Natural Computing
https://doi.org/10.1007/s11047-025-10030-z

Improving 5 G base station placement through precise rooftop detection using super-resolution diffusion models and satellite image analysis

Iván García-Aguilar[1,2,3] · Jesús Galeano-Brajones[3] · Francisco Luna-Valero[1] · Javier Carmona-Murillo[3] · Jose David Fernández-Rodríguez[1,2] · Rafael Marcos Luque-Baena[1,2]

Accepted: 25 May 2025
© The Author(s) 2025

Abstract
The accurate deployment of 5 G base stations (BSs) in urban environments is essential for achieving optimal network performance. In these scenarios, the most common positions for installing BSs are rooftops, which, however, given the complex topography and diverse building structures, present significant challenges when identifying suitable locations. This paper proposes an enhanced method for rooftop detection, integrating diffusion models based on super-resolution with segmentation using convolutional neural networks. Starting from the input image, a super-resolution model is applied to generate sliding windows on which re-inference is performed, thereby improving both the resolution and prediction accuracy for this type of object. By refining these detections, the placement of 5 G base stations is undertaken in a practical, industrial way, thus allowing network operators to perform a more real-world network optimization. The results demonstrate a significant improvement in detection accuracy, directly contributing to more efficient 5 G base station deployment in densely populated urban areas. This methodology offers a scalable, adaptable, and effective solution based on the context of the images it applies to.

Keywords 5 G base station deployment · Super-resolution diffusion models · Rooftop detection · Convolutional neural networks

✉ Iván García-Aguilar
ivangarcia@uma.es

Jesús Galeano-Brajones
jgaleanobra@unex.es

Francisco Luna-Valero
flv@lcc.uma.es

Javier Carmona-Murillo
jcarmur@unex.es

Jose David Fernández-Rodríguez
josedavid@uma.es

Rafael Marcos Luque-Baena
rmluque@uma.es

[1] ITIS Software, University of Málaga, C/ Arquitecto Francisco Peñalosa, 18, 29010 Málaga, Spain

[2] Biomedical Research Institute of Málaga (IBIMA), C/ Doctor Miguel Díaz Recio, 28, 29010 Málaga, Spain

[3] Merida University Center, Department of Computing and Telematics Engineering, C/ Sta. Teresa Jornet, 38, 06800 Mérida, Spain

1 Introduction

The deployment of 5 G networks in urban areas has become essential for meeting the growing demand for high-speed, low-latency wireless communications. This technology promises significant improvements over previous generations (Sindhushree and Naik 2023; Deepender et al. 2021), primarily by offering greater network capacity, lower latency, and the ability to connect a more significant number of devices simultaneously. However, one of the main challenges in this field lies in properly deploying base stations in usable locations that ensure efficient and continuous network coverage, especially in densely populated urban environments where automated analysis becomes more complex.

Urban areas present a complex topology, which, combined with the variety of architectural structures, complicates the identification of suitable sites for base station installation. Among these structures, rooftops are generally

Published online: 18 June 2025

the most viable locations for antenna deployment, as they provide a better line of sight and minimize interference caused by buildings. However, accurately detecting these elements is challenging due to the variability in the shapes, sizes, and orientations of buildings and the limited resolution of images, making this a critical issue to address.

It should be noted that the present work represents an extension of the study presented at IWINAC2024 – the 10th International Conference on the Interplay between Natural and Artificial Computation, in the Special Session on Machine Learning in Computer Vision and Robotics (MLCVR). In this extended version, significant enhancements have been incorporated, including the integration of super-resolution techniques and an improved sliding window approach to achieve more accurate rooftop detection.

The proposed methodology introduces an innovative approach to rooftop detection by integrating a pre-trained model with super-resolution techniques. The methodology is based on generating sliding windows over the input image, applying a diffusion model to improve its resolution, and enabling more precise inference in specific regions. This process significantly enhances prediction quality, particularly in areas where low resolution has traditionally hindered precise segmentation. This approach improves rooftop detection and provides a tool adaptable to different types of urban environments, facilitating automation in tasks related to infrastructure analysis and the deployment of new technologies.

The remainder of this article is organized as follows: Section 2 presents the state of the art and related work on the proposed methodology. Section 3 introduces the detailed methodology. Section 4 describes the experiments and results obtained, including the selected segmentation and super-resolution model, respectively, the dataset, the metrics used, the quantitative and qualitative outcomes, and a practical application case. Finally, Sect. 5 offers conclusions and outlines potential future research directions.

2 Related works

The deployment of base stations and the accurate detection of urban structures have become highly active research areas in recent years, particularly in infrastructure planning and 5 G network optimization. Advances in convolutional neural networks (CNNs) have led to significant progress in the precise detection and segmentation of elements in satellite and aerial images, such as rooftops and buildings, establishing a fundamental pathway for automating tasks related to 5 G network planning in complex urban environments (Zhu 2017; Audebert et al. 2018). Meanwhile, diffusion-based super-resolution models have shown promise in enhancing the quality of low-resolution images, presenting a viable approach to improve accuracy in urban environments where rooftop detection remains challenging (Xiao 2024; Luo et al. 2024). Integrating these technologies within the context of automated 5 G station deployments raises specific challenges, particularly emphasizing the need for methodologies that provide an integrated and adaptable approach (Almutairi 2022; Aloupogianni 2024).

2.1 Convolutional neural networks (CNNs) focused on segmentation

The use of convolutional neural networks (CNNs) for detection and segmentation has revolutionized the field of element identification across various domains, including identifying buildings and other urban structures in satellite images. Models like U-Net (Ronneberger et al. 2015) have been among the most prominent in segmentation tasks, effectively capturing both fine details and general patterns and inspiring improved models such as U-Net++ (Zhou et al. 2018). Initially designed for biomedical imaging applications (Azad 2024), this model has proven effective in other problems. It demonstrates its adaptability thanks to its U-shaped architecture, which combines high- and low-resolution features to optimize accuracy in complex environments (Liu et al. 2020).

Architectures like DeepLabV3 (Chen et al. 2017) and MANet (Li 2022) have addressed the challenges associated with variations in object scale within urban images. DeepLabV3 incorporates Atrous Spatial Pyramid (ASP) modules, particularly effective in capturing multi-scale contexts, and DeepLabV3+ (Chen et al. 2018) adds a decoder module to enhance the segmentation output. This is vital in densely urbanized areas where buildings exhibit significant variation in size and shape. MANet employs attention mechanisms to enhance segmentation accuracy, proving useful in contexts with high structural variability.

