Super-Resolution of Sentinel-2 Imagery Using Latent Diffusion Models for Photovoltaic Site Assessment

Ferdaous Chaabane[1], Mohamed Amin Chamtouri[1], Moez Zouari[1], Ghaith Kouki[2], Florence Tupin[3]

[1] COSIM Laboratory, SUP'COM, University of Carthage, Tunisia
[2] State University of New York College of Environmental Science and Forestry, Department of Environmental Resources and Engineering, United States.
[3] Department of Image and Signal Processing, Telecom ParisTech, France

Keywords: Sentinel-2, Super-Resolution, Latent Diffusion Models, Photovoltaic Site Assessment, Deep Learning, Remote Sensing

Abstract

The growing demand for renewable energy emphasizes the critical need for detailed geospatial information in photovoltaic (PV) site assessment and planning. While Sentinel-2 imagery provides a valuable resource, its native 10-meter spatial resolution limits the identification of small urban structures, such as individual rooftops and narrow roads, thereby constraining accurate solar suitability analyses.

To overcome this limitation, this paper presents a comprehensive PV assessment and optimization framework integrating a resolution enhancement module based on latent diffusion models. Operating in the latent space, this module utilizes an iterative diffusion process to accurately reconstruct fine urban structures. Cloud-filtered Sentinel-2 L2A scenes are processed to produce enhanced imagery with an effective 2.5-meter resolution. Pretrained on cross-sensor datasets, the model realistically recovers critical small features while maintaining spectral coherence.

This enhanced imagery enables precise rooftop segmentation, which drives a robust PV potential assessment. The subsequent installation optimization maximizes energy generation by integrating solar radiation, shading analysis, rooftop orientation, tilt angles, and panel layout efficiency, alongside technical and economic constraints. Qualitative evaluations demonstrate high-quality visual enhancement, confirming the relevance of this resolution-enhancement step for real-world PV site suitability analysis and solar deployment optimization.

1. INTRODUCTION

The global transition to renewable energy relies on accurate geospatial data for large-scale photovoltaic (PV) planning. Sentinel-2 imagery is essential for this, providing the multispectral capabilities and wide coverage needed to identify installation sites and monitor urban expansion (Drusch et al., 2012). These high-frequency insights allow researchers to optimize solar potential across diverse regions efficiently.

However, the native 10-meter spatial resolution of Sentinel-2 poses clear limitations, especially when the analysis requires precise detection of small structures such as rooftops, compact buildings, narrow roads, or detailed urban textures. These constraints impact the reliability of PV suitability assessment and highlight the need for enhanced-resolution imagery capable of capturing finer spatial and structural detail.

Several approaches have been explored to address these limitations. Traditional interpolation techniques offer computational simplicity but fail to reconstruct meaningful high-frequency information. Convolutional Neural Network (CNN)-based models introduced significant improvements by learning spatial patterns directly from large datasets (Dong et al., 2014). Later, Generative Adversarial Networks (GANs), such as SRGAN and ESRGAN, produced sharper and more detailed outputs through adversarial learning (Ledig et al., 2017; Wang et al., 2018). Recently, diffusion-based generative models have emerged as a powerful alternative for image restoration tasks, offering stable reconstruction quality and improved control over the generation process (Rombach et al., 2022).

In this context, the European Space Agency (ESA) has developed the OpenSR–SRLatentDiffusion model, specifically tailored for trustworthy enhancement of Sentinel-2 imagery (OpenSR Team, 2024). By operating in a compact latent space, the method reconstructs detailed structures while preserving multispectral coherence and physical realism. This resolution enhancement constitutes a crucial preprocessing stage within broader solar assessment workflows, enabling more precise characterization of built environments and improving the detection of areas suitable for photovoltaic installation.

The enhanced imagery subsequently supports comprehensive installation optimization, incorporating solar radiation data, shading analysis, and technical-economic factors to maximize energy yield viability (Pueblas et al., 2023). The present contribution focuses on both the enhancement phase and its integration within a complete PV assessment pipeline. We apply latent diffusion to generate effective 2.5-meter resolution products from Sentinel-2 L2A scenes and demonstrate how these outputs enable improved rooftop segmentation and installation optimization.

