

Satellite Imagery Super-resolution using GANs and Aerial Images

Magda Alexandra Trujillo-Jiménez , Francisco Ramiro Iaconis , Debora Pollicelli , Gisela Noelia Revollo Sarmiento , and Claudio Delrieux 

Abstract—Satellite imagery often suffers from limited spatial resolution and, in many cases, high acquisition costs. These factors restrict their use in applications such as urban monitoring, land management, and wildlife studies. This work proposes an AI-based super-resolution approach that leverages high-resolution aerial imagery to train a Generative Adversarial Network. Specifically, the ESRGAN (Enhanced Super-Resolution Generative Adversarial Network) architecture is adapted and trained using aerial orthophotos, enabling the transfer of learned spatial representations to low-resolution satellite images. The trained model is evaluated on satellite image patches at $\times 2$ and $\times 4$ super-resolution scales. Performance is assessed using structural, perceptual, and chromatic metrics, including SSIM-Y, MS-SSIM, LPIPS and CIEDE2000. The results show clear improvements, with increased sharpness, enhanced edge definition, and consistent reconstruction of urban structures and terrain features. From a quantitative perspective, the $\times 2$ scale achieves the best overall metric values, while the $\times 4$ scale maintains stable and meaningful performance despite the higher reconstruction difficulty. These findings demonstrate the feasibility of transferring super-resolution capabilities from aerial images to satellite imagery, even in the presence of spectral and geometric differences between acquisition domains. Overall, this study provides a solid foundation for the development of low-cost, AI-driven satellite image super-resolution models and outlines future research directions focused on dataset expansion, domain adaptation strategies, and sensor-specific architectural improvements.

Link to graphical and video abstracts, and to code:
<https://latam.ieee9.org/index.php/transactions/article/view/10603>

Index Terms—Adversarial Generative Networks, Satellite Imagery, Aerial Imagery, Deep Learning, Super-Resolution

I. INTRODUCTION

SPATIAL resolution of satellite imagery is a key factor in the performance of many remote sensing applications like urban monitoring, land-use planning, natural resource

management, precision agriculture, and environmental conservation. Recently, satellite data have also become essential for infrastructure inspection, land use and land cover, early disaster detection, identification of informal settlements, and monitoring of deforestation or illegal mining activities [1]–[3]. As a consequence, the demand for high-resolution Earth observation imagery has increased steadily. Despite continuous advances in sensor technology, satellite imaging systems face a well-known trade-off between spatial resolution, spectral coverage, temporal revisit, and acquisition cost. Public missions such as Landsat-9 and Sentinel-2 provide global coverage with spatial resolutions between 10 and 30 m per pixel. While these data are freely available, these spatial resolutions are often insufficiently detailed for high-precision tasks. In contrast, commercial platforms such as PlanetScope, Pléiades-Neo, and WorldView-3/4 offer sub-meter resolution imagery. However, their use is limited by high costs, restricted access, and lower revisit flexibility [4]. These trade-offs remain a major challenge for large-scale Earth observation [5]. In addition to resolution constraints, satellite images are affected by atmospheric noise, radiometric inconsistencies, and geometric distortions. These factors complicate cross-sensor comparisons and multi-temporal analysis [6]. As a result, computational approaches that may leverage latent information have gained relevance as complementary solutions to improve image quality.

Image super-resolution (SR) aims to enhance the spatial detail of low-resolution images through algorithmic reconstruction [7], [8]. Classical SR methods based on interpolation, such as bilinear or bicubic resampling, are computationally efficient but fail to recover high-frequency information. They often produce blurred edges and oversmoothed textures [9]. Pan-sharpening techniques partially address this limitation by combining multispectral and panchromatic data. However, this may introduce spectral distortions when sensor characteristics differ [10]. The emergence of Artificial Intelligence, and particularly Deep Learning (DL), advanced image SR research. DL architectures such as CNNs (SRCNN [11], EDSR [12], RCAN [13]) and transformer-based models (SwinIR [14]) have demonstrated strong capabilities in learning complex spatial representations. These models can reconstruct fine structural details by learning non-linear mappings from data. However, their performance strongly depends on the availability of large, high-quality, and well-aligned training datasets, which are scarce in remote sensing scenarios [15].

SR in satellite imagery using sensor-dependent approaches was recently addressed. For example, methods based on Sentinel-2 data leverage supervised learning with carefully

The associate editor coordinating the review of this manuscript and approving it for publication was Adriana Tellez (*Corresponding author: Magda Alexandra Trujillo-Jiménez*).

Magda Alexandra Trujillo-Jiménez and C. Delrieux are with the Laboratorio de Ciencias de las Imágenes, Departamento de Ciencias e Ingeniería de la Computación, Universidad Nacional del Sur, Bahía Blanca, Argentina (e-mails: mtrujillo@cenpat-conicet.gob.ar, and cad@uns.edu.ar).

F. R. Iaconis is with the Instituto de Física del Sur, Departamento de Física, Universidad Nacional del Sur, Bahía Blanca, Argentina (e-mail: francisco.iaconis@uns.edu.ar).

D. Pollicelli is with the Laboratorio de Investigación en Informática, Departamento de Informática, Universidad Nacional de la Patagonia San Juan Bosco, Puerto Madryn, Argentina (e-mail: pollicelli@cenpat-conicet.gob.ar).

G. N. Revollo Sarmiento is with the Instituto de Ecorregiones Andinas, Facultad de Ingeniería, Universidad Nacional de Jujuy, Jujuy, Argentina (e-mail: grevollo@fi.unju.edu.ar).

aligned multi-resolution bands and domain-specific calibration [6]. Similar approaches explored strategies to address discrepancies between training and target domains, including data fusion and domain-aware learning [16]. Even though these works achieved strong performance within their respective domains, they typically require sensor-specific training data or harmonization procedures, which may limit their scalability.

Generative Adversarial Networks (GANs) represent a further step forward in AI-based super-resolution. Unlike purely regression-based methods, GANs explicitly optimize perceptual quality through adversarial learning [17]. This strategy enables the generation of sharper textures and more realistic visual details. However, recent studies have shown that super-resolution models trained on synthetic degradations often overestimate their performance on real-world images. In practice, real images are affected by complex and unknown degradations that are difficult to model. Chen et al. [18] review real-world single image super-resolution (RSISR) methods and emphasize the need for approaches and evaluation protocols that better reflect realistic conditions. These challenges are especially relevant in remote sensing, where sensor characteristics and acquisition conditions strongly impact image quality. In this context, GAN-based models become particularly attractive for applications where visual fidelity is critical [8]. Among these models, the Enhanced Super-Resolution Generative Adversarial Network (ESRGAN) has achieved state-of-the-art results in general image super-resolution tasks [19], [20]. ESRGAN introduces architectural improvements and perceptual loss functions that enhance visual realism. However, its application to satellite imagery remains challenging due to domain-specific factors such as scale variability, spectral differences, and sensor-related degradations. These issues require adaptive training strategies and domain-aware learning.

High-resolution aerial imagery provides an alternative source of detailed spatial information that can be exploited for learning-based super-resolution. Among aerial products, orthophotos stand out due to their geometric correction, uniform scale, and accurate georeferencing [21], [22]. These properties make orthophotos well suited for supervised deep learning, as they preserve fine structural details while ensuring spatial consistency across large areas. From an AI perspective, orthophotos enable the construction of reliable low- and high-resolution image pairs through controlled downsampling processes, addressing the scarcity of high-quality satellite ground truth [15]. Moreover, the spatial richness of orthophotos facilitates the learning of transferable features that can generalize across sensing modalities [10].