In addition to the models above, various architectures such as PSPNet (Zhao et al. 2017), LinkNet (Chaurasia and Culurciello 2017), FPN (Quyen et al. 2023), and PAN (Li et al. 2018) have gained popularity in the field of image segmentation. PSPNet (Pyramid Scene Parsing Network) utilizes a scene pyramid module to capture contextual information at multiple scales, which is especially beneficial for segmenting dense urban areas with structural variability. LinkNet stands out for its lightweight structure and rapid image processing capabilities, making it ideal for detection tasks in urban environments. On the other hand, FPN (Feature Pyramid Network) employs a feature pyramid structure that facilitates the detection of objects of various sizes, optimizing segmentation accuracy for small and large elements within the same image. Finally, PAN (Path Aggregation Network) introduces a feature

aggregation approach across multiple scales, enhancing accuracy in detecting edges and contours in complex environments. These architectures extend the capabilities of CNNs by providing specific solutions for segmentation challenges in high-density urban settings.

2.2 Diffusion models for super-resolution

Super-resolution models have become essential tools for enhancing image quality across multiple fields, such as satellite and aerial imagery, particularly in contexts where low resolution limits the accuracy of element detection. Single Image Super-Resolution (SISR) has been widely adopted to increase detail levels in low-quality images, a crucial aspect in accurately identifying rooftops in urban environments (Chauhan 2023; Saharia 2023). Recently, diffusion models have emerged as state-of-the-art solutions for generating high-resolution images, preserving details essential for accurate segmentation. Methods like Denoising Diffusion Probabilistic Models (DDPM) (Ho et al. 2020) and Score-Based Generative Models have shown promising results in reconstructing low-resolution areas, generating more detailed representations (Zhu et al. 2023).

Among the most recent super-resolution methods, the ResShift model (Yue et al. 2024) has significantly improved the quality of low-resolution images, particularly in environments with high variability in fine details. This model uses a deep neural network-based approach to adjust residual features at multiple scales, allowing for greater reliability in image reconstruction.

In addition to ResShift, other diffusion models have shown great potential in the super-resolution of complex images. SR3 (Super-Resolution via Repeated Refinement) (Saharia 2023) and SR3+ (Sahak et al. 2023) are two notable examples. SR3 employs an iterative refinement approach through multiple diffusion steps, achieving high-quality detail reconstruction in low-resolution images by gradually enhancing clarity. SR3+, in turn, introduces a diffusion-based model for blind super-resolution, establishing a new state-of-the-art by combining composite, parameterized degradations for self-supervised training and noise-conditioning augmentation during training and testing. These models and ResShift illustrate the fundamental role of diffusion techniques in super-resolution, addressing quality and accuracy challenges critical for rooftop detection in complex urban environments.

2.3 Application of techniques in the context of 5 G base station deployment

The deployment of 5 G networks in urban areas presents challenges due to the dense distribution of structures, the diversity of buildings, and their variability, which significantly impact signal coverage and quality (Ahamed and Faruque 2021). In urban environments, precision in identifying rooftops and other structures is critical, as base stations must be strategically positioned to minimize interference and maximize reach. However, the topographical complexity of these areas, combined with noise in satellite images, represents a significant obstacle to automated and accurate network infrastructure planning.

Conventional network planning techniques have traditionally relied on model-based approaches, such as geometric modeling and ray tracing, to simulate signal propagation in various environments. While effective, these classical techniques have limitations in densely populated areas, as they cannot sufficiently capture the detailed layout of urban structures. Methods based on sensor networks or on-site inspection have attempted to address these limitations; however, depending on the scenario, they may lack accuracy (García-Aguilar et al. 2024).

The proposed solution addresses these challenges through advanced segmentation and super-resolution techniques based on diffusion models to detect rooftops in low-resolution images precisely. First, diffusion models focused on super-resolution, such as ResShift, are applied to improve the quality of input images, thus achieving a more detailed representation of urban structures. Next, convolutional neural networks, such as U-Net or FPN, segment areas of interest, generating accurate rooftop masks that help identify optimal locations for base stations. The proposed methodology also incorporates sliding windows on super-resolved images, allowing for detailed region-by-region analysis, enhancing inference accuracy, and identifying optimal rooftops for antenna deployment. This advanced approach optimizes base station deployment by ensuring strategic placement in urban environments and enables faster and more automated planning, reducing reliance on traditional methods that require direct inspection.

3 Methodology

The methodology consists of sequential steps for enhanced satellite image segmentation, as shown in Figure 1. The process includes initial super-resolution, patch generation via a sliding window, patch-level segmentation, recomposition of segmentation masks, weighted aggregation, and final contour extraction. Each step is detailed in the following subsections. The complete code is available at.[1]

[1] https://github.com/IvanGarcia7/Improving_5G_BS_Placement_Precise_Rooftop_Detection_SR_Diffusion_Models-and-Satellite-Image_Analysis/

3.1 Super-resolution of the input satellite image

Starting with the input satellite image I, a super-resolution is applied using the ResShift model, which enhances the resolution by a factor of $\times 4$. This process yields a higher-resolution image I_{HR}, facilitating more accurate segmentation in subsequent steps. The transformation can be mathematically described as follows:

$$I_{HR} = \text{ResShift}(I, \times 4) \tag{1}$$

After applying super-resolution to the image I_{HR}, a smoothing operation is performed to reduce noise and artifacts that may have been introduced during the enhancement process. This step improves the overall quality of the high-resolution image, facilitating more accurate segmentation results. The smoothed high-resolution image is represented mathematically as:

$$I_{HR}^{s} = \text{Smoothing}(I_{HR}) \tag{2}$$

3.2 Patch extraction using a sliding window approach

The high-resolution image I_{HR}^{s} is divided into overlapping patches using a sliding window technique. Let w x w denote the size of each window, and let s be the stride length, defining the distance between adjacent windows. For each window position (i, j), a patch $P_{i,j}$ is extracted from I_{HR}^{s} as follows:

$$P_{i,j} = I_{HR}^{s}[i \cdot s : i \cdot s + w,\ j \cdot s : j \cdot s + w] \tag{3}$$

In our implementation, the window size w is chosen based on the typical scale of the target objects within the dataset, ensuring that each patch captures sufficient contextual information for accurate segmentation. The stride s is then determined to guarantee an adequate overlap between adjacent patches–typically set to achieve an overlap of around 25–50%–which helps mitigate border effects and ensures that no critical features are missed. These parameters were initially determined through preliminary