The optimization framework considers multiple factors including solar radiation patterns, atmospheric conditions, rooftop orientation, tilt angles, and shading effects to determine optimal panel placement and system configuration. By restoring fine spatial details in a reliable and physically consistent manner, the method addresses the intrinsic limitations of traditional approaches and provides improved inputs for end-to-end solar energy analysis. The following section presents the integrated methodology and outlines the resulting improvements across the complete workflow.

2. METHODOLOGY

The procedure developed in this study aims to enhance the spatial resolution of Sentinel-2 imagery and enable a more precise assessment of photovoltaic (PV) suitability in both urban and rural environments. The methodology integrates concepts from

Earth observation, deep learning for image resolution enhancement, rooftop segmentation, and solar potential analysis. The overall workflow is summarized in Figure 1, which presents all processing phases and the data sources involved at each stage.

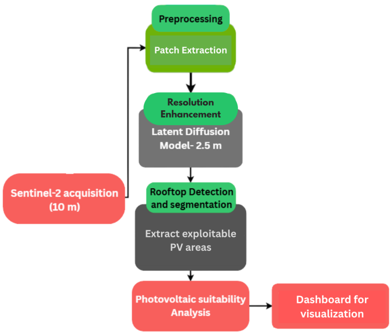


Figure 1: Flowchart of the proposed methodology

Figure 1 illustrates the overall workflow adopted in this study and summarizes the data sources and processing stages involved in the analysis. The methodology is structured into four main steps, each addressing a specific component of the photovoltaic site assessment pipeline, followed by a final visualization stage.

The first step is preprocessing, which begins with the extraction of Sentinel-2 Level-2A imagery via the Microsoft Planetary Computer (MPC). This process utilizes the Spatio-Temporal Asset Catalog (STAC) API to query specific geographic and temporal constraints, automatically selecting the least-cloudy scene through the Sentinel-2 Scene Classification Layer (SCL) (ESA, 2015; Gascon et al., 2017). During this phase, fixed-size multispectral patches are extracted, converted to Bottom-of-Atmosphere (BOA) reflectance values, and normalized to provide a standardized input for the subsequent deep learning models.

The second step enhances the spatial resolution of the preprocessed patch using a latent diffusion–based resolution enhancement model (OpenSR Team, 2024; Song et al., 2020). This produces a high-resolution (2.5 m) multispectral output that preserves fine structural details, effectively reconstructing urban features that are otherwise lost at the native 10-meter resolution.

In the third step, this enhanced image is processed by a convolutional neural network (U-Net) for rooftop detection and semantic segmentation. This enables the automated extraction of precise building footprints and other relevant land-surface categories from the resolution-enhanced imagery.

The fourth step focuses on photovoltaic suitability analysis, where the detected rooftops are evaluated according to their geometric and contextual characteristics to determine their potential for solar energy installation (Pueblas et al., 2023). Finally, all outputs generated throughout the workflow are integrated into a web-based visualization platform, providing an interactive interface for exploring the resolution-enhanced imagery, rooftop footprints, and PV-suitable surfaces.

2.1 Resolution Enhancement Using Latent Diffusion Models

This stage focuses on enhancing the spatial resolution of Sentinel-2 multispectral imagery to enable fine-scale rooftop analysis and photovoltaic site assessment. The original 10 m input patch is processed using a latent diffusion–based resolution enhancement model (OpenSR Team, 2024). Figure 2 illustrates the complete latent diffusion resolution enhancement pipeline.

The model operates in a compressed latent space, where high-frequency spatial details are progressively restored through an iterative reverse-diffusion process (Song et al., 2020), producing an output with an effective resolution of 2.5 m. This approach allows structural consistency to be preserved while introducing realistic high-resolution textures suitable for downstream segmentation tasks. Pretraining on cross-sensor datasets such as SEN2NAIP can improve detail fidelity and texture realism (SEN2NAIP, 2023).

