www.anaconda.com/docs/getting-started/anaconda/main>

Aprendizaje profundo | Aula En Abierto. (s. f.). Recuperado 1 de junio de 2025, de

<https://formacion.intef.es/aulaenabrierto/mod/book/tool/print/index.php?id=5077>

Balter, S. (1993). Fundamental properties of digital images. *RadioGraphics*, 13(1), 129-141.

<https://doi.org/10.1148/radiographics.13.1.8426915>

Dong, C., Loy, C. C., He, K., & Tang, X. (2016a). Image Super-Resolution Using Deep Convolutional Networks.

IEEE Transactions on Pattern Analysis and Machine Intelligence, 38(2), 295-307.

<https://doi.org/10.1109/TPAMI.2015.2439281>

Dong, C., Loy, C. C., He, K., & Tang, X. (2016b). Image Super-Resolution Using Deep Convolutional Networks.

IEEE Transactions on Pattern Analysis and Machine Intelligence, 38(2), 295-307.

<https://doi.org/10.1109/TPAMI.2015.2439281>

Dong, C., Loy, C. C., He, K., & Tang, X. (2016c). Image Super-Resolution Using Deep Convolutional Networks.

IEEE Transactions on Pattern Analysis and Machine Intelligence, 38(2), 295-307.

<https://doi.org/10.1109/TPAMI.2015.2439281>

FileZilla—The free FTP solution. (s. f.). Recuperado 21 de mayo de 2025, de <https://filezilla-project.org/>

Gomi, T., Hara, H., Watanabe, Y., & Mizukami, S. (2020). Improved digital chest tomosynthesis image quality by use of a projection-based dual-energy virtual monochromatic convolutional neural network with super resolution. *PLoS ONE*, 15(12), e0244745. <https://doi.org/10.1371/journal.pone.0244745>

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014, junio 10). Generative Adversarial Networks. arXiv.Org. <https://arxiv.org/abs/1406.2661v1>

- Heaton, J. (2018). Ian Goodfellow, Yoshua Bengio, and Aaron Courville: Deep learning. *Genetic Programming and Evolvable Machines*, 19(1), 305-307. <https://doi.org/10.1007/s10710-017-9314-z>
- Hwang, J. H., Park, C. K., Kang, S. B., Choi, M. K., & Lee, W. H. (2024). Deep Learning Super-Resolution Technique Based on Magnetic Resonance Imaging for Application of Image-Guided Diagnosis and Surgery of Trigeminal Neuralgia. *Life*, 14(3), Article 3. <https://doi.org/10.3390/life14030355>
- Hwang, J. J., Jung, Y.-H., Cho, B.-H., & Heo, M.-S. (2020). Very deep super-resolution for efficient cone-beam computed tomographic image restoration. *Imaging Science in Dentistry*, 50(4), 331-337. <https://doi.org/10.5624/isd.2020.50.4.331>
- Image Quality Metrics—MATLAB & Simulink. (s. f.). Recuperado 15 de mayo de 2025, de <https://www.mathworks.com/help/images/image-quality-metrics.html>
- Incio-Flores, F. A., Capuñay-Sánchez, D. L., & Estela-Urbina, R. O. (2023). Modelo de red neuronal artificial para predecir resultados académicos en la asignatura Matemática II. *Revista Electrónica Educare*, 27(1), 1-21.
- Jayanarayan, A., Sowmya, V., & Soman, K. P. (2020). Remote Sensing Image Super-Resolution Using Residual Dense Network. En V. S. Reddy, V. K. Prasad, J. Wang, & K. T. V. Reddy (Eds.), *Soft Computing and Signal Processing* (pp. 721-729). Springer. https://doi.org/10.1007/978-981-15-2475-2_66
- Kaku, K. (2019). Satellite remote sensing for disaster management support: A holistic and staged approach based on case studies in Sentinel Asia. *International Journal of Disaster Risk Reduction*, 33, 417-432. <https://doi.org/10.1016/j.ijdrr.2018.09.015>
- Kim, J., Lee, J. K., & Lee, K. M. (2016a). Accurate Image Super-Resolution Using Very Deep Convolutional Networks. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1646-1654. <https://doi.org/10.1109/CVPR.2016.182>
- Kim, J., Lee, J. K., & Lee, K. M. (2016b). Accurate Image Super-Resolution Using Very Deep Convolutional Networks (No. arXiv:1511.04587; Número arXiv:1511.04587). arXiv. <https://doi.org/10.48550/arXiv.1511.04587>