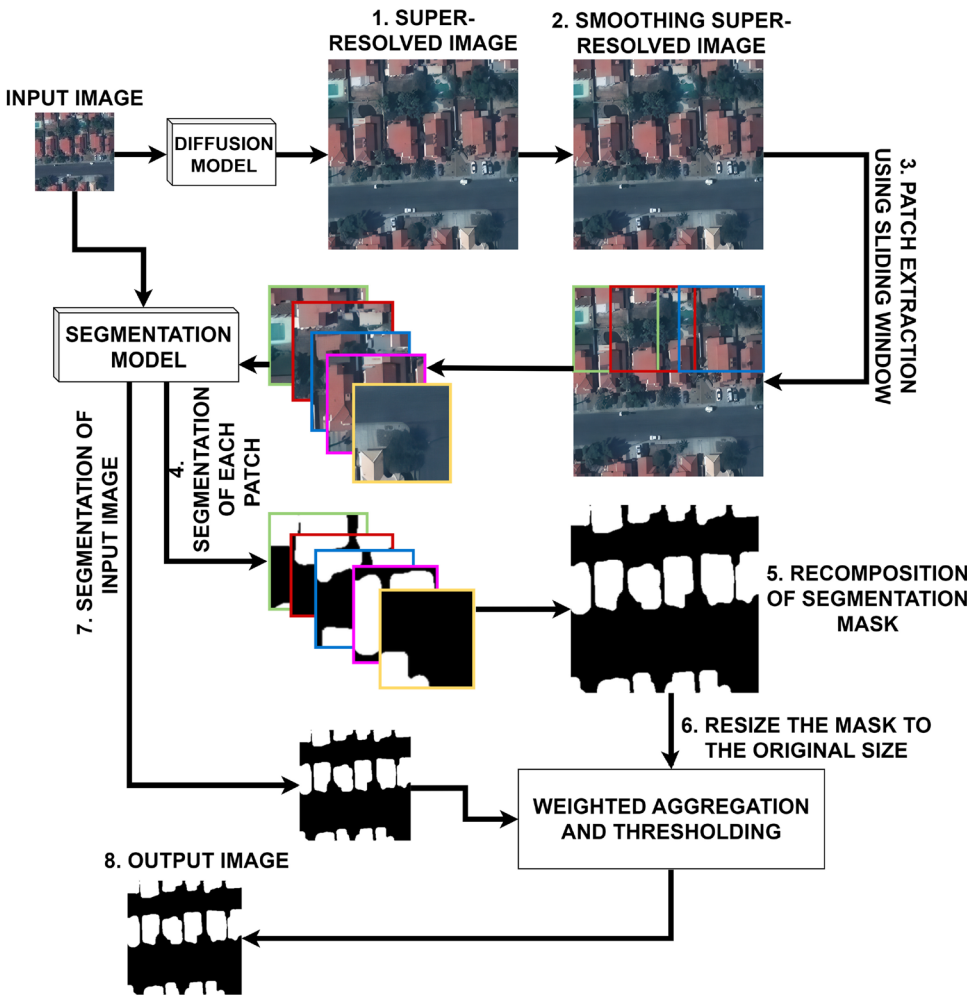

Fig. 1 Workflow of the proposed methodology

experiments, balancing detection accuracy with computational efficiency.

In cases where the window extends beyond the image boundary, reflection padding is applied, which mirrors pixels from the opposite edge of I_{HR}^{s}. The padded patch $P_{i,j}$ is defined as:

$$P_{i,j} = \begin{cases} \text{if within bounds} \\ I_{HR}^{s}[i \cdot s : i \cdot s + w,\ j \cdot s : j \cdot s + w] \\ \text{otherwise} \\ \text{reflect}\left(I_{HR}^{s}\right) \end{cases} \tag{4}$$

In addition, when processing the segmentation outputs, if the corrected coordinates (x, y) (obtained via the stride s) fall outside the x $w \times w$ boundaries of a patch, the corresponding value of the function $S_{i,j}^{\text{original}}(x, y)$ is defined to be 0. Furthermore, the overlapping segmentation outputs $S_{i,j}^{\text{original}}$ are aggregated by summing the contributions from each patch and subsequently normalizing the result to ensure that the final segmentation values lie between 0 and 1. This summation approach is preferred over a maximum function because it incorporates the contributions of all overlapping patches, resulting in a more robust segmentation outcome.

3.3 Segmentation of each patch with a pre-trained model

Each patch $P_{i,j}$ is then processed by a pre-trained segmentation model M (see Section 4.1), producing a segmentation mask $S_{i,j}$ for each patch:

$$S_{i,j} = M(P_{i,j}) \tag{5}$$

In this expression, $S_{i,j}$ denotes the segmented output of the patch $P_{i,j}$, which contains pixel-level classifications for the targeted objects. The segmentation output $S_{i,j}$ is a binary mask in which a value of 1 indicates that a roof has been detected, while a value of 0 denotes the absence of a roof.

3.4 Recomposition of segmentation masks: offset removal and rescaling

The segmented patches $S_{i,j}$ are recombined to reconstruct the segmentation for the complete high-resolution image. The scaling introduced by the super-resolution step is reversed by downsampling each patch by a factor of 4. In our implementation, the downsampling factor u is set to 4, so that the rescaled segmentation $S_{i,j}^{\text{original}}$ is given by:

$$S_{i,j}^{\text{original}} = \text{downscale}(S_{i,j}, \frac{1}{u}) \tag{6}$$

The complete segmented image $S_{\text{final}}(x, y)$ is constructed by summing each patch $S_{i,j}^{\text{original}}$ at its respective offset:

$$S_{\text{final}}(x, y) = \sum_{i,j} S_{i,j}^{\text{original}}(x - i \cdot s, y - j \cdot s) \tag{7}$$

In our implementation, the segmentation outputs $S_{i,j}^{\text{original}}(x, y)$ are normalized to lie between 0 and 1, ensuring that the recomposed segmentation $S_{\text{final}}(x, y)$ is also bounded within this range. This operation aligns the masks by reversing the offset applied during patch extraction, achieving a coherent global segmentation.

3.5 Weighted aggregation of masks

A weighted scoring scheme is introduced to balance contributions between global and local segmentations. Each pixel in the full image segmentation $S_{\text{full}}(x, y)$ receives a score of 3, while pixels detected within individual patch segmentations $S_{i,j}^{\text{original}}(x, y)$ are assigned a score of 1. This weighting reflects the importance of global context in achieving accurate segmentation. The model performs better when inferring from the complete image due to the broader visual context available, which helps recognize relationships and structures across the entire scene. This greater context reduces errors that might arise from limited information.

In contrast, patches $S_{i,j}^{\text{original}}(x, y)$, while valuable for detecting finer details, operate with a restricted field of view. This limited context can sometimes lead to errors, especially when the structures, such as rooftops, are larger or more complex than the model encountered during training. The reduced weight of 1 for patch segmentations minimizes the impact of such potential inaccuracies.

The total score at each pixel location (x, y) is given by:

$$W(x, y) = 3 \cdot S_{\text{full}}(x, y) + \sum_{i,j} S_{i,j}^{\text{original}}(x, y) \tag{8}$$

$S_{\text{full}}(x, y)$ denotes the global segmentation output of the complete image and is also a normalized mask with values between 0 and 1. Moreover, since both $S_{\text{full}}(x, y)$ and $S_{i,j}^{\text{original}}(x, y)$ are normalized to the range [0, 1], the weighted score $W(x, y)$ is theoretically bounded between 0 and a value that depends on the number of overlapping patches. In our experiments, the effective range of $W(x, y)$ was observed to be approximately [0, L]. We performed empirical cross-validation to set the threshold T for binarization, and T was chosen accordingly to maximize segmentation performance.