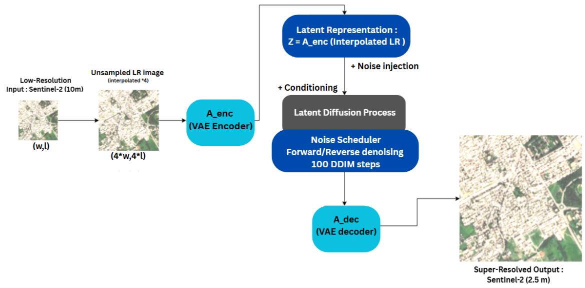


Figure 2: Architecture of the latent diffusion–based resolution enhancement model.

The process begins with the low-resolution Sentinel-2 image, which serves as the multispectral input requiring enhancement for fine-scale rooftop and surface analysis. The image is first interpolated by a factor of four, a step whose objective is to match the spatial dimensions of the target output and prepare the data for subsequent latent-space processing.

The interpolated image is then projected into a compact latent representation using the VAE encoder:

$$z_0 = A_{\text{enc}}(x_{\text{LR}}), \tag{1}$$

where A_{enc} denotes the encoder and x_{LR} the interpolated low-resolution input. Here, z_0 represents the initial latent encoding from which the diffusion process begins.

Inside the latent domain, the model performs a deterministic reverse diffusion process, progressively removing noise and reconstructing high-frequency spatial content. This stage aims to reintroduce details that are missing in the original Sentinel-2 resolution while ensuring that the reconstructed textures remain physically plausible. Each denoising step follows the DDIM update rule:

$$z_{t-1} = \sqrt{\alpha_{t-1}}\, z_t + \sqrt{1-\alpha_{t-1}}\, D_\theta(z_t, t), \tag{2}$$

where D_θ is the denoising model and α_t the noise-schedule coefficient.

After iterating through the full diffusion trajectory, the model produces the final refined latent variable z_{SR}, which represents the high-resolution version of the scene in latent space. The enhanced image is then obtained by decoding this latent representation:

$$x_{\text{SR}} = A_{\text{dec}}(z_{\text{SR}}), \tag{3}$$

where A_{dec} denotes the VAE decoder, z_{SR} is the final denoised latent embedding containing the reconstructed high-frequency information, and x_{SR} is the resulting resolution-enhanced multispectral image at 2.5 m effective resolution.

The objective of this decoding stage is to reconstruct an enhanced multispectral image that preserves spectral fidelity while providing the spatial detail necessary for rooftop segmentation and photovoltaic suitability analysis. This latent-space formulation enables realistic texture reconstruction and high structural fidelity, while ensuring spectral consistency with Sentinel-2 multispectral properties.

2.2 Rooftop and Building Segmentation

After improving the image resolution to 2.5 m, these clear images are used to find buildings for solar panel installation. The sharper edges make it much easier to detect rooftops. However, a major problem was the lack of free, ready-to-use datasets (images with matching masks) at this 2.5 m resolution.

To solve this, a custom dataset was created using an automated algorithm. First, the algorithm reads the exact GPS location directly from the georeferenced imagery. It uses this location data to download the map outlines of the buildings from open-source databases like OpenStreetMap and Overture Maps. Next, it mathematically projects these GPS shapes onto the high-resolution image pixels. This process creates a precise black-and-white mask that perfectly matches the georeferenced imagery, ensuring the roofs are clearly separated from the background.

The resulting custom dataset (shown in Figure 3) consists of 3,521 total patches, with each image sized at 256 $\times$ 256 pixels and 3 color channels (RGB). To properly train and evaluate the deep learning model, the dataset was split into 2,816 patches for training and 705 patches for validation.