In this work, we propose an AI-driven super-resolution approach based on training an ESRGAN model with high-resolution aerial imagery. Specifically, orthophotos are used to generate paired low and high-resolution samples for supervised learning. The learned spatial representations are then transferred to low-resolution satellite images acquired over the same geographic area. This strategy alleviates the lack of high-resolution satellite ground truth data and enables cross-domain knowledge transfer between aerial and satellite modalities. Experimental results show consistent quantitative improvements and noticeable visual enhancement in satel-

lite images reconstructed using the proposed method. These findings demonstrate the feasibility of training GAN-based super-resolution models with aerial data and applying them to satellite imagery. The proposed approach offers a cost-effective and scalable alternative for satellite image enhancement. It also highlights the potential of Artificial Intelligence for developing data-efficient and domain-adaptive super-resolution solutions in Earth observation. The specific contributions of this work are organized along three main axes:

1. **Cross-Domain Feasibility:** We demonstrate that a super-resolution model can learn spatial representations (edges, textures, and urban patterns) from aerial orthophotos and effectively transfer them to satellite imagery from different sensors without requiring explicit domain adaptation or satellite-specific fine-tuning.

2. **Public Data Synergy:** We show that freely available public orthophotos (such as those from IGN Argentina, IBGE, or ICGC) can serve as a scalable and accessible training source, replacing the need for scarce and expensive high-resolution satellite ground truth.

3. **Multi-Metric Evaluation:** We propose a comprehensive evaluation protocol combining structural (SSIM-Y, MS-SSIM), perceptual (LPIPS), and chromatic (CIEDE2000) metrics to assess the robustness of cross-domain super-resolution in remote sensing scenarios.

Recent SR approaches for satellite imagery often rely on sensor-specific training data, such as Sentinel-2 or PlanetScope, requiring carefully aligned ground truth and domain-specific calibration. In contrast, the proposed method explores a cross-domain learning strategy, where high-resolution aerial orthophotos are used as a surrogate training source. This enables the model to learn transferable spatial representations without requiring satellite-specific supervision.

These findings demonstrate the feasibility of training GAN-based super-resolution models with aerial data and applying them to satellite imagery. The proposed approach offers a cost-effective and scalable alternative for satellite image enhancement.

II. METHODS

Fig. 1 summarizes the main stages of the proposed pipeline, from data preparation to super-resolution inference.

A. Study Area

The study area selected for this work is the city of Lleida, located in the autonomous community of Catalonia, in north-eastern Spain. Lleida presents a heterogeneous landscape that includes dense urban areas, peri-urban zones, and extensive agricultural land. This diversity makes it a suitable testbed for evaluating super-resolution performance across different land surface types. The terrain is predominantly flat, which reduces geometric distortions and facilitates the comparison between aerial and satellite imagery. At the same time, the spatial distribution of buildings, road networks, and vegetation provides a wide range of structural patterns. These patterns are valuable for training and validating AI-based super-resolution models. Figure 2 shows a geospatial view

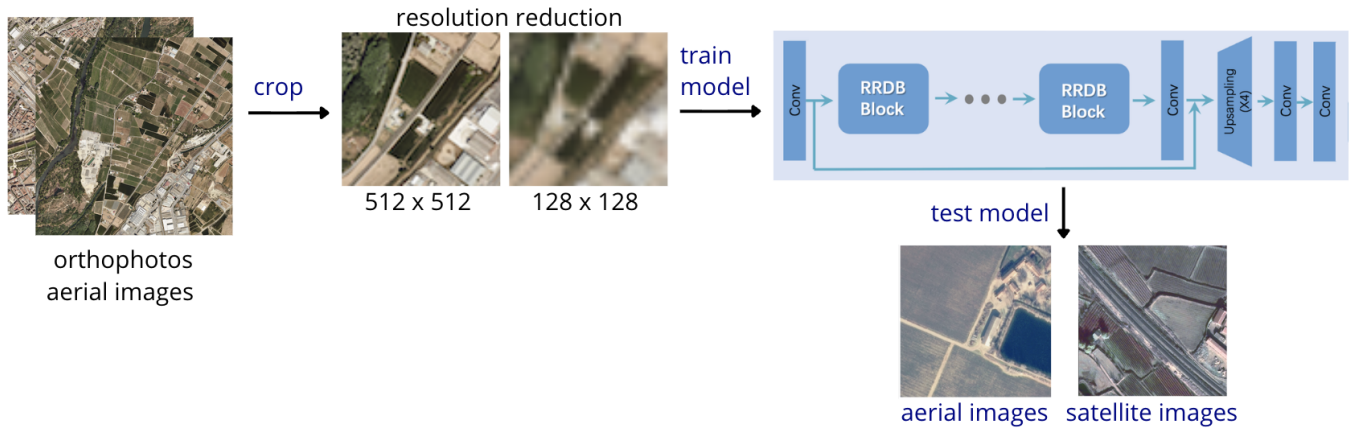


Fig. 1. Overview of the proposed AI-based super-resolution workflow. High-resolution aerial orthophotos are first cropped into fixed-size patches and synthetically degraded to generate paired high- and low-resolution samples. These pairs are used to train the model. The trained network is then evaluated on aerial and satellite images, enabling cross-domain super-resolution without satellite-specific training data.

of the city of Lleida and the selected study area. The main urban zones and relevant spatial features considered in this work are highlighted. The map was generated using data provided by the Institut Cartogràfic i Geològic de Catalunya (ICGC). Lleida was selected not only for data availability but also for its representative land cover diversity and complexity within a compact geographic extent, enabling efficient model training while capturing varied spectral signatures essential for generalization. These orthophotos offer high spatial resolution and accurate georeferencing, making them well suited for supervised learning and cross-domain transfer experiments.

B. Training Data Preparation and Pre-processing

A dataset composed of two high-resolution aerial orthophotos of the city of Lleida was used to train the proposed AI-based super-resolution model. These orthophotos were acquired through drone flights and subsequently orthorectified to remove geometric distortions caused by terrain relief and acquisition geometry. This preprocessing step ensures a geometrically consistent and spatially accurate representation of the Earth's surface, which is essential for supervised deep learning tasks [21]–[23]. The aerial images have a spatial resolution of 25 cm per pixel and were provided in RGB format. They were accessed through the *Open ICGC* plugin of the QGIS geographic information system [24]. Orthophotos represent vertically projected aerial images with uniform scale, making them well suited for learning spatial patterns in AI-based computer vision models. The RGB composition captures visible-spectrum information and preserves the natural appearance of urban structures, roads, and vegetation. High-resolution aerial orthophotos (25 cm/pixel) served as the ground truth for training. Low-resolution counterparts were generated via synthetic downsampling to 50 cm ($\times 2$ scale) and 100 cm ($\times 4$ scale). This allows the model to learn the reconstruction of 25 cm structures from coarser inputs.

The original images, with dimensions of 15715×12261 and 13480×14564 pixels, were processed to generate paired high- and low-resolution samples for supervised training.

High-resolution reference patches of 512×512 pixels were extracted using a sliding window strategy. Each patch was then downsampled to 128×128 pixels to create its low-resolution counterpart. Area-based interpolation was employed during downsampling, as it better preserves structural information when reducing image size. This procedure enabled the creation of 1418 paired samples at scale factors of $\times 2$ and $\times 4$. This ensures a quantitative assessment based on a statistically significant sample of over 370 million pixels, capturing the complex spatial features required for robust super-resolution in remote sensing. The resulting dataset provides consistent input–target pairs, allowing the GAN-based model to learn the mapping between low- and high-resolution representations and to capture fine-grained spatial details from aerial imagery.

For reproducibility, the dataset was divided into training and testing subsets at the image level, preventing overlap between patches extracted from the same regions. Patch extraction was performed using a sliding window strategy with fixed size (512×512 pixels) and consistent stride, generating a total of 1418 paired samples. Low-resolution counterparts were obtained through area-based downsampling, following a controlled degradation process consistent with standard SR training pipelines. In addition, fixed random seeds were used during data preparation and training to ensure consistent patch sampling and reproducible model behavior across runs.