The assignment of a weight of 3 to the global segmentation $S_{\text{full}}(x, y)$ and 1 to the patch-based segmentations $S_{i,j}^{\text{original}}(x, y)$ can be quantitatively justified by considering the relative confidence and accuracy levels typically observed in these two approaches. Empirical analysis during model validation indicated that full-image

segmentations consistently demonstrated an approximately threefold increase in accuracy compared to individual patch segmentations. This increase can be attributed to the complete visual context available in global segmentation, which reduces ambiguity and enhances the model's ability to segment larger structures correctly.

By assigning a weight of 3, the contribution of the global segmentation is effectively scaled to reflect its higher confidence level and accuracy. This weighting ensures that the final combined score prioritizes the more reliable global detections while still integrating the local patch-based segmentation to enhance detail. The choice of the weight ratio (3:1) was determined through cross-validation experiments, where this configuration maximized the overall performance metrics such as pixel accuracy and intersection over union (abbreviated as IoU, the ratio of pixels correctly detected as part of the ground truth to all pixels either detected or in the ground truth), demonstrating that it provides an optimal balance between leveraging the broad context of global inference and the granularity of local detections.

3.6 Thresholding and contour extraction

A threshold T is applied to the weighted mask $W(x, y)$ to refine the segmentation, creating a binary mask $B(x, y)$ that identifies segmented regions. Pixels with scores above T are set to 1, while all others are set to 0:

$$B(x,y) = \begin{cases} 1 & \text{if } W(x,y) \geq T \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

The threshold T was determined empirically through cross-validation on a validation dataset to optimize segmentation performance. Contour extraction is then applied to $B(x, y)$ to identify the boundaries of each detected object. This process yields a set of contours C representing the perimeters of segmented regions:

$$C = \text{ContourDetection}(B) \tag{10}$$

The function ContourDetection is implemented using the OpenCV `findContours` algorithm, which is widely adopted for extracting object boundaries. These contours serve as the final delineation of each detected object within the high-resolution image, offering a precise outline of segmented regions.

It is worth noting that the proposed approach is a meta-model based on re-inference through patches on super-resolved images using diffusion models to verify that this method indeed aids the model in achieving higher accuracy. In this context, the weighted aggregation of segmentation masks, thresholding, and contour extraction constitute the specific contributions of our meta-model. The underlying segmentation model and diffusion network used for super-resolution are considered part of the broader application framework in which our meta-model operates.

4 Experiments

This section presents the experiments conducted to evaluate the proposed methodology. The focus is on the segmentation models, the selected super-resolution model, and the dataset employed for testing. Additionally, the results obtained are discussed, providing insights into the performance and effectiveness of the approach.

4.1 Segmentation models

This work leverages a selection of models chosen for their diverse architectures and established performance in image segmentation tasks to achieve accurate segmentation. Below is a summary of each model used in this study:

- **UNet**: Recognized for its symmetric encoder-decoder structure, UNet is extensively applied in biomedical and satellite image segmentation. It employs skip connections to enhance segmentation accuracy by retaining spatial context.
- **UNet++**: This variant improves skip connections and introduces dense, nested connections, facilitating more effective multi-scale feature fusion.
- **MAnet**: By incorporating attention mechanisms, MAnet focuses on relevant spatial information, which is advantageous for segmenting specific priority regions within images.
- **LinkNet**: Featuring an encoder-decoder architecture with residual blocks, LinkNet enables efficient segmentation with lower computational demands, making it suitable for applications requiring rapid inference.
- **FPN (Feature Pyramid Network)**: Designed with a pyramid structure, FPN enhances multi-scale feature extraction, thereby supporting object detection across various resolutions.
- **PSPNet (Pyramid Scene Parsing Network)**: Through pyramid pooling, PSPNet captures contextual information at multiple scales, supporting analysis in complex scenes containing diverse objects.
- **PAN (Pyramid Attention Network)**: PAN achieves high-resolution segmentation with optimized efficiency by combining a feature pyramid architecture with spatial attention mechanisms.
- **DeepLabV3**: Leveraging atrous (dilated) convolutions, DeepLabV3 captures multi-scale context, facilitating the segmentation of objects with varied shapes and sizes.

- **DeepLabV3+**: Extending DeepLabV3, this model integrates atrous spatial pyramid pooling (ASPP) with a decoder module to improve boundary delineation, delivering precise segmentation in complex images.

Each of these models brings distinct advantages to the segmentation task, contributing to a comprehensive analysis of satellite imagery.

4.2 Super-resolution model

ResShift is a diffusion-based super-resolution model that enhances image resolution through iterative detail refinement. Its diffusion mechanism allows for restoring high-quality details while minimizing artifacts, a key advantage for satellite images where structural clarity of objects such as rooftops is essential. Additionally, the ResShift noise reduction approach enhances clarity in satellite imagery, effectively preserving edges and contrasts, which is crucial for accurate rooftop interpretation in urban areas.

ResShift's adaptability to various scales is highly advantageous for applications focused on rooftop resolution enhancement, as satellite images often include rooftops of different sizes, shapes, and orientations. The model's strong texture and structure enhancement capabilities reveal realistic details, making it particularly effective for tasks such as urban planning. Combined with its emphasis on maintaining texture fidelity, these features make ResShift an ideal choice for improving rooftop visibility and clarity in satellite images.

As shown in Figure 2, the left side displays a section of the original image, while the right side presents the same area after applying super-resolution. Applying this technique significantly enhances the level of detail, providing more precise features and improved visualization. The selection of the ResShift super-resolution model in the proposed methodology is non-exclusive. Other super-resolution models may be incorporated if project requirements or image characteristics require it. In our experiments, the patch size w was set to 960 pixels and the stride s to 480 pixels. Furthermore, the threshold T used for binarization in Sect. 3.6 was set to 0.5 based on cross-validation results.

4.3 Datasets

For this project, two datasets have been selected to support segmentation tasks and enhance image resolution models, ensuring robustness and adaptability across different satellite image contexts.

The first dataset, sourced from Kaggle's Synthetic Word OCR Dataset, entitled Mapping Challenge (Mohanty 2020) comprises labeled images designed for segmentation tasks, primarily focused on object detection. This extensive dataset offers a large sample size to support optimal model training and validation for this project. Originally posted on the crowdAI platform as part of a competition, this dataset includes RGB satellite images annotated with building locations, facilitating the development of models capable of detecting structures within new images. It has been divided into training, testing, and validation sets to implement this dataset effectively. Using the K-Fold cross-validation technique, the data is split with 75% allocated for training and 25% for testing in each experiment, while 20% of the training set is reserved for validation. The training set contains 280,741 satellite images, each sized at 300x300 pixels in RGB format, with annotations in MS-COCO format, accessible via a JSON file. These annotations highlight segmentation details within each image.