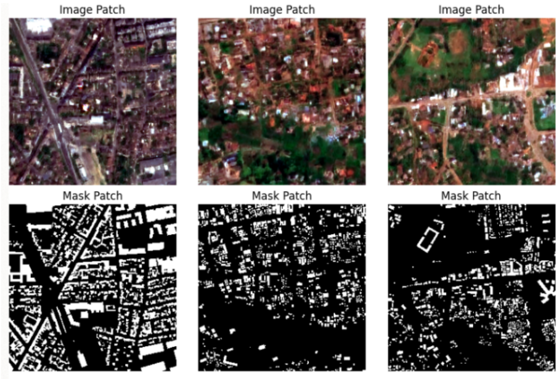


Figure 3: Sample from the custom dataset showing the 2.5 m resolution-enhanced image alongside its generated mask.

To detect the roofs automatically, a deep learning model called U-Net was utilized.

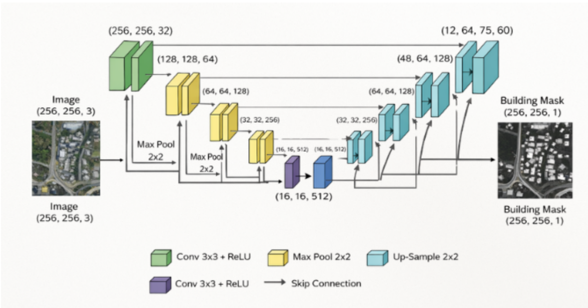


Figure 4: The employed U-Net architecture for rooftop segmentation, featuring contracting and expansive paths with skip connections.

Figure 4 shows the employed architecture. The U-Net architecture features a symmetric, “U”-shaped design consisting of an encoder that shrinks the image to extract deep patterns, and a decoder that rebuilds the image to its original size. Crucially, it uses skip connections that directly link the encoder’s high-resolution details to the decoder. This structure allows the model to combine the broad context of the neighborhood with the fine spatial details of the building edges, resulting in highly precise segmentation masks.

The model was trained on the generated dataset to classify every pixel as either “rooftop” or “background”. It successfully learned to identify complex building shapes while ignoring surrounding trees and roads. Finally, these detected roofs are saved as geometric shapes (polygons) so that the available space for solar panels can be easily calculated.

2.3 Photovoltaic Suitability Analysis

In the final stage, the extracted rooftop geometries are processed to estimate their actual photovoltaic (PV) production potential. The objective is to transform the physical surface area of each detected rooftop into a realistic estimation of monthly and annual energy generation (measured in kWh).

To achieve this, the open-source Python library PVLib is utilized to physically model the solar energy system. The process begins by automatically calculating the exact latitude and longitude from the center of the original georeferenced imagery. Using these geographic coordinates, historical meteorological and solar data are automatically retrieved from PVGIS (Photovoltaic Geographical Information System). This provides a Typical Meteorological Year (TMY) profile for the specific location.

By combining the weather data with the available rooftop area, PVLib simulates the complete energy generation process. It precisely calculates the sun’s position (azimuth and zenith) throughout the year, the solar irradiance falling on inclined surfaces, estimated cell temperatures, and expected system losses. Ultimately, these physical calculations convert the segmented building footprints into highly accurate energy yield estimates. The final outputs including the enhanced imagery, the detected rooftops, and the estimated solar production are then integrated into dashboards, allowing users to easily explore the photovoltaic potential of the entire study area.

3. RESULTS

This section presents qualitative results obtained from the proposed workflow. The focus is on two main aspects: the visual improvement provided by the latent diffusion–based resolution enhancement step and the extraction of building and rooftop structures from the enhanced imagery. All examples are taken from an urban test area in Nairobi, Kenya, which provides a representative mix of residential buildings, streets, and small open spaces suitable for photovoltaic site assessment.

3.1 Resolution Enhancement

The first set of results evaluates the visual improvements introduced by the latent diffusion resolution enhancement model. The objective is to highlight how fine spatial structures such as buildings, roads, and vegetation boundaries become more distinguishable when upgrading the native 10 m Sentinel-2 imagery to an effective 2.5 m resolution. The comparison is presented both at the scene level and through zoomed-in details.