C. Satellite Imagery for Evaluation

For the evaluation stage, high-resolution pan-sharpened satellite images acquired by the SuperView-1 platform were used. Unlike the native 2 m multispectral product, we employed the pan-sharpened 0.5 m resolution RGB composites to establish a more challenging evaluation baseline that better approximates the quality of aerial reference data. The 0.5 m pan-sharpened images were used as the reference (GT). To test the models, these images were downsampled to 1 m (for $\times 2$ testing) and 2 m (for $\times 4$ testing). The selected scenes present approximately 10% cloud coverage. The blue (450–520 nm), green (520–590 nm), and red (630–690 nm) bands were



Fig. 2. Geospatial view of the city of Lleida showing the study area used in this analysis. Main urban zones and spatial features relevant to this work are highlighted. Source: Institut Cartogràfic i Geològic de Catalunya (ICGC).

selected and combined to generate RGB images with an effective spatial resolution of 0.5 m per pixel, derived from the fusion of 2 m multispectral and 0.5 m panchromatic data. From these images, RGB patches of 615×615 pixels were used as input for the super-resolution experiments. These patches represent heterogeneous land cover conditions, including urban areas, vegetation, and open terrain. A critical requirement for training deep learning models for image super-resolution is the availability of a large and diverse dataset. This diversity ensures that both frequent and less common spatial patterns are statistically represented. When this condition is not met, the model may favor dominant structures present in the training data. For example, in agricultural regions with uncommon planting geometries, such as circular sowing patterns instead of typical linear arrangements, the model may tend to reconstruct the prevailing linear patterns, introducing reconstruction bias. However, in this study, the interpolation scale is smaller than the characteristic size of the spatial patterns of interest. As a result, this limitation does not constitute a critical constraint for the overall quality of the super-resolved outputs. The generated images remain visually coherent and structurally consistent across the evaluated scenes.

D. Evaluation Methods and Comparison

To assess the performance of the proposed super-resolution model, a set of widely used quantitative and perceptual evaluation metrics was employed. These metrics allow measuring both the structural similarity and the perceived visual quality between the super-resolved images and their corresponding reference images [25]. First, the Structural Similarity Index Measure (SSIM) was computed on the luminance channel (Y) of the images. SSIM evaluates image similarity by jointly considering luminance, contrast, and structural information. The Y channel was selected because it plays a dominant role in human visual perception and is commonly used for quality assessment in super-resolution tasks [26]. In addition, the Multi-Scale Structural Similarity Index (MS-SSIM) was included to extend the structural evaluation across multiple spatial scales. MS-SSIM captures consistency in structure and contrast at different resolution levels, providing a more robust perceptual assessment of image quality [27]. To account for perceptual differences that are difficult to capture with traditional metrics, the Learned Perceptual Image Patch Similarity (LPIPS) metric was also used [13], [28]. LPIPS relies on deep feature representations extracted from pre-trained neural networks. In this work, AlexNet was employed as the backbone network. Lower

LPIPS values indicate higher perceptual similarity between the generated and reference images. Finally, color fidelity was evaluated using the CIEDE2000 color difference metric, computed in the CIE Lab color space [29]. This metric quantifies average chromatic differences and is particularly relevant for aerial and satellite imagery, where color consistency is critical for interpretation and downstream analysis. Together, these four metrics provide a balanced and complementary evaluation framework. They jointly assess structural accuracy, perceptual quality, and chromatic stability of the super-resolved images. For result presentation, the evaluation was conducted on nine aerial image patches and nine satellite image patches. These samples were selected to represent diverse textures, contrast conditions, and levels of spatial detail.

III. MODEL DESCRIPTION

A. Model Adoption

To perform the SR task, we adopted the Real-ESRGAN framework [19], [20], which is specifically designed to handle complex, non-idealized degradations. As our study focuses on the empirical feasibility of cross-domain transfer (from aerial orthophotos to satellite imagery) rather than architectural modification, we utilize the standard Real-ESRGAN configuration for a robust and reproducible baseline. The generator is based on the Residual-in-Residual Dense Block (RRDB) architecture, consisting of 16 blocks that facilitate deep feature reuse and high-frequency detail reconstruction. At the input stage, a pixel-unshuffle operation is employed to reduce spatial dimensions while increasing channel depth, optimizing GPU memory efficiency during the processing of large-scale remote sensing patches. The discriminator follows a U-Net design with skip connections, providing pixel-level feedback that is essential for maintaining the local realism of urban textures and edges. Training stability is further enforced through spectral normalization in the discriminator and a composite loss function. This includes a pixel-wise L_1 loss for structural consistency, a VGG-based perceptual loss to enhance visual fidelity, and a relativistic GAN (RaGAN) loss to promote the generation of sharp, realistic textures. By maintaining the original architecture, we isolate the experimental variable to the training data source, evaluating how effectively features learned from high-resolution aerial orthophotos generalize to the satellite domain. This cross-domain approach serves as a necessary methodological alternative to a traditional ablation study. In satellite super-resolution, the lack of sub-decimeter satellite Ground Truth (GT) creates a ‘resolution ceiling’ that prevents learning ultra-fine structural details. By utilizing aerial imagery, the model is exposed to high-frequency features physically absent in native satellite products, demonstrating that the learned representations successfully transfer to different sensors like SuperView-1 (see Fig. 3).

B. Training Configuration

Although Real-ESRGAN was originally designed and optimized for general-purpose natural images, it has been successfully adapted to other application domains, such as medical imaging and degraded visual data [30], [31]. In this work, the

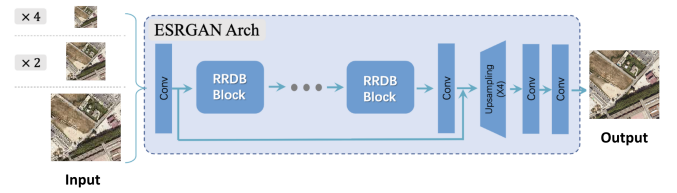


Fig. 3. Overall architecture of the Real-ESRGAN model, based on a generative adversarial network with 16 Residual-in-Residual Dense Blocks (RRDB) in the generator. The model integrates pixel-unshuffle operations, residual connections, and spectral normalization to improve training stability and super-resolution image quality. Adapted from [20].

training procedure was specifically adjusted to accommodate the characteristics of aerial and satellite imagery, which present different spatial patterns, textures, and degradation processes compared to standard photographic datasets. To optimize computational efficiency and ensure stable adversarial training, several hyperparameters were carefully configured. The batch size was initially set to two images during the preliminary training phase and later reduced to one image in the optimized configuration to limit GPU memory usage. Training was accelerated using GPU-based parallelization, and the total number of iterations was fixed at 100,000, providing a suitable balance between convergence quality and computational cost.

The optimization strategy combined multiple loss functions to jointly enhance structural accuracy and perceptual realism. A pixel-wise L_1 loss was employed to preserve low-frequency content and overall image structure, while a perceptual loss based on deep feature representations encouraged visually consistent textures. In addition, an adversarial loss using the relativistic average GAN (RaGAN) formulation was incorporated to promote sharper and more realistic high-frequency details. The relative weights of these losses were tuned to favor image sharpness without compromising training stability. Both the generator and discriminator were trained using an initial learning rate of 1×10^{-4} , with the Adam optimizer configured with $\beta_1 = 0.9$ and $\beta_2 = 0.99$. A MultiStepLR scheduling strategy was applied, reducing the learning rate by a factor of 0.5 after 50,000 iterations. All experiments were conducted on a dedicated server equipped with an Intel Core i9-13900KF CPU and an NVIDIA RTX 3080 Ti GPU with 12 GB of GDDR6 VRAM (ASUS TUF). The full implementation developed for this study, including training, evaluation, and inference scripts, is publicly available at <https://github.com/aletrujim/Super-Reso>. Regarding model initialization, the generator and discriminator weights were not trained from scratch. Instead, we performed fine-tuning starting from the provided pre-trained weights of Real-ESRGAN, which were originally trained on large-scale natural image datasets (DF2K and OST). This transfer learning strategy allows the model to leverage robust low-level feature representations (such as basic edges and textures) before adapting them to the specific spatial patterns of aerial orthophotos and satellite imagery. The training process then refined these weights over 100,000 iterations using our specific aerial dataset.