The second dataset, Satellite Dataset I (Global Cities) (Ji et al. 2019), is also utilized to provide a comprehensive testing ground for model performance in diverse geographic contexts. This dataset includes high-resolution satellite images of buildings across major cities worldwide, allowing the model to be evaluated on diverse urban layouts and building structures. This additional dataset aids in assessing the model's generalization capabilities and ensures accuracy across a variety of urban landscapes.

Fig. 2 Comparison between a normal image section (left) and a super-resolved image section (right)

4.4 Metrics

For the evaluation of the models in this project, several metrics have been employed to ensure a comprehensive assessment of performance. The primary metric utilized is the Mean Average Precision (mAP), calculated using the COCO evaluation framework (Everingham et al. 2010). This metric provides a robust measure of model performance by evaluating precision and recall across different thresholds for Intersection over Union (IoU) (Padilla et al. 2020), allowing for quantifying how well the model detects and segments objects within images.

In addition to mAP, the Dice coefficient is employed to further assess model accuracy, which is particularly relevant for segmentation tasks and quantifies the overlap between the predicted and ground truth segmentation masks. This metric ranges from 0 to 1, with values closer to 1 indicating better agreement between the model's predictions and the actual data.

4.5 Results

This section comprehensively analyzes the experimental results of applying various models and strategies to the specified datasets. The results are structured to facilitate an understanding of the models' performance differences and the proposed methods' effectiveness. In particular, three variants have been evaluated:

- RAW: This variant corresponds to performing inference directly on the original image without any additional preprocessing or modification.
- RAW Composite: In this approach, the original image is cropped into sub-images, each of which is processed separately. The segmentation results from these sub-images are then recomposed to form a complete segmentation map.
- Ours: This variant represents the proposed methodology detailed in this article. It integrates a super-resolution step, a sliding window patch generation process, and refined inference, resulting in improved segmentation performance compared to the other methods.

The following subsections provide a detailed explanation of the performance differences among these variants and demonstrate the effectiveness of the proposed method.

4.5.1 Comparison of mean average precision (mAP)

Table 1 illustrates the Mean Average Precision (mAP) across different models, using the first dataset (Mohanty 2020) post-split into training, validation, and testing sets. The mAP is a crucial metric for evaluating the performance of object detection models, providing insight into their accuracy across multiple Intersection over Union (IoU) thresholds. The results indicate that the proposed strategy (labeled as OURS) consistently outperforms the raw inference method (labeled as RAW) across most models.

Table 1 Mean Average Precision (mAP) comparison across different models using the Mapping Challenge Dataset (Mohanty 2020). The best results are highlighted in **bold**

Model	Strategy	AP@[0.50:0.95] (all)	AP@[0.50] (all)	AP@[0.75] (all)
UNET	RAW	0.357	0.594	0.398
	OURS	**0.376**	**0.632**	**0.419**
UNET++	RAW	0.362	0.599	0.405
	OURS	**0.392**	**0.648**	**0.442**
MAnet	RAW	0.357	0.598	0.394
	OURS	**0.375**	**0.631**	**0.413**
Linknet	RAW	0.301	0.505	0.340
	OURS	**0.385**	**0.648**	**0.432**
FPN	RAW	0.363	0.616	0.399
	OURS	**0.389**	**0.663**	**0.429**
PSPNet	RAW	0.306	0.547	0.324
	OURS	**0.324**	**0.578**	**0.344**
PAN	RAW	0.297	0.537	0.315
	OURS	**0.319**	**0.584**	**0.335**
DeepLabV3	RAW	0.350	0.593	**0.389**
	OURS	**0.350**	**0.606**	0.383
DeepLabV3+	RAW	0.346	0.594	0.384
	OURS	**0.374**	**0.638**	**0.416**

For example, in the UNET model, the mAP@[0.50:0.95] (that is, the mean of average precisions for IoU thresholds from 0.5 to 0.95, as computed using the COCO framework (Padilla et al. 2020)) increased from 0.357 to 0.376, demonstrating a notable improvement. Similar trends were observed in other architectures, such as UNET++ and MAnet, where the proposed method significantly enhanced the precision values.

Particularly noteworthy is the performance of the Linknet model, where the mAP@[0.50:0.95] exhibited an increase from 0.301 in the RAW strategy to 0.385 with the OURS strategy. This improvement suggests that the sliding window approach combined with super-resolution and mask reconstruction contributes to a more refined detection capability, even in models that initially displayed lower performance. Moreover, the FPN model also showed substantial enhancements, with a rise in mAP@[0.50:0.95] from 0.363 to 0.389, further supporting the effectiveness of the proposed method. Conversely, the PSPNet and PAN models demonstrated less improvement, highlighting that specific architectures may benefit more from the proposed enhancements than others.

The results across the table underscore the overall trend that integrating advanced techniques such as super-resolution and structured reconstruction significantly improves model performance, as reflected in the higher precision scores across all evaluated metrics. This reinforces the importance of model architecture selection and method application in achieving optimal detection results. It is important to note that the RAW COMPOSITE strategy was not evaluated in this study. This decision was based on the initial size of the images, which was already suitable for direct input into the models. Consequently, generating patches directly from the original images was not feasible, as the dimensions needed to allow for effective segmentation through this method. This limitation emphasizes the need to carefully consider image dimensions when selecting preprocessing strategies for different model architectures.

4.5.2 Evaluation of DICE score

Table 2 presents a comparative analysis of strategies using DICE scores for the second dataset (Ji et al. 2019), as COCO-format annotations were unavailable, making this alternative metric necessary (but please note that the DICE coefficient is very similar to the IoU (Nazzal et al. 2024)). The evaluation categorizes each model's performance into three strategies: RAW SIMPLE, RAW COMPOSITE, and OURS.

The results reveal that the OURS strategy consistently outperforms both RAW strategies across most models. For example, in the UNET model, the DICE score improved from 0.48 in the RAW SIMPLE method to 0.60 in OURS, signifying a substantial enhancement in segmentation accuracy. In the case of UNET++, the performance gap was also pronounced, with a DICE score of 0.36 for RAW SIMPLE advancing to 0.54 in OURS. This indicates that incorporating the proposed strategy yields significant gains in segmentation effectiveness, particularly for architectures that traditionally struggle with raw inference methods.