Figure 5: Comparison between low-resolution (LR) and resolution-enhanced (SR) Sentinel-2 imagery, including zoomed-in details.

Figure 5 compares the original low-resolution patch with the corresponding resolution-enhanced output. The LR image and its zoomed detail appear strongly pixelated, with blurred building blocks, indistinct road networks, and limited separation between built-up and vegetated areas. In contrast, the enhanced result shows sharper edges, clearer block structures, and more continuous road segments, revealing spatial information that was not distinguishable at the native resolution.

Figure 6 presents the complete resolution-enhanced Sentinel-2 scene. Individual rooftops, building shapes, and street patterns

Figure 6: Resolution-enhanced Sentinel-2 scene at 2.5 m effective resolution.

become visually discernible, demonstrating a significant gain in spatial interpretability. These qualitative results confirm that the latent diffusion model enhances spatial detail while maintaining the global radiometric consistency of the original Sentinel-2 imagery.

To complement the qualitative evaluation, quantitative metrics were computed to assess the reconstruction quality of the resolution-enhanced imagery. The model achieved a PSNR of 45.88 dB and a SSIM of 0.994 on the test dataset. These metrics are defined as follows:

The Peak Signal-to-Noise Ratio (PSNR) measures the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation. For an $m \times n$ image I and its approximation K, with $\text{MSE} = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - K(i,j)]^2$, it is defined as:

$$\text{PSNR} = 10 \cdot \log_{10}\left(\frac{\text{MAX}_I^2}{\text{MSE}}\right) \tag{4}$$

where MAX_I is the maximum possible pixel value of the image.

The Structural Similarity Index (SSIM) considers image degradation as perceived change in structural information. For two windows x and y of common size, where μ is the average, σ^2 the variance, σ_{xy} the covariance, and c_1, c_2 stabilizing constants:

$$\text{SSIM}(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{5}$$

These results indicate a high level of fidelity between the reconstructed and reference images, confirming that the proposed approach effectively preserves both pixel-level accuracy and structural information. The high SSIM value, in particular, demonstrates that the spatial organization and geometric consistency of urban features are well maintained after resolution enhancement.

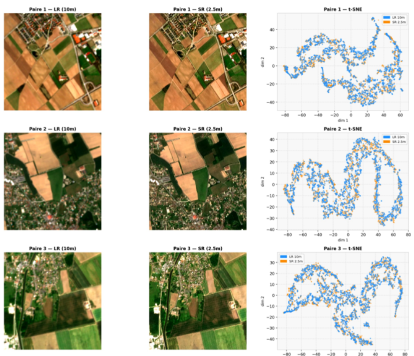
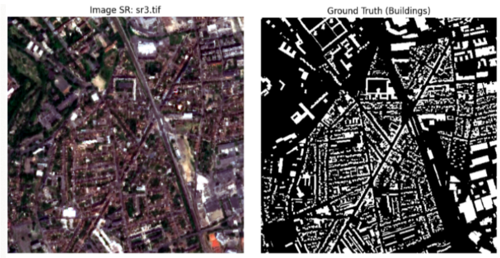


Figure 7: t-SNE feature distributions of LR (blue) and SR (orange) patches confirming spectral coherence after enhancement.

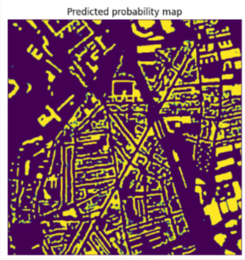
In addition to the pixel-level and structural metrics (PSNR of 45.88 dB and SSIM of 0.994), the preservation of data integrity was evaluated using t-distributed Stochastic Neighbor Embedding (t-SNE). Figure 7 illustrates visual pairs of LR and SR patches alongside their corresponding t-SNE feature distributions. The point clouds show a strong and consistent overlap between the original 10 m data (blue) and the enhanced 2.5 m data (orange). In these plots, the axes (dim 1 and dim 2) simply represent the abstract 2D projection of the complex, high-dimensional image features. This tight alignment in the projected lower-dimensional space visually guarantees that the latent diffusion model respects the original spectral distribution. Consequently, it confirms that the resolution enhancement process successfully synthesizes high-frequency textures without hallucinating completely new features or losing the underlying manifold information of the native Sentinel-2 imagery.

