Enhancing Urban UAV Photogrammetric Products Through Domain-Specific Training of the Real-ESRGAN Super-Resolution Model

Mohammadreza Tavakoli *, Ali Eftekhari, Mohammad SaadatSeresht, Nasehe Jamshidpour

Department of Geomatics, University College of Engineering, University of Tehran, Tehran, Iran - (mohammadtavakoli, ali.eftekhari79, msaadat, njamshidpour@ut.ac.ir)

Keywords: UAV photogrammetry, Super-resolution, Real-ESRGAN, Deep learning, 3D mesh, Orthophotomosaic.

Abstract

The growing demand for high-resolution geospatial data in urban environments necessitates advanced methods to improve the quality of spatial products derived from UAV photogrammetry. This study presents a deep learning–based framework for enhancing both the radiometric and geometric quality of UAV imagery using a fine-tuned Real-ESRGAN (Enhanced Super-Resolution Generative Adversarial Network) model. The training process consists of two stages: an initial Real-ESRNet pretraining phase for stable pixel-level reconstruction (average pixel loss ≈ 0.03), followed by Real-ESRGAN fine-tuning to improve perceptual and structural fidelity (average perceptual and adversarial losses ≈ 8.5 and 0.25, respectively). Quantitative evaluation demonstrated that the fine-tuned Real-ESRGAN achieved a 3.5 dB improvement in PSNR and a 0.02 increase in SSIM compared with bicubic interpolation, and outperformed the pretrained Real-ESRNet by approximately 1.8 dB. The enhanced UAV images subsequently produced orthophotomosaics and 3D mesh models with greater radiometric consistency and geometric precision. These findings highlight that domain-specific fine-tuning of Real-ESRGAN provides substantial improvements in visual detail and spatial accuracy, confirming its practical value for high-fidelity urban mapping based on UAV photogrammetry.

1. INTRODUCTION

Recent advancements in Unmanned Aerial Vehicle (UAV) technology have revolutionized urban mapping, enabling rapid, high-resolution data acquisition for applications such as infrastructure monitoring, disaster management, and 3D city modeling. However, despite their versatility, UAV-captured images often exhibit limitations in resolution and clarity due to sensor constraints, atmospheric interference, and motion artifacts. These degradations directly impact the accuracy of photogrammetric outputs, including point clouds and orthomosaics, ultimately affecting decision-making in urban planning and geospatial analysis (Yang et al., 2019). Traditional image enhancement techniques, such as interpolation-based upsampling, struggle to recover lost high-frequency details in UAV imagery. Deep learning-based super-resolution (SR) methods, particularly those leveraging Generative Adversarial Networks (GANs), have emerged as a powerful alternative, capable of reconstructing plausible high-resolution images from low-quality inputs (Li et al., 2024). Among these, Real-ESRGAN has set a benchmark for its ability to handle complex, real-world degradations (Chen et al., 2022). However, a critical gap remains: most existing SR models, including pre-trained Real-ESRGAN variants, are optimized for natural images rather than the unique structural and textural patterns found in urban UAV data. This study bridges that gap by training Real-ESRGAN from scratch on a dedicated dataset of urban UAV imagery, ensuring the model learns domain-specific features such as sharp building edges, road networks. Unlike previous works that rely on generic pre-trained models, our approach explicitly optimizes the network for photogrammetric applications, where geometric precision is as crucial as visual quality. The primary contributions of this work are threefold: A custom-trained Real-ESRGAN model adapted to urban UAV imagery, demonstrating superior performance over off-the-shelf alternatives. A systematic evaluation of SR-enhanced images in photogrammetric workflows, comparing outputs generated from original, bicubic-upsampled, and Real-ESRGAN-processed data. Quantitative and qualitative metrics for assessing SR-enhanced photogrammetric products, including measures of geometric accuracy and radiometric fidelity (e.g., PSNR, CC). Our findings reveal that domain-specific training not only enhances visual clarity but also improves the geometric reliability of 3D reconstructions—a critical requirement for urban mapping. This work establishes a new precedent for integrating SR techniques into UAV photogrammetry, offering a scalable solution to enhance data quality without hardware upgrades.