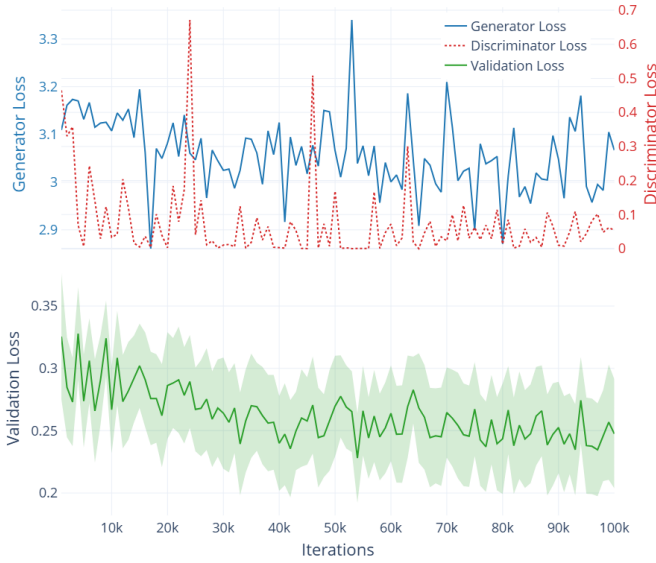


Fig. 4. Training dynamics of the GAN-based super-resolution model. The generator loss (blue) and discriminator loss (red, dashed) are shown across iterations, indicating stable adversarial training without divergence. The validation loss (green) exhibits a gradual decrease and stabilization over time, suggesting consistent convergence and generalization performance. The shaded region represents variability across validation samples.

IV. RESULTS

Both qualitative and quantitative analyses are reported. The qualitative assessment focuses on the visual evolution of the super-resolved outputs across different training stages. The quantitative evaluation is based on standardized image quality metrics. Together, these analyses provide a comprehensive assessment of the model's ability to recover high-resolution details and to faithfully reconstruct relevant spatial structures of the terrain.

A. Model Performance

Prior to the quantitative evaluation, the model's convergence was monitored to ensure robustness. As shown in Fig. 4, the adversarial losses present a characteristic oscillatory yet bounded behavior, typical of stable GAN training. Complementarily, the sustained reduction in validation loss suggests a progressive improvement in reconstruction capability, providing empirical evidence that the model avoids overfitting and generalizes effectively across the training iterations.

To analyze the learning dynamics and the effect of training checkpoints on reconstruction quality, intermediate outputs were generated at three key stages of the training process: 1,000, 50,000, and 100,000 iterations. Display images (see Fig. 5) allow a visual comparison between the low-quality inputs (LQ), the super-resolved outputs (SR), and the high-resolution ground truth (GT). At the early stage of training (1,000 iterations), the SR images exhibit smooth edges and limited texture recovery. This behavior is consistent with an initial learning phase dominated by content loss optimization,

where the model primarily focuses on coarse structural reconstruction. At 50,000 iterations, a noticeable improvement is observed. Textural details, agricultural patterns, and parcel boundaries become more clearly defined, reflecting the increasing influence of perceptual and adversarial losses. At the final checkpoint (100,000 iterations), the model produces sharper and more coherent reconstructions. Fine-scale details such as roads, buildings, and crop rows are more accurately recovered and show strong visual consistency with the GT images. These qualitative results illustrate the gradual convergence of the model and highlight the complementary role of the different loss components in guiding the learning process. Quantitative results are presented in the following subsections.

B. Model Evaluation on Aerial Images

The quantitative evaluation highlights clear performance differences between the $\times 2$ and $\times 4$ upscaling factors, as summarized in Table I. At the $\times 2$ scale, the model achieves its best overall performance across all evaluated metrics. In particular, SSIM-Y reaches a mean value of 0.812 with a low standard deviation (0.044), indicating strong structural consistency and stable luminance preservation with respect to the high-resolution references. This behavior confirms the model's ability to preserve urban geometry, edge continuity, and surface textures in aerial imagery. Similarly, MS-SSIM reports a high mean value of 0.969, reflecting robust multi-scale structural coherence and consistent contrast reproduction across different spatial frequencies. From a perceptual standpoint, LPIPS obtains a relatively low mean value of 0.137, suggesting limited perceptual divergence from the ground truth, especially in complex regions such as vegetated areas and agricultural patterns. Regarding chromatic fidelity, the CIEDE2000 metric shows a mean value of 3.206, indicating good color consistency and controlled chromatic deviations, even in regions prone to tonal amplification such as shadows and metallic surfaces.

At the $\times 4$ scale, all metrics exhibit a moderate degradation, which is expected given the increased difficulty of reconstructing high-frequency information from a more severely downsampled input. SSIM-Y decreases to a mean value of 0.779, while MS-SSIM reaches 0.964, still maintaining acceptable structural coherence but with a reduced capacity to recover fine details. The perceptual metric LPIPS increases to 0.187, reflecting more noticeable perceptual differences, particularly in highly textured areas. Likewise, CIEDE2000 rises to a mean value of 3.979, indicating increased sensitivity to chromatic variations when extrapolating details absent from the low-resolution input. Despite these differences, the results at $\times 4$ remain consistent and visually plausible. Overall, the quantitative metrics confirm that the model performs optimally at $\times 2$, while still delivering robust and meaningful improvements at $\times 4$. These findings demonstrate a balanced trade-off between structural fidelity, perceptual quality, and color stability, validating the effectiveness of the proposed AI-based super-resolution approach on aerial imagery.

These quantitative results are supported by the qualitative examples shown in Fig. 6. The figure illustrates representative

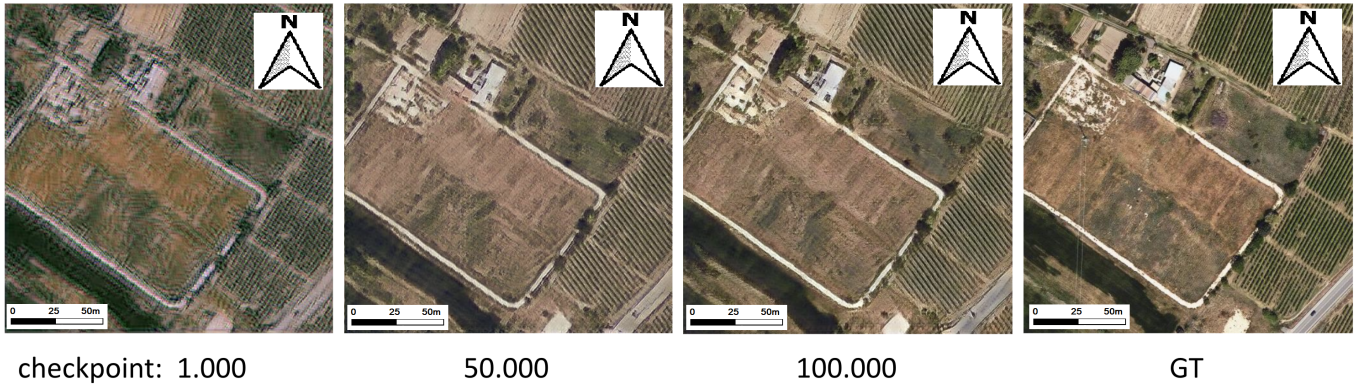


Fig. 5. Visual comparison of super-resolution results at different training checkpoints. From left to right: super-resolved outputs obtained after 1,000, 50,000, and 100,000 training iterations, followed by the high-resolution ground truth (GT).