3.2 Rooftop and Building Segmentation

The second set of results evaluates the performance of the trained U-Net model in delineating buildings and rooftops from the resolution-enhanced imagery. The segmentation is performed directly on the enhanced 2.5 m product generated in the previous step. The aim is to extract coherent building footprints that serve as the foundation for the subsequent photovoltaic potential analysis.



(a) Input 2.5 m **(b)** Ground truth



(c) Predicted mask

High Probability Low Probability

Figure 8: Visual evaluation of the segmentation model comparing the resolution-enhanced input image, the actual ground-truth mask, and the model's predicted output (probability map).

Figure 8 illustrates the visual performance of the model over an urban area. The comparison between the ground-truth mask and the predicted mask demonstrates a high level of accuracy. The model successfully detects complex roof shapes, individual houses, and dense building clusters while effectively ignoring background elements such as roads and vegetation. The predicted geometric footprints closely mirror the actual ground-truth structures.

To quantify this performance, two standard evaluation metrics were calculated: the Intersection over Union (IoU) and the Dice coefficient. The model achieved a validation IoU of 0.8205 and a validation Dice coefficient of 0.7860. These strong quantitative results confirm that the U-Net architecture, combined with the enhanced spatial resolution, provides a highly reliable method for rooftop detection. This precision ensures accurate estimations of the available area for solar panel installations in the final processing stage.

3.3 Photovoltaic Suitability Analysis Results

The final step of the workflow successfully translates the detected rooftops into practical solar energy estimations. Building directly upon the segmentation results presented previously, the entire urban area extracted from the resolution-enhanced imagery (shown in Figure 8) is processed to evaluate its collective energy generation potential. By combining the physical surface area of all identified structures with local meteorological data, the system models the complete energy yield for the neighborhood.

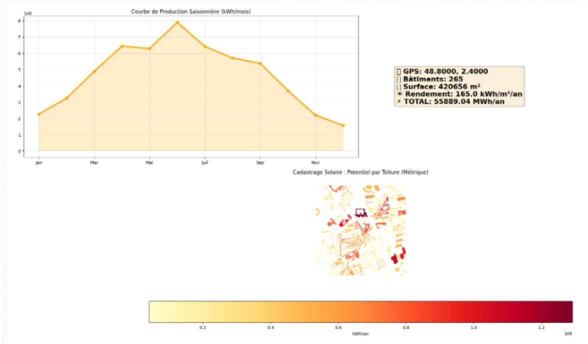


Figure 9: Photovoltaic suitability analysis for the segmented area, featuring the seasonal energy production curve (top) and a solar cadastre map highlighting the potential of individual rooftops (bottom).

Figure 9 presents the quantitative outcome of this regional analysis. The system successfully processed 265 detected buildings within the sample area, encompassing a total usable rooftop surface of 420,656 m^2. The generated results feature a seasonal production curve, which illustrates the expected monthly variations in energy yield, peaking during the summer months. Additionally, a solar cadastre map is produced, color-coding individual building footprints based on their specific energy generation capacity.