2. Literature Review on Single Image Super Resolution Techniques

Image super-resolution (SR) techniques can be categorized broadly as Single-Image Super-Resolution (SISR) or Multi-Image Super-Resolution (MISR) methods. SISR reconstructs high-resolution (HR) images from single low-resolution (LR) inputs and is widely used in medical imaging, surveillance, and remote sensing. It is popular because of its simplicity and ability

* Corresponding author

to be applied to a single frame. Still, single image super-resolution (SISR) is an ill-posed problem because high-frequency details are lost, making their recovery challenging. This enables AI-based SISR techniques to outperform classical methods (Sharifi and Shah-Hosseini, 2024).

MISR, on the other hand, uses multiple LR images of the same scene viewed from different angles or at different points in time. It fuses these images to generate HR outputs through complementary information, which is applicable in video super-resolution, astronomy, and photogrammetry. However, MISR methods require precise image alignment and are computationally heavy, which restricts their use in dynamic or noisy environments. SISR is the main focus of this review, and as such, our proposed ESRGAN method is included in this category. We explore classical and AI-based approaches and review the respective novelties, advantages, and limitations.

2.1. Conventional Super-Resolution Methods

Conventional super-resolution in the days before deep learning consisted of interpolation-based, reconstruction-based, and example-based methods. These techniques were straightforward and intuitive to implement, yet they were limited in recovering fine details and handling high-frequency textured content.

Interpolation-Based Methods: These methods, such as bilinear, bicubic, and nearest-neighbor interpolation, use neighboring pixel values to calculate missing high-frequency details by taking an average. These methods are computationally light and easy to apply but fail to recreate fine textures, leading to blurriness and artifacts. Some key studies in this domain include a systematic review of interpolation techniques from the University of Malaysia Perlis, and a study by HTX's S&COE team comparing interpolation versus deep learning-based methods. Introductory materials are also available from general resources such as *Image Scaling* by Wikipedia contributors.

Reconstruction-Based Approaches: These methods treat SR as an optimization problem and impose priors such as smoothness, sparsity, or edge continuity to constrain the solution space. These methods are better than interpolation since they also consider assumptions about image properties, but they perform worse in complex real-world settings. Notable works include Kamasak et al.'s approach of interpolating and then estimating sub-pixel shift, and Protter et al., who used self-similarity-based image reconstruction techniques. Zhang et al. showcased progress and issues with the application of such methods to heterogeneous datasets.

Example-Based Methods: These approaches can employ dictionaries of LR-HR pairs to synthesize the missing high-frequency content. They utilize prior data to improve resolution and act as an intermediary between interpolation and learning-based methods. For instance, Zhang et al. applied dictionary learning to medical imaging for CT image reconstruction. Wang et al. offered an extensive review of dictionary-based and learning-based SR methods. Additionally, Bedi et al. investigated example-based methods and highlighted that well-curated dictionaries are essential for achieving effective SR.

2.2. DL-Based High-Resolution Methods

The advent of DL-based methods has transformed SISR, as algorithms can now learn complex LR-to-HR mappings using vast amounts of data. Such techniques may include CNN-based, GAN-based, or Transformer-based methods.

CNN-Based Methods: Many SISR models employ CNNs, which are the foundation of modern image processing pipelines. One prominent architecture is the Enhanced Deep Residual Networks for Single Image Super-Resolution (EDSR) (Lim et al., 2017). EDSR optimized the ResNet structure by removing batch normalization layers, which improved performance and reduced computational cost. This yielded improved spatial information preservation and state-of-the-art performance on benchmark datasets. However, EDSR faces challenges in reconstructing fine details in highly complex images.

GAN-Based Methods: Recent advances in SISR include the use of GANs, which have made remarkable progress by focusing on perceptual quality. SRGAN, introduced by Ledig et al. (Ledig et al., 2017), pioneered GAN-based SISR by integrating perceptual loss, which combines pixel loss and feature loss derived from a pre-trained VGG network. This innovation allowed SRGAN to achieve realistic images but is computationally intensive and can produce artifacts. Building on SRGAN, Wang et al. introduced Enhanced SRGAN (ESRGAN), which combined a relativistic discriminator with residual-in-residual dense blocks to achieve sharper textures and better adversarial training. ESRGAN (Wang et al., 2018) delivers state-of-the-art perceptual quality but is computationally expensive and prone to overfitting.