TABLE I
DESCRIPTIVE STATISTICS OF THE EVALUATION METRICS
COMPUTED ON AERIAL AND SATELLITE IMAGES USED TO
VALIDATE THE PROPOSED MODEL

Dataset	Scale	Metric	Mean	Std	Min	Max
Aerial	$\times 2$	SSIM-Y	0.812	0.044	0.740	0.865
		MS-SSIM	0.969	0.005	0.961	0.977
		LPIPS	0.137	0.026	0.098	0.171
		CIEDE2000	3.206	0.334	2.382	3.557
	$\times 4$	SSIM-Y	0.779	0.064	0.701	0.866
		MS-SSIM	0.964	0.012	0.948	0.983
		LPIPS	0.187	0.044	0.111	0.255
		CIEDE2000	3.979	0.420	3.457	4.755
Satellite	$\times 2$	SSIM-Y	0.755	0.096	0.577	0.852
		MS-SSIM	0.965	0.030	0.887	0.980
		LPIPS	0.180	0.084	0.092	0.342
		CIEDE2000	2.842	1.033	1.807	5.046
	$\times 4$	SSIM-Y	0.729	0.120	0.519	0.847
		MS-SSIM	0.970	0.009	0.951	0.982
		LPIPS	0.216	0.082	0.107	0.319
		CIEDE2000	3.162	0.333	2.625	3.623

aerial image patches processed with the proposed model at $\times 2$ and $\times 4$ scale factors, including the low-resolution input, the super-resolved outputs, and the corresponding regions of interest highlighted with red boxes. At $\times 2$, the model recovers sharper edges, clearer object boundaries, and more consistent textures, closely matching the visual appearance of the reference images. At $\times 4$, although the reconstruction task is more challenging, the model still enhances fine-scale structures such as roads, crop patterns, and building outlines, producing visually coherent results with improved spatial detail. The zoomed-in regions emphasize the model's ability to reconstruct high-frequency information and fine textures, in agreement with the trends observed in the SSIM, LPIPS, and CIEDE2000 metrics.

C. Model Evaluation on Satellite Images

The quantitative and qualitative evaluation on satellite image patches demonstrates that, despite being trained exclusively on aerial imagery, the proposed model exhibits a strong generalization capability when applied to satellite data. For both

super-resolution scales, the model produces visually coherent and stable reconstructions, achieving noticeable improvements in spatial detail and overall image quality (see Fig. 7). The generated images present sharper contours and increased clarity compared to their low-resolution counterparts, particularly in urban structures such as buildings, rooftops, and road networks, as well as in terrain elements and vegetated areas.

At the $\times 2$ scale, the model achieves a solid and consistent performance. SSIM-Y values predominantly fall within the range of 0.66–0.85, indicating effective recovery of structural luminance information from the original images. This behavior is reinforced by the high MS-SSIM scores, which remain above 0.96 in most cases and exhibit low variability, reflecting stable preservation of structural information across multiple spatial scales. From a perceptual perspective, LPIPS values are relatively low (≈ 0.09 – 0.34), with a mean around 0.18, suggesting perceptually meaningful improvements in texture reconstruction and edge definition. Regarding color fidelity, CIEDE2000 values are mainly concentrated between 2.0 and 3.2, with a single outlier close to 5, associated with regions of high spectral heterogeneity where chromatic deviations are more likely to occur.

For the $\times 4$ scale, as expected given the increased reconstruction difficulty, a moderate performance degradation is observed. SSIM-Y exhibits a broader range (≈ 0.52 – 0.85), although the model still maintains satisfactory structural recovery for most image patches. MS-SSIM remains high and stable (≈ 0.95 – 0.98), indicating that structural coherence is largely preserved even at higher upscaling factors. LPIPS values increase slightly compared to $\times 2$ (≈ 0.11 – 0.32), reflecting more noticeable perceptual differences in highly detailed textures, yet remaining within acceptable ranges for $\times 4$ super-resolution tasks. Similarly, CIEDE2000 shows a moderate increase (≈ 2.6 – 3.6), highlighting the additional challenge of preserving chromatic consistency when extrapolating fine-scale details from lower-resolution inputs.

It is important to note that some metrics exhibit relatively high variability, particularly in the satellite domain. For instance, SSIM-Y at $\times 4$ shows a higher standard deviation (0.120), and CIEDE2000 at $\times 2$ presents a standard deviation of 1.033. This is mainly associated with the heterogeneous

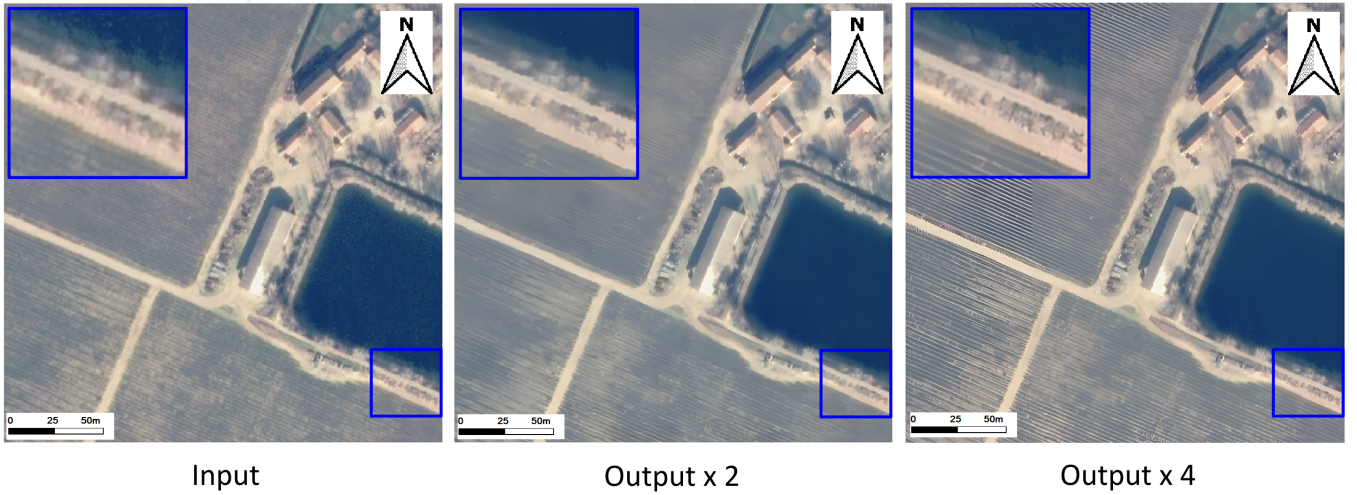


Fig. 6. Super-resolution results on aerial images. The input image and the outputs at $\times 2$ and $\times 4$ are shown. Highlighted regions emphasize improvements in edge sharpness, texture reconstruction, and structural continuity at different spatial scales.

nature of satellite imagery, which includes diverse land cover types, contrast conditions, and spectral responses. In particular, an outlier value of CIEDE2000 (5.046) was observed in regions with high spectral heterogeneity. This can be attributed to the increased difficulty of preserving chromatic consistency when reconstructing fine details from low-resolution inputs, especially in areas with complex textures or mixed materials. These results highlight the sensitivity of color-based metrics to local variations and reinforce the importance of complementing them with perceptual and structural measures such as LPIPS and MS-SSIM, which remain more stable across samples. This behavior is consistent with previous findings in remote sensing super-resolution, where color-based metrics are more sensitive to spectral variability.

Overall, these results indicate that the this approach maintains a balanced trade-off between structural fidelity, perceptual quality, and chromatic stability when applied to satellite imagery, confirming its robustness and suitability for SR in cross-domain scenarios.