Table 2 Comparison of Strategies: DICE Score using the Satellite Dataset I (Global Cities) (Ji et al. 2019). The best results are highlighted in **bold**

	Strategy		
Model	RAW SIMPLE	RAW COMPOSITE	OURS
Unet	0.48	0.50	**0.60**
Unet++	0.36	0.37	**0.54**
MAnet	0.45	0.48	**0.51**
Linknet	0.40	0.42	**0.60**
FPN	0.43	0.46	**0.49**
PSPNet	0.37	0.41	**0.56**
PAN	0.51	**0.53**	0.51
DeepLabV3	0.45	**0.48**	0.44
DeepLabV3+	0.45	0.49	**0.56**

The performance of the Linknet model is particularly striking, showcasing an increase from 0.40 in RAW SIMPLE to 0.60 in OURS for DICE scores. This improvement highlights the potential of the proposed method to enhance models that rely heavily on precise segmentation capabilities. In contrast, models such as DeepLabV3 and PAN showed more modest improvements, suggesting that specific architectures may not fully leverage the benefits of the proposed techniques.

Overall, the results reaffirm the effectiveness of the OURS strategy in enhancing segmentation metrics such as DICE scores. The clear trend across multiple models indicates that the proposed enhancements contribute significantly to improved performance, thereby underscoring the relevance of innovative approaches in advancing model capabilities in image segmentation tasks.

Figure 3 compares segmentation results from different strategies. The top part features the original image with its corresponding ground truth mask, providing an apparent reference for evaluation. The lower section contrasts three methods: RAW SIMPLE, RAW COMPOSITE, and OURS. Notably, the OURS strategy successfully segments more objects than the other two methods, demonstrating enhanced accuracy and effectiveness in object detection. This improvement highlights the benefits of incorporating advanced techniques in the segmentation process.

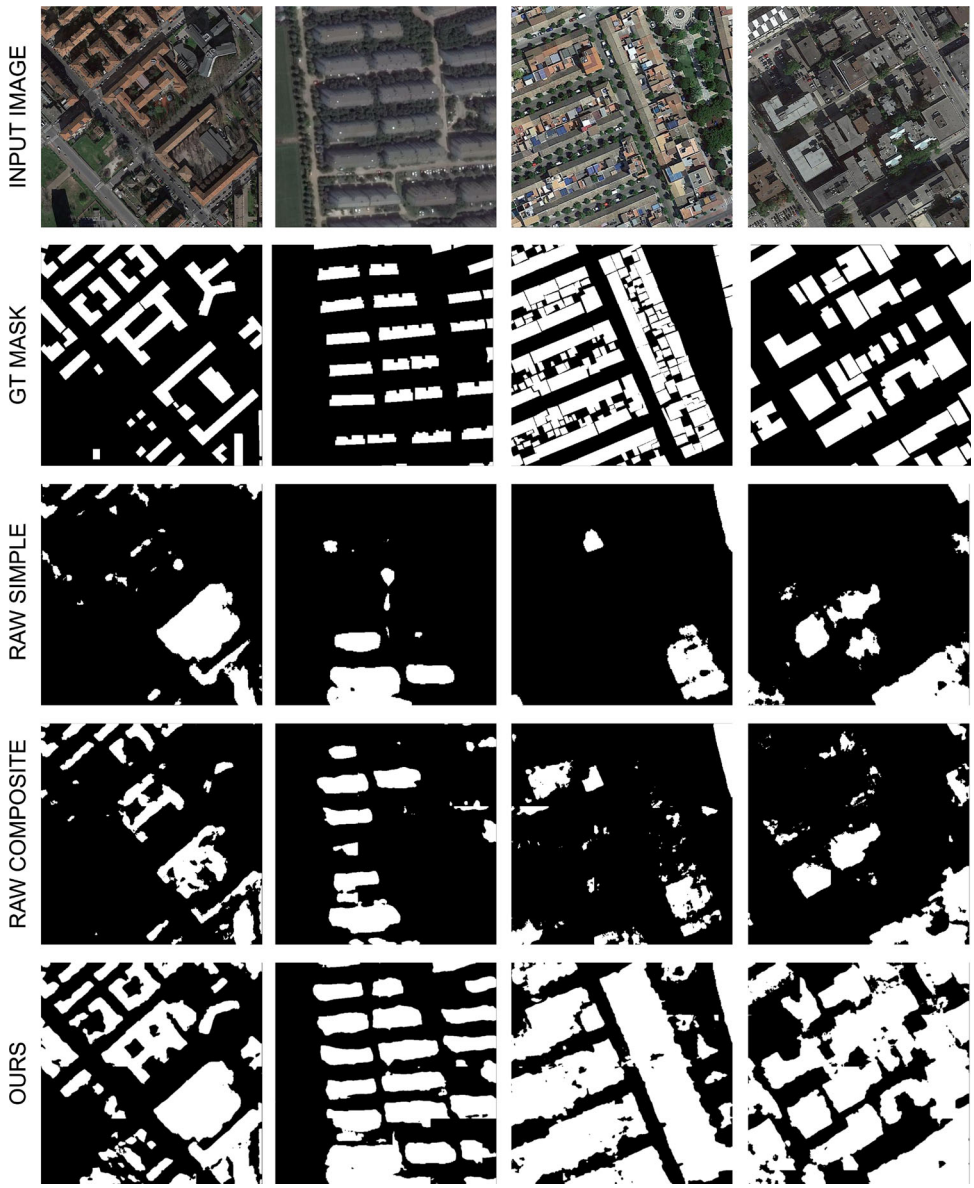

Fig. 3 The images are displayed in a vertical sequence, starting with the original image and its ground truth mask, followed by the three segmentation alternatives: RAW SIMPLE, RAW COMPOSITE, and OURS, in that order using the Unet Model

4.5.3 A practical application case: the effect of realistic deployment strategies on the cell switch-off problem

To practically test the proposed solution within a 5 G network context, we analyze its impact on the cell switch-off (CSO) problem, a multi-objective optimization problem that aims to minimize energy consumption of the network while maximizing network capacity. The goal is to deactivate a subset of deployed BSs while maintaining an adequate level of service for user equipment. This problem has been widely studied in the literature, and our approach is based on multi-objective evolutionary algorithms (MOEAs), specifically using NSGA-II (Deb et al. 2002), as presented in Galeano-Brajones et al. (2023, 2024). This algorithm efficiently explores the trade-offs between conflicting objectives, just as in the case of the CSO problem, by evolving a diverse population of candidate solutions over multiple generations. The selection process is based on non-dominated sorting, ensuring that better trade-off solutions receive higher priority, while crowding distance preservation maintains solution diversity across the Pareto front approximation.

A key aspect of solving the CSO problem is the accurate modeling of base station deployment, as the placement of these elements directly affects energy consumption and network coverage. Among the various strategies developed for realistic network deployments and traffic modeling, our approach builds upon the method introduced in Mirahsan et al. (2015). This method models traffic heterogeneity using three independent Poisson Point Processes (PPPs) for

base stations, user equipment, and social attractors (points of interest for network users). PPPs are widely used in the literature as they provide a balance between methodological simplicity and realistic spatial distributions, making them suitable for modeling the irregular placement of network infrastructure in urban environments.