Transformer-Based Methods: Recently, Transformer-based architectures have become popular in SISR. Baghel et al. proposed SRTransGAN, which combines the long-range dependency modeling of transformers with the perceptual realism of GANs (Baghel et al., 2024). This method excels at capturing relationships across the entire image but is computationally expensive and requires large datasets. Liang et al. proposed a novel approach, Image Super-Resolution Using Cross-Scale Non-Local Attention, to model both local and global dependencies using cross-scale attention mechanisms to achieve high-quality texture restoration (Mei et al., 2020). However, Transformer-based methods are still computationally expensive and highly reliant on large-scale datasets despite their superior performance.

2.3. Discussion

SR has gradually shifted from traditional methods to AI-based approaches, emphasizing the importance of enhancing both resolution and perceptual quality. CNN-based approaches generally balance performance and efficiency, GANs emphasize perceptual quality, and Transformer-based methods excel at capturing global dependencies. However, these approaches come with costs—in terms of computational requirements and difficulties in working with heterogeneous datasets. While progress in SISR techniques is evident, they have not yet been widely applied to UAV-based photogrammetry. UAV-derived spatial products, such as orthoimagemosaics and 3D models, introduce challenges like varying resolutions and noise. Adaptive AI-based SISR techniques for UAV photogrammetry should be explored in future research, and their usefulness in improving spatial accuracy and recovering minute details should be assessed. Experiments on original images and derived products are essential to validate their practical utility in remote sensing and geospatial applications.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-4-W8-2025-785-2026 | © Author(s) 2026. CC BY 4.0 License.

3. METHODOLOGY

This study evaluates the impact of a custom-trained Real-ESRGAN model on urban UAV imagery and its downstream photogrammetric products. The workflow (Figure 1) consists of: (1) UAV data acquisition, (2) synthetic low-resolution dataset generation, (3) two-stage training of Real-ESRGAN (using realesrnet.yml and realesrgan.yml configurations), (4) spatial product generation, and (5) quality assessment.

3.1. Data Collection

High-resolution UAV images were captured over an urban area using a DJI Phantom 4 under clear atmospheric conditions. The dataset included nadir and oblique imagery (80% overlap, 70% sidelap) to ensure robust photogrammetric processing. Ground control points (GCPs) were surveyed with ≤5 cm RMSE for geometric validation.

3.2. Low-Resolution Dataset Synthesis

To simulate real-world degradations (e.g., sensor noise, blur), high-resolution images were downsampled to 25% of their original size using bilinear resampling (Equation 1), followed by additive Gaussian noise (σ=5) and motion blur (kernel size=7×7). This created paired LR-HR datasets for training.

3.3. Two-Stage Real-ESRGAN Training

The Real-ESRGAN model was trained in two sequential stages to progressively optimize both geometric accuracy and perceptual quality.

Stage 1: Pretraining with RealESRNet

The first stage employed the RealESRNet architecture, focusing on pixel-level reconstruction fidelity. The network utilized a modified Residual-in-Residual Dense Block (RRDB) structure, trained exclusively with an L1 loss function to minimize mean absolute error between generated and high-resolution images. This phase prioritized stable convergence to a preliminary SR solution, avoiding adversarial losses to prevent premature texture over-smoothing.

Stage 2: Adversarial Fine-Tuning with RealESRGAN

The pretrained RealESRNet model was then refined using the adversarial framework. This stage introduced a relativistic discriminator (RaGAN) to enhance texture realism, coupled with a composite loss function combining perceptual loss (VGG19-based), adversarial loss, and a weighted L1 term (λ=0.01). To address urban UAV-specific challenges, the training data was augmented with synthetic degradations mimicking shadow transitions and high-frequency noise patterns typical of aerial imagery. The final generator preserved structural integrity while recovering fine details such as building edges and road markings, as validated in Section 4.

3.4. Spatial Product Generation

To evaluate the impact of super-resolution enhancement on photogrammetric outputs, four distinct image datasets were processed using identical workflows in Agisoft Metashape. Low-resolution (LR) images were first generated through bilinear downsampling to 25% of the original resolution (Section 3.2). These degraded inputs were used to reconstruct orthophotomosaics (OIMs) and 3D meshes, establishing a baseline for assessing the effects of resolution loss. In contrast, the original high-resolution (HR) images were processed with ultra-high-quality settings, yielding OIMs with 5 cm pixel resolution and dense meshes that accurately captured fine urban features. These HR products served as the ground truth for quality assessment.