The proposed method is particularly suitable in contexts where perceptual quality is prioritized over strict pixel-wise fidelity. Table II presents a quantitative comparison between bicubic interpolation, a GAN-based SR baseline [32], and the proposed model on both aerial and satellite images at $\times 2$ scale. Bicubic interpolation consistently achieves the highest SSIM and MS-SSIM values across both datasets, indicating superior pixel-wise structural similarity. It also provides the lowest CIEDE2000 values, reflecting better color fidelity. However, the proposed model significantly outperforms both baselines in terms of LPIPS, achieving the lowest perceptual error in both aerial (0.137) and satellite (0.180) domains. Compared to bicubic interpolation, this represents a substantial improvement in perceptual quality, indicating that the model generates more visually realistic textures and details. The GAN-based SR baseline provides intermediate performance between bicubic and the proposed method, improving perceptual quality compared to bicubic while maintaining higher structural similarity than the proposed model. These results

TABLE II
COMPARISON OF BASELINE METHODS ON AERIAL AND SATELLITE IMAGES AT $\times 2$ SCALE. RESULTS OF THE PROPOSED MODEL ARE REPORTED IN TABLE I

Dataset	Method	Metric	Mean
Aerial	Bicubic	SSIM-Y	0.912
		MS-SSIM	0.996
		LPIPS	0.462
		CIEDE2000	1.704
	GAN-based SR	SSIM-Y	0.884
		MS-SSIM	0.994
		LPIPS	0.318
		CIEDE2000	2.039
Satellite	Bicubic	SSIM-Y	0.907
		MS-SSIM	0.996
		LPIPS	0.389
		CIEDE2000	1.303
	GAN-based SR	SSIM-Y	0.842
		MS-SSIM	0.993
		LPIPS	0.221
		CIEDE2000	1.751

highlight a clear trade-off between pixel-wise fidelity and perceptual realism. While bicubic interpolation favors smooth reconstructions that maximize structural similarity, the proposed model prioritizes perceptual quality. This behavior is consistent with findings reported in [32], where learning-based SR methods improve perceptual quality at the expense of pixel-wise accuracy. Such a trade-off is particularly relevant in applications where visual interpretability is more important than exact pixel reconstruction.

D. Statistical Significance Analysis

To validate the reliability of the performance metrics reported in Table II and address the inherent variability in aerial and satellite datasets, a comprehensive statistical analysis was conducted.

1. Normality Test and Method Selection: Prior to significance testing, the distribution of the results for each metric (SSIM-Y, MS-SSIM, LPIPS, and CIEDE2000) was evaluated

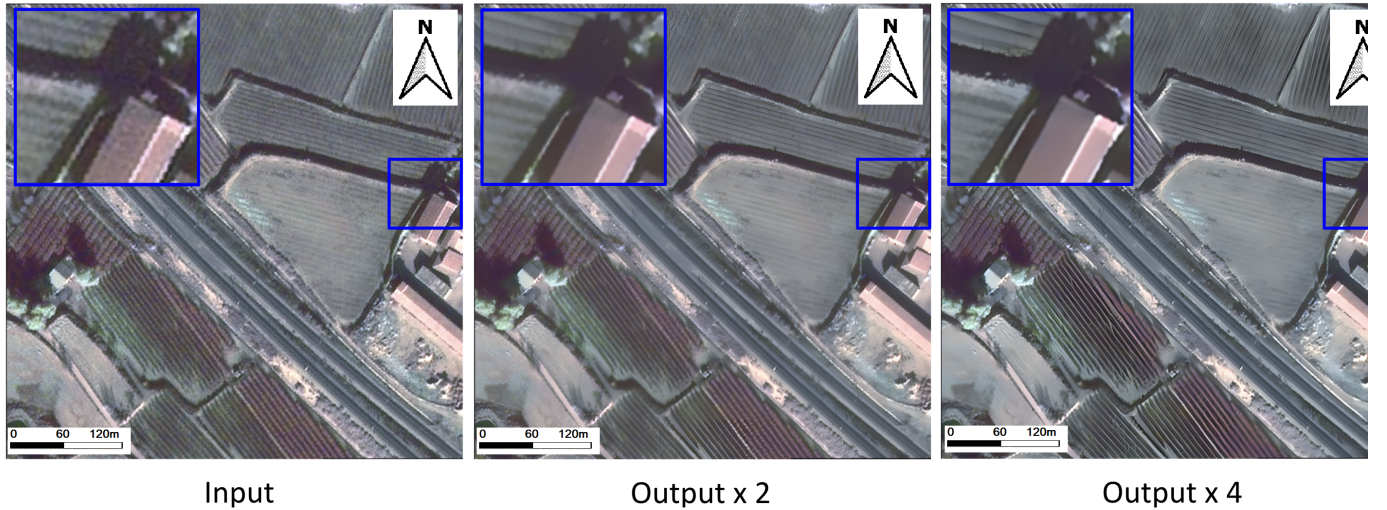


Fig. 7. Super-resolution results on satellite images at $\times 2$ and $\times 4$. Highlighted regions show enhanced structural details and improved texture reconstruction compared to the low-resolution input.

using the Shapiro-Wilk test. The results indicated that the majority of the groups did not follow a normal distribution ($p < 0.05$), justifying the use of non-parametric statistical methods, which are more robust for small sample sizes ($N = 9$) and the presence of potential outliers.

2. Analysis of Scaling Factors ($\times 2$ vs. $\times 4$) The Wilcoxon signed-rank test was used to evaluate the impact of the scaling factor on model performance. The results revealed a distinct behavior between the two domains. In aerial domain the transition from $\times 2$ to $\times 4$ resulted in a statistically significant change in quality for SSIM-Y ($p = 0.043$), LPIPS ($p = 0.0039$), and CIEDE2000 ($p = 0.0039$). This confirms a clear scaling effect where structural and perceptual fidelity are sensitive to the magnification factor. Interestingly, MS-SSIM ($p = 0.437$) did not show significant degradation, suggesting that multi-scale structural features are effectively preserved even at higher scales. In satellite domain, in contrast, no metrics showed statistically significant differences between scales ($p > 0.05$), with LPIPS ($p = 0.0546$) being the closest to the significance threshold. This lack of statistical divergence suggests that the proposed model maintains a highly consistent performance level in the satellite domain regardless of the scaling factor. This stability is a key finding, as it indicates that the learned representations for satellite imagery are robust enough to handle $\times 4$ upsampling with a quality loss that is not statistically distinct from the $\times 2$ case. The high p -values in the satellite domain, particularly for MS-SSIM ($p = 0.91$) and SSIM-Y ($p = 0.28$), further emphasize the model's architectural stability when dealing with the lower signal-to-noise ratio typical of satellite sensors.

3. Domain Invariance (Aerial vs. Satellite): To assess the model's robustness across different sensing domains, the Mann-Whitney U test was performed, comparing metrics between the aerial and satellite domains for each scaling factor. At scale $\times 2$ no statistically significant differences were found across any of the evaluated metrics ($p > 0.05$), indicating that at moderate magnification, the model performs consistently

regardless of the image source. At scale $\times 4$, the structural and perceptual metrics—SSIM-Y ($p = 0.452$), MS-SSIM ($p = 0.231$), and LPIPS ($p = 0.331$)—showed no significant differences between domains. This supports our hypothesis that the proposed architecture achieves a high degree of structural domain invariance, maintaining consistent geometry and detail reconstruction across both aerial and satellite imagery.

4. Chromatic Variance at $\times 4$: In contrast, the CIEDE2000 metric showed a statistically significant difference ($p = 0.001478$) between the two domains at the $\times 4$ scale. This indicates that while the structural integrity is preserved, color fidelity is more sensitive to the domain-specific spectral characteristics when higher upsampling is required. This can be attributed to the inherent differences in atmospheric interference and sensor calibration between satellite and aerial platforms, which are more pronounced at higher magnification levels.

V. DISCUSSION AND CONCLUSIONS

This work demonstrates that AI-based super-resolution models trained exclusively on UAV aerial imagery can effectively enhance low-resolution satellite images. Beyond the quantitative metrics reported in the previous section, the results indicate that the features learned from aerial orthophotos—specifically high-frequency edges and urban textures—are sufficiently domain-invariant to be projected onto satellite sensors. The fact that the model maintains structural coherence across scales suggests that the Residual-in-Residual Dense Blocks (RRDB) are effective at capturing multi-scale spatial dependencies that exist regardless of the acquisition platform.

From a methodological perspective, the main contribution of this study lies in demonstrating cross-domain generalization in AI-based super-resolution without explicit domain adaptation or satellite-specific fine-tuning. Although trained solely on aerial imagery, the model successfully transfers learned representations to satellite data. This suggests that the network captures domain-invariant visual features, such as

edges, textures, and spatial organization, which are essential for high-quality super-resolution. Compared to state-of-the-art approaches that often rely on sensor-specific training data, this work highlights a more flexible and data-efficient alternative.