- Kim, K. G. (2016). Book Review: Deep Learning. *Healthcare Informatics Research*, 22(4), 351-354.
<https://doi.org/10.4258/hir.2016.22.4.351>
- Kramer, P., Steinhardt, A., & Pedretsch, B. (2024). Enhancing Sentinel-2 Image Resolution: Evaluating Advanced Techniques based on Convolutional and Generative Neural Networks (No. arXiv:2410.00516). arXiv. <https://doi.org/10.48550/arXiv.2410.00516>
- LauAI. (2024). Galego: Esquema dunha rede neural convolucional [Graphic]. Own work.
https://commons.wikimedia.org/wiki/File:Esquema_RedeNeuralConvolucional.svg
- Ledig, C., Theis, L., Huszar, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., & Shi, W. (2016, septiembre 15). Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. arXiv.Org. <https://arxiv.org/abs/1609.04802v5>
- Ledig, C., Theis, L., Huszar, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., & Shi, W. (2017). Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network (No. arXiv:1609.04802). arXiv. <https://doi.org/10.48550/arXiv.1609.04802>
- Lim, B., Son, S., Kim, H., Nah, S., & Lee, K. M. (2017). Enhanced Deep Residual Networks for Single Image Super-Resolution (No. arXiv:1707.02921). arXiv. <https://doi.org/10.48550/arXiv.1707.02921>
- Lorna. (2025). Lornatang/VDSR-PyTorch [Python]. <https://github.com/Lornatang/VDSR-PyTorch> (Obra original publicada en 2020)
- Loss Functions in Deep Learning: A Comprehensive Review. (s. f.). Recuperado 15 de mayo de 2025, de <https://arxiv.org/html/2504.04242v1>
- Machine Learning textbook. (s. f.). Recuperado 14 de mayo de 2025, de <https://www.cs.cmu.edu/~tom/mlbook.html>
- Nasrollahi, K., & Moeslund, T. B. (2014). Super-resolution: A comprehensive survey. *Machine Vision and Applications*, 25(6), 1423-1468. <https://doi.org/10.1007/s00138-014-0623-4>
- Nielsen, M. (s. f.). Neural Networks and Deep Learning. Deep Learning.
- Notepad++. (s. f.). Recuperado 21 de mayo de 2025, de <https://notepad-plus-plus.org/>

Optimization Algorithm—An overview | ScienceDirect Topics. (s. f.). Recuperado 15 de mayo de 2025, de
https://www.sciencedirect.com/topics/computer-science/optimization-algorithm?utm_source=chatgpt.com

Panagiotopoulou, A., Grammatikopoulos, L., Charou, E., Bratsolis, E., Madamopoulos, N., & Petrogonas, J. (2020). Very Deep Super-Resolution of Remotely Sensed Images with Mean Square Error and Variance Estimators as Loss Functions (No. arXiv:2007.15417). arXiv.
<https://doi.org/10.48550/arXiv.2007.15417>

(PDF) Inteligencia artificial en la enfermedad de Parkinson y otros trastornos del movimiento. (s. f.). ResearchGate. <https://doi.org/10.24875/KRANION.M23000055>

Pouliot, D., Latifovic, R., Pasher, J., & Duffe, J. (2018). Landsat Super-Resolution Enhancement Using Convolution Neural Networks and Sentinel-2 for Training. *Remote Sensing*, 10(3), Article 3.
<https://doi.org/10.3390/rs10030394>

PyTorch. (s. f.). PyTorch. Recuperado 21 de mayo de 2025, de <https://pytorch.org/>