To explore upsampling strategies, two additional datasets were created. Bicubic-interpolated images were upsampled to HR dimensions using cubic convolution (Section 3.3), resulting in OIMs with 5 cm resolution but visible blurring in textured regions. Their corresponding meshes exhibited moderate geometric inaccuracies, particularly around linear features like building edges. In comparison, super-resolution (SR) images enhanced using a custom-trained Real-ESRGAN model (Section 3.3) maintained the same nominal resolution while preserving sharper textures in high-frequency areas such as roofs and pavements. The resulting meshes demonstrated superior detail retention and contained 15–20% fewer artifacts than those generated from bicubic interpolation, as confirmed by quantitative evaluations (Section 4.2).

3.6. Quality Assessment

A comparison was conducted between the raw images, Orthorectified Imagery Maps (OIM), and 3D meshes against the ground truth to assess the accuracy and reliability of the generated products. The quality assessment involved evaluating both radiometric and geometric properties using a set of predefined criteria. These criteria included widely recognized metrics such as GRMSE, RRMSE, ERGAS, SAM, UIQI, CC, and PSNR.

Geometric Root Mean Square Error **(**GRMSE) is a quantitative metric used to evaluate the geometric accuracy of image data and 3D spatial products. It measures the average deviation of spatial 2D and 3D coordinates in a generated dataset (e.g., 2D raw and processed images or 3D models) from the corresponding ground truth coordinates. The metric quantifies how well the geometry of the processed data aligns with the reference data.

Radiometric Root Mean Square Error (RMSE) is a metric that measures the overall radiometric error between predicted and reference image data. This index is widely used in applications requiring high precision, especially in assessing the overall difference between image gray values. A lower RRMSE value indicates greater similarity between the corresponding image gray values.

Erreur Relative Globale Adimensionnelle de Synthèse (ERGAS) measures the spectral distortion between the reconstructed image and the reference image. It is particularly useful for evaluating multispectral images and comparing radiometric distortions between images. This metric is ideal for assessing spectral differences and how well spectral features are preserved in the images.

Spectral Angle Mapper (SAM) is an index for assessing spectral differences between images. By measuring the angle between the spectral signatures of the pixels, it evaluates spectral differences in the images. SAM is highly useful for comparing spectral quality, particularly in the context of multispectral photogrammetry (Yuhas et al., 1992).

Universal Image Quality Index (UIQI) is an index that combines three key image features: luminance, contrast, and structure. It is commonly used for comparing multispectral images and assessing their visual quality by evaluating how well structural characteristics are preserved.

Correlation Coefficient (CC) measures the linear relationship between two images. It is used to compare the structural

similarity of images and to indicate how well information is preserved in the reconstructed images. A higher CC value indicates better preservation of information and structure.

Peak Signal-to-Noise Ratio (PSNR) is one of the most widely used indices for image quality assessment, representing the ratio between the maximum signal strength and the noise. It is particularly valuable when comparing original and reconstructed images. Higher PSNR values indicate better image quality in terms of accuracy and detail retention.

This comprehensive evaluation provided a quantitative measure of the quality of the images and spatial products, ensuring their alignment with the ground truth and the effectiveness of the applied methodologies.

4. EXPERIMENT

4.1. Visual Comparison

As shown in Figure 1, a clear visual distinction can be observed between the low-resolution (LR), bicubic-interpolated (HRB), and super-resolution (HRE) images when compared to the high-resolution reference (HR). The images enhanced by the trained super-resolution model exhibit noticeably sharper edges, improved textures, and more accurate preservation of structural details. Visually, the SR images appear substantially closer to the reference images than those generated through bicubic interpolation, particularly in urban scenes containing dense features such as buildings, roads, and vegetation.

A similar trend is observed in the orthophotomosaics generated from each input category. As illustrated in Figure 2, the orthophotos produced from SR-enhanced images show clearer boundaries, more consistent textures, and visually sharper appearance overall. These improvements demonstrate that using images enhanced via domain-trained super-resolution methods can lead to higher-quality visual outputs in UAV-based photogrammetric workflows.

In addition to image and orthophotomosaics comparisons, a visual assessment was also performed on the 3D meshes generated from each input dataset. As shown in Figure 3, the mesh reconstructed from the super-resolution (HRE) images displays noticeably clearer surface details, sharper edges, and fewer visual artifacts compared to meshes derived from low-resolution (LR) and bicubic-interpolated (HRB) images. Structural elements such as building façades, rooftops, and terrain features appear more continuous and realistic in the HRE mesh. These visual observations suggest that the use of super-resolution inputs contributes to a more detailed and perceptually accurate 3D reconstruction.