Nevertheless, certain limitations are observed. Differences in spectral response and radiometric characteristics between aerial and satellite sensors introduce challenges in color consistency. While aerial images typically present more uniform visible-spectrum properties, satellite imagery exhibits higher radiometric variability due to sensor design and atmospheric effects. These factors partially explain the chromatic deviations observed in some $\times 4$ reconstructions. Furthermore, the absence of native high-resolution satellite ground truth required the use of pan-sharpened products as a proxy reference. While practical, this approach implies that some recovered details might be influenced by priors learned from aerial imagery or fusion artifacts inherent to the pan-sharpening process, rather than representing the absolute ground truth of the scene.

Future work will focus on strengthening the proposed framework through several extensions. Expanding the training dataset to include diverse geographic regions, seasonal conditions, and acquisition scenarios is expected to improve robustness. In addition, exploring state-of-the-art AI architectures, such as transformer-based or hybrid convolutional–transformer models, may further enhance the recovery of fine-scale details. Finally, incorporating domain adaptation strategies represents a promising direction to explicitly align learned representations with the spectral and radiometric characteristics of different sensors.

In conclusion, this study confirms the potential of AI-driven super-resolution as a cross-domain enhancement strategy for remote sensing imagery. By leveraging knowledge learned from aerial data, the proposed approach offers a scalable and cost-effective alternative for improving satellite image quality, contributing to the advancement of state-of-the-art AI-based super-resolution methods in Earth observation.

VI. FUTURE WORK

Future work will focus on several directions to further strengthen the proposed AI-based super-resolution framework. Recent studies on real-world super-resolution emphasize the importance of accurately modeling sensor-specific degradations during training, rather than relying solely on idealized downsampling operations [18]. While our synthetic degradation through area-based interpolation preserves structural information, real satellite imagery exhibits additional degradations including sensor MTF effects, photon noise, and atmospheric scattering. Future work will incorporate more realistic degradation models as suggested by Chen *et al* [18]. Such strategies could better approximate real satellite acquisition processes and act as effective data augmentation mechanisms. A broader validation across different geographic regions, land-cover types, and acquisition conditions would further strengthen the assessment of the proposed framework's robustness and transferability. In addition, comparisons with remote sensing-specific super-resolution methods, as well as analyses of the impact of cloud cover and shadowed areas

on reconstruction quality, will be considered. Finally, future studies will explore alternative evaluation protocols based on native high-resolution satellite imagery and no-reference perceptual quality metrics. Another aspect to be addressed concerns the size of the training dataset. Although the current results demonstrate stable and consistent performance, the limited number of aerial images may be considered restrictive for GAN-based learning. To mitigate this limitation, future work will investigate the use of extensive data augmentation techniques, including geometric transformations and photometric variations, in order to increase data diversity and improve training robustness [33]. Finally, an important direction for future research is to evaluate the impact of the proposed super-resolution approach on downstream remote sensing tasks. In particular, the use of super-resolved satellite images as input for automatic analysis pipelines, such as object detection or semantic segmentation, appears promising. Recent foundation models for segmentation, such as the Segment Anything Model (SAM) [34], [35], could benefit from enhanced spatial detail, potentially improving target delineation and overall task performance. Assessing the contribution of AI-based super-resolution in such downstream applications would provide additional evidence of its practical value in real-world Earth observation scenarios.

ACKNOWLEDGMENTS

The authors would like to acknowledge the Institut Cartogràfic i Geològic de Catalunya (ICGC) for providing the geospatial data used to generate the map of the study area.

This manuscript benefited from the use of an artificial intelligence (AI) system for English-language writing support. The AI tool was used exclusively to assist in language refinement, grammar correction, and stylistic improvement of the manuscript text. No AI system was used to generate scientific results, experimental design, data analysis, or conclusions.

REFERENCES

- [1] J. Shan, Q. Weng, M. Ehlers, D. A. Quattrochi, G. Zhou, U. Stilla, and S. Anderson, *Urban Remote Sensing*. CRC Press, 2018, doi: <https://doi.org/10.1201/9781315166612>.
- [2] S. Basheer, X. Wang, A. A. Farooque, R. A. Nawaz, K. Liu, T. Adekanmbi, and S. Liu, "Comparison of land use land cover classifiers using different satellite imagery and machine learning techniques," *Remote Sensing*, vol. 14, no. 19, p. 4978, 2022, doi: <https://doi.org/10.3390/rs14194978>.
- [3] M. A. Trujillo-Jiménez, A. L. Liberoff, N. Pessacq, C. Pacheco, L. Díaz, and S. Flaherty, "Satred: New classification land use/land cover model based on multi-spectral satellite images and neural networks applied to a semiarid valley of patagonia," *Remote Sensing Applications: Society and Environment*, vol. 26, p. 100703, 2022, doi: <https://doi.org/10.1016/j.rsase.2022.100703>.
- [4] Y. Kang, E. Jang, J. Im, and C. Kwon, "A deep learning model using geostationary satellite data for forest fire detection with reduced detection latency," *GIScience & Remote Sensing*, vol. 59, no. 1, pp. 2019–2035, 2022, doi: <https://doi.org/10.1080/15481603.2022.2143872>.
- [5] S. Wang, K. Guan, C. Zhang, Q. Zhou, S. Wang, X. Wu, and Z. Ma, "Cross-scale sensing of field-level crop residue cover: Integrating field photos, airborne hyperspectral imaging, and satellite data," *Remote Sensing of Environment*, vol. 285, p. 113366, 2023, doi: <https://doi.org/10.1016/j.rse.2022.113366>.
- [6] L. Salgueiro, J. Marcello, and V. Vilaplana, "Single-image super-resolution of sentinel-2 low resolution bands with residual dense convolutional neural networks," *Remote Sensing*, vol. 13, no. 24, p. 5007, 2021, doi: <https://doi.org/10.3390/rs13245007>.