Rouault, E., Warmerdam, F., Schwehr, K., Kiselev, A., Butler, H., Łoskot, M., Szekeres, T., Tourigny, E., Landa, M., Miara, I., Elliston, B., Chaitanya, K., Plesea, L., Morissette, D., Jolma, A., Dawson, N., Baston, D., de Stigter, C., & Miura, H. (2025). GDAL (Versión v3.11.0) [Software]. Zenodo.
<https://doi.org/10.5281/ZENODO.5884351>

Rubel, A., Ieremeiev, O., Lukin, V., Fastowicz, J., & Okarma, K. (2022). Combined No-Reference Image Quality Metrics for Visual Quality Assessment Optimized for Remote Sensing Images. *Applied Sciences*, 12(4), Article 4. <https://doi.org/10.3390/app12041986>

Ruder, S. (2017). An overview of gradient descent optimization algorithms (No. arXiv:1609.04747). arXiv.
<https://doi.org/10.48550/arXiv.1609.04747>

Spatial without Compromise · QGIS Web Site. (s. f.). Recuperado 21 de mayo de 2025, de <https://qgis.org/>

Stanford University CS231n: Deep Learning for Computer Vision. (s. f.). Recuperado 15 de mayo de 2025, de
<https://cs231n.stanford.edu/>

Sueldo: Geospatial Engineer en España 2025. (s. f.). Glassdoor. Recuperado 12 de mayo de 2025, de

https://www.glassdoor.es/Sueldos/geospatial-engineer-sueldo-SRCH_KO0,19.htm

Superresolución de Imágenes. (s. f.). Glosarix. Recuperado 9 de mayo de 2025, de

<https://glosarix.com/glossary/superresolucion-de-imagenes/>

Wang, X., Yi, J., Guo, J., Song, Y., Lyu, J., Xu, J., Yan, W., Zhao, J., Cai, Q., & Min, H. (2022). A Review of Image Super-Resolution Approaches Based on Deep Learning and Applications in Remote Sensing. *Remote Sensing*, 14, 5423. <https://doi.org/10.3390/rs14215423>

Wang, Z., Chen, J., & Hoi, S. C. H. (2020). Deep Learning for Image Super-resolution: A Survey (No. arXiv:1902.06068). arXiv. <https://doi.org/10.48550/arXiv.1902.06068>

Welcome to Python.org. (2025, mayo 7). Python.Org. <https://www.python.org/>

Wellmann, T., Lausch, A., Andersson, E., Knapp, S., Cortinovis, C., Jache, J., Scheuer, S., Kremer, P., Mascarenhas, A., Kraemer, R., Haase, A., Schug, F., & Haase, D. (2020). Remote sensing in urban planning: Contributions towards ecologically sound policies? *Landscape and Urban Planning*, 204, 103921. <https://doi.org/10.1016/j.landurbplan.2020.103921>

White, J., Codoreanu, A., Zuleta, I., Lynch, C., Marchisio, G., Petrie, S., & Duffy, A. R. (2021). Super-Resolving Beyond Satellite Hardware Using Realistically Degraded Images (No. arXiv:2103.06270). arXiv. <https://doi.org/10.48550/arXiv.2103.06270>

Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., & Fu, Y. (2018). Image Super-Resolution Using Very Deep Residual Channel Attention Networks (No. arXiv:1807.02758). arXiv. <https://doi.org/10.48550/arXiv.1807.02758>

Zhang, Y., Tian, Y., Kong, Y., Zhong, B., & Fu, Y. (2018). Residual Dense Network for Image Super-Resolution (No. arXiv:1802.08797). arXiv. <https://doi.org/10.48550/arXiv.1802.08797>

Zhang, Z. (2018). Multivariate Time Series Analysis in Climate and Environmental Research. Springer International Publishing. <https://doi.org/10.1007/978-3-319-67340-0>

Zhu, Y., Geiß, C., & So, E. (2021). Image super-resolution with dense-sampling residual channel-spatial attention networks for multi-temporal remote sensing image classification. *International Journal of*

Applied Earth Observation and Geoinformation, 104, 102543.

<https://doi.org/10.1016/j.jag.2021.102543>

Zotero | About. (s. f.). Recuperado 21 de mayo de 2025, de <https://www.zotero.org/about/>