To evaluate the impact of realistic deployment constraints, we compare two deployment strategies in the CSO optimization process:

- Baseline strategy (Mirahsan et al. 2015). A conventional deployment model where BSs are placed without constraints, potentially leading to unrealistic locations (e.g., roads or open areas).
- Feasibility mask strategy. A deployment model where BSs are only placed on rooftops, as determined by our segmentation model. This strategy follows the same deployment model as the baseline, but ensuring that all base stations are placed within feasible rooftop locations by resampling any generated points that fall outside identified rooftops.

Given that the effectiveness of the feasibility mask strategy depends directly on the accuracy of the segmentation model used to identify valid deployment locations, we conducted our experiments using the segmentation approach proposed in this study. According to the results obtained in this work, our segmentation model provides the most precise identification of rooftops, ensuring that the feasibility masks used in the optimization process closely match real-world deployment constraints. This choice guarantees that the comparison between the baseline and feasibility mask strategies is not biased by segmentation inaccuracies, which allows us to better isolate the impact of realistic deployment constraints on the optimization process.

Figure 4 compares the Pareto fronts obtained for both strategies. The blue triangles represent the feasibility mask strategy, ensuring base stations are only deployed in valid rooftop locations. The orange circles correspond to the baseline strategy, where no feasibility constraints are applied. Since the CSO problem is multi-objective, both strategies approximate the Pareto front, balancing the trade-offs between reducing energy consumption and maintaining network capacity. To ensure statistical robustness, we performed 10 independent executions of the optimization process. The reported front corresponds to the 50% attainment surface, which includes all objective vectors attained in at least half of the runs. This representation, computed using the visualization tool from Knowles (2005), provides a representative and probabilistically meaningful view of the algorithm's typical behavior across multiple executions.

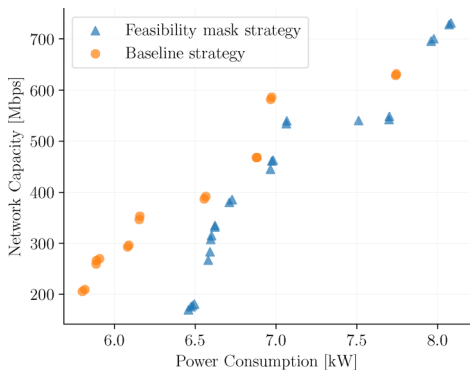

Fig. 4 Comparison between deployment strategies on the CSO problem

The results indicate that using the feasibility mask strategy leads to solutions that achieve higher network capacity, but only in the higher power consumption range. In contrast, the baseline strategy achieves a lower energy consumption, as base stations can be placed more freely, allowing greater flexibility in the CSO process. This increased flexibility enables more aggressive cell switch-off, leading to lower power consumption but at the cost of reduced network capacity, and highlights the fundamental trade-off in CSO optimization: as base stations are deactivated to reduce energy consumption, network capacity decreases since fewer active base stations mean fewer radio resources available to users. However, the baseline strategy appears more energy-efficient in the lower consumption range, suggesting that greater flexibility in base station placement allows configurations that optimize energy consumption more effectively. In contrast, the feasibility mask strategy produces higher-capacity solutions but only in the high-energy space, reinforcing the importance of considering realistic deployment constraints when optimizing 5 G networks. However, the differences between both strategies remain relatively small, suggesting that PPP-based deployment models provide a sufficiently accurate approximation of real-world base station distributions while maintaining methodological simplicity. This validates the use of PPP-based approaches for network optimization, as they capture the essential trade-offs between energy efficiency and capacity without introducing excessive complexity.

5 Conclusions and future lines

The experimental results demonstrate the significant impact of the proposed methodologies on model performance, as evidenced by the quantitative analyses presented in Tables 1 and 2. The Mean Average Precision (mAP) results indicate consistent improvements across various models, particularly integrating advanced techniques such

as super-resolution and mask reconstruction. For example, models like Linknet exhibited substantial gains in mAP, starting at 30.1% to 38.5% when transitioning from the raw inference method to the proposed strategy. This underscores the effectiveness of the proposed enhancements in enhancing detection accuracy.

Similarly, the evaluation based on the DICE score in the second dataset reaffirms these findings. The OURS strategy notably improved the performance metrics for models such as UNET and Unet++, with DICE scores rising from 0.48 to 0.60 and 0.36 to 0.54, respectively. These improvements highlight the potential of our approach to refine segmentation capabilities across different architectures. Visual enhancements are also evident in the qualitative results in Figure 3, where the proposed methods yield clearer and more accurate segmentations than traditional RAW strategies. The visual improvements align with the quantitative results, further validating the effectiveness of the implemented techniques.

Future research directions will explore the application of these methodologies to predict building heights from aerial imagery. By adapting the proposed strategies to extract features relevant to height estimation, it is anticipated that the models can be refined to offer precise predictions, thereby contributing to urban planning and development efforts. Furthermore, continued investigation into other architectural designs and data augmentation techniques could yield additional improvements and broaden the applicability of the proposed methods across various domains.

Acknowledgements This work is partially supported by the Ministry of Science and Innovation of Spain under grants PID2022-136764OA-I00, PID2023-151462OB-100, TED2021-131699B-I00 (MCIN / AEI / 10.13039 / 501100011033, FEDER) and PID2020-112545RB-C54, by the University of Málaga (Spain) under grants B1-2021_20, B4-2023_13, B1-2022_14 and by the Fundación Unicaja under project PUNI-003_2023.

Author Contributions I.G.-A. conceptualized the study, led the methodology, and conducted the main experiments. J.G.-B. and F.L.-V. developed and implemented the practical case study. J.C.-M. and J.D.F.-R. participated in data analysis and provided critical insights for interpretation. R.M.L.-B. supervised the project, contributed to revisions, and ensured the overall integrity of the manuscript. All authors contributed to the writing of the manuscript and reviewed the final version.

Funding For open access charge: Universidad de Málaga/CBUA.

Data availibility statement The first dataset used in this study, titled Mapping Challenge, was sourced from Kaggle's Synthetic Word OCR Dataset and is publicly available at https://www.kaggle.com/datasets/kmader/synthetic-word-ocr. This dataset has been referenced from the work of Mohanty et al. (2020) (Deep Learning for Understanding Satellite Imagery: An Experimental Survey). The second dataset, titled Satellite Dataset I (Global Cities), is available at http://gpcv.whu.edu.cn/data/building_dataset.html and was detailed in the publication by Ji et al. (2019) (Fully Convolutional Networks for Multisource Building Extraction From an Open Aerial and Satellite Imagery Data Set), published in IEEE Transactions on Geoscience and Remote Sensing.