Overall, the physical simulation estimates an average yield of 165.0 kWh/m^2/year. To obtain this result and ensure the estimations reflect realistic operational conditions, several key technical parameters were fixed within the simulation model. First, a standard photovoltaic module efficiency of 20% was assumed, representing typical modern commercially available monocrystalline silicon panels. Second, the installation was modeled with an optimal fixed tilt angle of 35° oriented South, which maximizes the annual solar energy capture for the specific latitude of the region. Third, an urban surface albedo coefficient of 0.2 a metric representing the fraction of sunlight reflected back by surrounding surfaces like adjacent roofs and asphalt was applied to capture the indirect light reaching the panels. Finally, a comprehensive system loss factor of 14% was incorporated. This factor accounts for the inevitable reduction in usable energy caused by real-world physical penalties, specifically the energy lost when converting the solar panels' native Direct Current (DC) electricity into grid-compatible Alternating Current (AC) through inverters, as well as electrical resistance in wiring and natural panel soiling. For this single analyzed neighborhood, applying these rigorous fixed parameters translates to a substantial total estimated production of 55,889.04 MWh per year. These findings clearly demonstrate the pipeline's capability to seamlessly convert raw satellite imagery into large-scale, actionable renewable energy assessments grounded in physical modeling.

4. CONCLUSION

This paper introduced a hybrid framework combining latent diffusion–based resolution enhancement, semantic segmentation, and physically grounded photovoltaic modeling to enhance solar site assessment from Sentinel-2 imagery. By addressing the intrinsic limitation of the 10 m spatial resolution, the proposed approach enables the reconstruction of fine-scale urban structures, which are critical for accurate rooftop detection and solar potential estimation.

The results highlight that the use of latent diffusion models provides a substantial improvement in spatial interpretability, allowing the extraction of detailed rooftop geometries that were not accessible at the original resolution. When coupled with a U-Net–based segmentation model, the enhanced imagery leads to reliable building footprint extraction, forming a robust basis for downstream photovoltaic analysis. The integration of these outputs into a physics-based modeling pipeline further demonstrates the practical relevance of the approach for real-world energy applications.

However, the study also reveals important challenges. The reconstruction process inherently relies on learned priors, which may introduce uncertainties in the generated high-frequency details. Moreover, the dependence on automatically generated ground-truth data and the limited geographic scope of the experiments may affect the generalization of the results. These aspects underline the need for careful validation when deploying such methods in operational contexts.

References

Dong, C., Loy, C.C., He, K. and Tang, X., 2014. Learning a deep convolutional network for image super-resolution. In: *European Conference on Computer Vision (ECCV)*, pp. 184–199.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F. and Bargellini, P., 2012. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sensing of Environment*, 120, pp. 25–36.

European Space Agency (ESA), 2015. *Sentinel-2 User Handbook*. ESA Standard Document.

Gascon, F., Bouzinac, C., Thépaut, O., Jung, M., Francesconi, B., Louis, J., Lonjou, V., Lafrance, B., Massera, S., Gaudel-Vacaresse, A., Languille, F., Alhammoud, B., Viallefont, F., Pflug, B., Bieniarz, J., Clerc, S., Devignot, O., Pessiot, L., Trémas, T., Cadau, E., De Bonis, R., Isola, C., Martimort, P. and Fernandez, V., 2017. Copernicus Sentinel-2 product evolution: calibration and validation status. *Remote Sensing*, 9(6), 584.

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z. and Shi, W., 2017. Photo-realistic single image super-resolution using a generative adversarial network. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4681–4690.

European Space Agency (ESA) OpenSR Team, 2024. Trustworthy super-resolution of multispectral Sentinel-2 imagery with latent diffusion. *ESA Technical Report*.

Pueblas, R., Kuckertz, P., Weinand, J.M., Linssen, J. and Stolten, D., 2023. ETHOS.PASSION: An open-source workflow for rooftop photovoltaic potential assessments from satellite imagery. *Solar Energy*, 265, 112094.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P. and Ommer, B., 2022. High-resolution image synthesis with latent diffusion models. In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10684–10695.

SEN2NAIP, 2023. SEN2NAIP: Cross-sensor dataset pairing Sentinel-2 L2A with NAIP for super-resolution research. Dataset resource.

Song, J., Meng, C. and Ermon, S., 2020. Denoising diffusion implicit models (DDIM). *arXiv:2010.02502*.

Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Loy, C.C., Qiao, Y. and Tang, X., 2018. ESRGAN: Enhanced super-resolution generative adversarial networks. In: *ECCV Workshops (ECCVW)*.