- [7] W. Yang, X. Zhang, Y. Tian, W. Wang, J. H. Xue, and Q. Liao, "Deep learning for single image super-resolution: A brief review," *IEEE Transactions on Multimedia*, vol. 21, no. 12, pp. 3106–3121, 2019, doi: <https://doi.org/10.1109/TMM.2019.2919431>.
- [8] K. Karwowska and D. Wierzbicki, "Using super-resolution algorithms for small satellite imagery: A systematic review," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 3292–3312, 2022, doi: <https://doi.org/10.1109/JSTARS.2022.3167646>.
- [9] G. Vivone, L. J. Deng, S. Deng, D. Hong, M. Jiang, C. Li, and A. Plaza, "Deep learning in remote sensing image fusion: Methods, protocols, data, and future perspectives," *IEEE Geoscience and Remote Sensing Magazine*, vol. 13, no. 1, pp. 269–310, 2025, doi: <https://doi.org/10.1109/MGRS.2024.3495516>.
- [10] C. Lanaras, J. Bioucas-Dias, S. Galliani, E. Baltsavias, and K. Schindler, "Super-resolution of sentinel-2 images: Learning a globally applicable deep neural network," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 146, pp. 305–319, 2018, doi: <https://doi.org/10.1016/j.isprsjprs.2018.09.018>.
- [11] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *European Conference on Computer Vision (ECCV)*. Springer, 2016, pp. 391–407, doi: https://doi.org/10.1007/978-3-319-46475-6_25.
- [12] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017, pp. 136–144, doi: <https://doi.org/10.48550/arXiv.1707.02921>.
- [13] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 286–301, doi: <https://doi.org/10.48550/arXiv.1807.02758>.
- [14] J. Liang, H. Zeng, and L. Zhang, "Efficient and degradation-adaptive network for real-world image super-resolution," in *European Conference on Computer Vision (ECCV)*. Springer, 2022, pp. 574–591, doi: https://doi.org/10.1007/978-3-031-19797-0_33.
- [15] F. Pineda, V. Ayma, and C. Beltran, "A generative adversarial network approach for super-resolution of sentinel-2 satellite images," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLIII-B1-2020, pp. 9–14, 2020, doi: <https://doi.org/10.5194/isprs-archives-XLIII-B1-2020-9-2020>.
- [16] G. Vivone, L.-J. Deng, S. Deng, D. Hong, M. Jiang, C. Li, W. Li, H. Shen, X. Wu, J.-L. Xiao et al., "Deep learning in remote sensing image fusion: Methods, protocols, data, and future perspectives," *IEEE geoscience and remote sensing magazine*, vol. 13, no. 1, pp. 269–310, 2024.
- [17] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, p. 105, doi: <https://doi.org/10.48550/arXiv.1609.04802>.
- [18] H. Chen, X. He, L. Qing, Y. Wu, C. Ren, R. E. Sherif, and C. Zhu, "Real-world single image super-resolution: A brief review," *Information Fusion*, vol. 79, pp. 124–145, 2022, doi: <https://doi.org/10.1016/j.inffus.2021.09.005>.
- [19] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. Change Loy, "Esrgan: Enhanced super-resolution generative adversarial networks," in *Proceedings of the European conference on computer vision (ECCV) workshops*, 2018, pp. 0–0, doi: <https://doi.org/10.48550/arXiv.1809.00219>.
- [20] X. Wang, L. Xie, C. Dong, and Y. Shan, "Real-esrgan: Training real-world blind super-resolution with pure synthetic data," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021, pp. 1905–1914, doi: <https://doi.org/10.48550/arXiv.2107.10833>.
- [21] A. F. Habib, E.-M. Kim, and C.-J. Kim, "New methodologies for true orthophoto generation," *Photogrammetric Engineering & Remote Sensing*, vol. 73, no. 1, pp. 25–36, 2007, doi: <https://doi.org/10.14358/PERS.73.1.25>.
- [22] Y. Hu, D. Stanley, and Y. Xin, "True ortho generation of urban area using high resolution aerial photos," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 3, pp. 3–10, 2016, doi: <https://doi.org/10.5194/isprs-annals-III-4-3-2016>.
- [23] L. Garcia Diaz, "Pueblos abandonados de cataluña. análisis espacial de su dinámica territorial y de los cambios en sus edificaciones," *Bellaterra: Universitat Autònoma de Barcelona*, 2023. [Online]. Available: <https://ddd.uab.cat/record/283023>
- [24] Institut Cartogràfic i Geològic de Catalunya (ICGC), "Ortofoto 25 cm v7r0 color," *Infraestructura de Dades Espacials de Catalunya (IDE)*, n.d., disponible en línea. [Online]. Available: <https://ddd.uab.cat/record/283023>
- [25] M. Arabbboev, S. Begmatov, M. Rikhsivoev, K. Nosirov, and S. Saydiakbarov, "A comprehensive review of image super-resolution metrics: classical and ai-based approaches," *Acta IMEKO*, vol. 13, no. 1, pp. 1–8, 2024, doi: <https://doi.org/10.21014/actaimeko.v13i1.1679>.
- [26] M. Hassan and C. Bhagvati, "Structural similarity measure for color images," *International Journal of Computer Applications*, vol. 43, no. 14, pp. 7–12, 2012, doi: <https://doi.org/10.5120/6169-8590>.
- [27] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *The thirty-seventh asilomar conference on signals, systems & computers*, 2003, vol. 2. Ieee, 2003, pp. 1398–1402, doi: <https://doi.org/10.1109/ACSSC.2003.1292216>.
- [28] H. Park and S. Park, "Improving monocular depth estimation with learned perceptual image patch similarity-based image reconstruction and left-right difference image constraints," *Electronics*, vol. 12, no. 17, p. 3730, 2023, doi: <https://doi.org/10.3390/electronics12173730>.
- [29] G. Sharma, W. Wu, and E. N. Dalal, "The ciede2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations," *Color Research & Application: Endorsed by Inter-Society Color Council, The Colour Group (Great Britain), Canadian Society for Color, Color Science Association of Japan, Dutch Society for the Study of Color, The Swedish Colour Centre Foundation, Colour Society of Australia, Centre Français de la Couleur*, vol. 30, no. 1, pp. 21–30, 2005, doi: <https://doi.org/10.1002/col.20070>.
- [30] P. Nandal, S. Pahal, A. Khanna, and P. Rogério Pinheiro, "Super-resolution of medical images using real esrgan," *IEEE Access*, vol. 12, pp. 176 155–176 170, 2024, doi: <https://doi.org/10.1109/ACCESS.2024.3497002>.
- [31] Z. Zhu, Y. Lei, Y. Qin, C. Zhu, and Y. Zhu, "Ire: Improved image super-resolution based on real-esrgan," *IEEE Access*, vol. 11, pp. 45 334–45 348, 2023, doi: <https://doi.org/10.1109/ACCESS.2023.3256086>.
- [32] R. Berral-Soler, F. J. Madrid-Cuevas, S. Ventura, R. Muñoz-Salinas, and M. J. Marín-Jiménez, "A comparison of neural network-based super-resolution models on 3d rendered images," in *International Conference on Computer Analysis of Images and Patterns*. Springer, 2023, pp. 45–55, doi: https://doi.org/10.1007/978-3-031-44237-7_5.
- [33] L. M. Hopkins, W.-K. Wong, H. Kerner, F. Li, and R. A. Hutchinson, "Data augmentation approaches for satellite imagery," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, no. 27, 2025, pp. 28 097–28 105, doi: <https://doi.org/10.1609/aaai.v39i27.35028>.
- [34] N. Ravi, V. Gabeur, Y.-T. Hu, R. Hu, C. Ryali, T. Ma, H. Khedr, R. Rädle, C. Rolland, L. Gustafson, E. Mintun, J. Pan, K. V. Alwala, N. Carion, C.-Y. Wu, R. Girshick, P. Dollár, and C. Feichtenhofer, "Sam 2: Segment anything in images and videos," *arXiv preprint arXiv:2408.00714*, 2024, doi: <https://doi.org/10.48550/arXiv.2408.00714>.
- [35] S. Ren, F. Luzzi, S. Lahrichi, K. Kassaw, L. M. Collins, K. Bradbury, and J. M. Malof, "Segment anything, from space?" in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2024, pp. 8355–8365, doi: <https://doi.org/10.48550/arXiv.2304.13000>.

Magda Alexandra Trujillo-Jiménez received the Ph.D. degree in Engineering. She is a specialist in 2D and 3D image processing, computer vision, and Deep Learning. Her research focuses on low-cost 3D anthropometric reconstruction using geometric morphometrics and Artificial Intelligence. She develops high-performance AI methods for image-based modeling, super-resolution, and spatial analysis, with applications in remote sensing and biomedical imaging.



Francisco Ramiro Iaconis received the Ph.D. degree in Physics. He is a technical lead for innovative projects and a consultant in research and development. His expertise includes data analysis, machine learning, and Artificial Intelligence applications in neuroscience and image processing.





Debora Pollicelli received the Ph.D. degree in Engineering. She specializes in Artificial Intelligence-driven research and development, with a focus on image and data processing applied to biology, health, and environmental sciences. She has contributed to interdisciplinary projects involving biomedical imaging, environmental monitoring, and intelligent data-driven systems.



Gisela Noelia Revollo Sarmiento received the Ph.D. degree in Engineering. She is a researcher specialized in intelligent processing of satellite imagery, remote sensing, and geomorphological modeling of coastal and intertidal environments. Her doctoral research focused on computer vision techniques for the automatic classification of geographic features and the analysis of remote sensing data.



Claudio Delrieux received the Ph.D. degree in Computer Science. He is a specialist in image processing and Artificial Intelligence, with more than 20 years of experience in research and academic leadership. His work focuses on computer vision, machine learning, and advanced image analysis, with applications across scientific and technological domains. He has led and collaborated in numerous research and development projects, contributing to the advancement of intelligent imaging systems and data-driven methodologies. His interests include

pattern recognition, high-performance computing, and the integration of AI techniques into real-world problem solving.