Declarations

Conflict of interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Ahamed MM, Faruque S (2021) 5g network coverage planning and analysis of the deployment challenges. Sensors 21:6608. https://doi.org/10.3390/s21196608

Almutairi MS (2022) Deep learning-based solutions for 5g network and 5g-enabled internet of vehicles: advances, meta-data analysis, and future direction. Math Probl Eng 2022:1–27. https://doi.org/10.1155/2022/6855435

Aloupogianni E et al (2024) Ai-driven optimization of small cell deployment for beyond 5g networks. Procedia Computer Science 238:908–913. The 15th International Conference on Ambient Systems, Networks and Technologies Networks (ANT) / The 7th International Conference on Emerging Data and Industry 4.0 (EDI40), April 23-25, 2024, Hasselt University, Belgium https://www.sciencedirect.com/science/article/pii/S1877050924013474

Audebert N, Le Saux B, Lefèvre S (2018) Beyond rgb: very high resolution urban remote sensing with multimodal deep networks. ISPRS J Photogramm Remote Sens 140:20–32 (**Geospatial Computer Vision**)

Azad R et al (2024) Medical image segmentation review: the success of u-net. IEEE Trans Pattern Anal Mach Intell 1–20

Chauhan K et al (2023) Deep learning-based single-image super-resolution: a comprehensive review. IEEE Access 11:21811–21830

Chaurasia A, Culurciello E (2017) Linknet Exploiting encoder representations for efficient semantic segmentation, pp 1–4

Chen L-C, Papandreou G, Schroff F, Adam H (2017) Rethinking atrous convolution for semantic image segmentation arXiv:abs/1706.05587

Chen L-C, Zhu Y, Papandreou G, Schroff F, Adam H, Ferrari V, Hebert M, Sminchisescu C, Weiss Y (2018) (eds) Encoder-decoder with atrous separable convolution for semantic image segmentation. (eds Ferrari, V., Hebert, M., Sminchisescu, C. & Weiss, Y.) Computer Vision – ECCV 2018, pp 833–851 (Springer International Publishing, Cham)

Deb K, Pratap A, Agarwal S, Meyarivan T (2002) A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Trans Evol Comput 6:182–197

Deepender, Manoj, Shrivastava, U. & Verma, J. K. (2021) A study on 5g technology and its applications in telecommunications 365–371

Everingham M, Van Gool L, Williams CK, Winn J, Zisserman A (2010) The pascal visual object classes (voc) challenge. Int J Comput Vision 88:303–338

Galeano-Brajones J, Luna-Valero F, Carmona-Murillo J, Cano PHZ, Valenzuela-Valdés JF (2023) Designing problem-specific operators for solving the cell switch-off problem in ultra-dense 5G networks with hybrid MOEAs. Swarm Evol Comput 78:101290

Galeano-Brajones J et al (2024) Landscape-enabled algorithmic design for the cell switch-off problem in 5G ultra-dense networks. Engineering Optimization, pp 1–23

García-Aguilar I et al (2024) Ferrández Vicente, J. M., Val Calvo, M. & Adeli, H (eds) Prediction of optimal locations for 5g base stations in urban environments using neural networks and satellite image analysis. (eds Ferrández Vicente, J. M., Val - Calvo, M. & Adeli, H.) Bioinspired Systems for Translational Applications: From Robotics to Social Engineering, pp 33–43 (Springer Nature Switzerland, Cham)

Ho J, Jain A, Abbeel P (2020) Denoising diffusion probabilistic models. arXiv:abs/2006.11239

Ji S, Wei S, Lu M (2019) Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set. IEEE Trans Geosci Remote Sens 57:574–586

Knowles J (2005) A summary-attainment-surface plotting method for visualizing the performance of stochastic multiobjective optimizers, pp 552–557

Li R et al (2022) Multiattention network for semantic segmentation of fine-resolution remote sensing images. IEEE Trans Geosci Remote Sens 60:1–13. https://doi.org/10.1109/TGRS.2021.3093977

Li H, Xiong P, An J, Wang L (2018) Pyramid attention network for semantic segmentation. arXiv:abs/1805.10180

Liu Z, Chen B, Zhang A (2020) Building segmentation from satellite imagery using u-net with resnet encoder 1967–1971

Luo Z, Song B, Shen L (2024) Satdiffmoe: A mixture of estimation method for satellite image super-resolution with latent diffusion models arXiv:abs/2406.10225

Mirahsan M, Schoenen R, Yanikomeroglu H (2015) HetHetNets: heterogeneous traffic distribution in heterogeneous wireless cellular networks. IEEE J Sel Areas Commun 33:2252–2265

Mohanty SP et al (2020) Deep learning for understanding satellite imagery: an experimental survey. Frontiers in Artificial Intelligence 3

Nazzal W, Thurnhofer-Hemsi K, López-Rubio E (2024) Improving medical image segmentation using test-time augmentation with medsam. Mathematics 12:4003

Padilla R, Netto SL, da Silva EAB (2020) A survey on performance metrics for object detection algorithms pp 237–242

Quyen VT, Lee JH, Kim MY (2023) Enhanced-feature pyramid network for semantic segmentation, pp 782–787

Ronneberger O, Fischer P, Brox T, Navab N, Hornegger J, Wells WM, Frangi AF (2015) U-net: Convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF (eds) Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015. Springer International Publishing, Cham, pp 234–241

Sahak H, Watson D, Saharia C, Fleet D (2023) Denoising diffusion probabilistic models for robust image super-resolution in the wild arXiv:abs/2302.07864

Saharia C et al (2023) Image super-resolution via iterative refinement. IEEE Trans Pattern Anal Mach Intell 45:4713–4726

Sindhushree K, Naik DC (2023) Advancements and challenges in 5g networks 1–6

Xiao Y et al (2024) Ediffsr: an efficient diffusion probabilistic model for remote sensing image super-resolution. IEEE Trans Geosci Remote Sens 62:1–14

Yue Z, Wang J, Loy CC (2024) Resshift: efficient diffusion model for image super-resolution by residual shifting

Zhao H, Shi J, Qi X, Wang X, Jia J (2017) Pyramid scene parsing network, pp 6230–6239

Zhou Z, Rahman Siddiquee MM, Tajbakhsh N, Liang J Stoyanov D et al (2018) (eds) Unet++: A nested u-net architecture for medical image segmentation. (eds Stoyanov, D. et al.) Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, pp 3–11 (Springer International Publishing, Cham)

Zhu XX et al (2017) Deep learning in remote sensing: a comprehensive review and list of resources. IEEE Geosci Remote Sensing Magazine 5:8–36

Zhu Y et al (2023) Denoising diffusion models for plug-and-play image restoration

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.