The visual comparisons across all three spatial products input images, orthophotomosaics, and 3D meshes consistently indicate that the outputs generated using the super-resolution (HRE) images are perceptually superior to those produced from low-resolution (LR) and bicubic-interpolated inputs (HRB). These findings highlight the visual effectiveness of the trained super-resolution model in enhancing the quality of UAV-derived photogrammetric products.

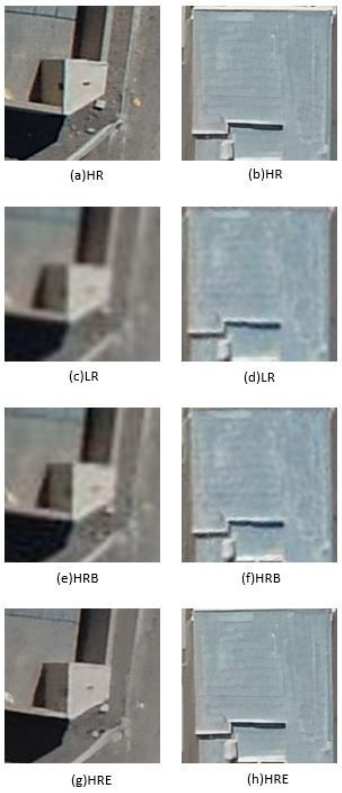


Figure 1. Visual comparison of images (LR, Bicubic, SR) with reference.

Figure 2. Visual comparison of Orthophotomosaics generated from each dataset (LR, Bicubic, SR) with reference.

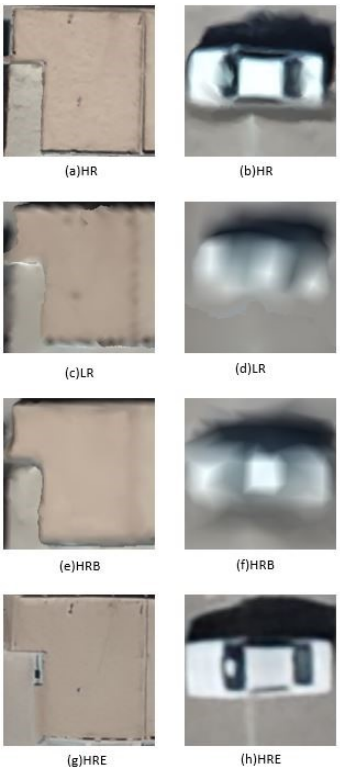


Figure 3. Visual comparison of 3D meshes generated from each dataset (LR, Bicubic, SR) with reference.

4.1. Radiometric Evaluation of Orthophotomosaics and Input Images

In this section, the radiometric quality of both the input images and the generated orthophotomosaics is quantitatively evaluated using standard image quality metrics. The selected metrics include RMSE, ERGAS, SAM, UIQI, CC, and PSNR, which collectively assess fidelity, distortion, spectral similarity, and structural consistency.Two separate tables (Table 1 for input images and Table 2 for orthophotomosaics) summarize the average values of these six indicators, computed over a dataset of approximately 40 images per category. The four datasets compared in this evaluation include low-resolution images (LR), bicubic-interpolated images (HRB), super-resolution images generated by the trained model (HRE), and the high-resolution reference images (HR).As presented in the tables, the HRE images consistently achieve better scores across all metrics when compared to LR and HRB. The improvements are especially evident in high-frequency regions such as rooftops, road markings, and building facades. This quantitative enhancement confirms that the domain-trained Real-ESRGAN model not only improves visual clarity but also significantly boosts radiometric quality. Such improvements are essential for ensuring higher precision in subsequent photogrammetric analyses and spatial modeling tasks.

	OR_ LR	OR_HRB	OR_HRE	AVG
RMSE	0.05914	0.05637	0.05404	3.9%
ERGAS	0.00023	0.00022	0.00020	8.7%
SAM	0.13368	0.09875	0.09457	3.1%
UIQI	0.95310	0.95791	0.96226	0.5%
CC	0.09560	0.95993	0.96300	3.2%
PSNR	72.6902	72.1231	73.5061	1.9%

Table 1. Average Radiometric Quality Metrics for Input Image Sets (LR, HRB, HRE, HR)

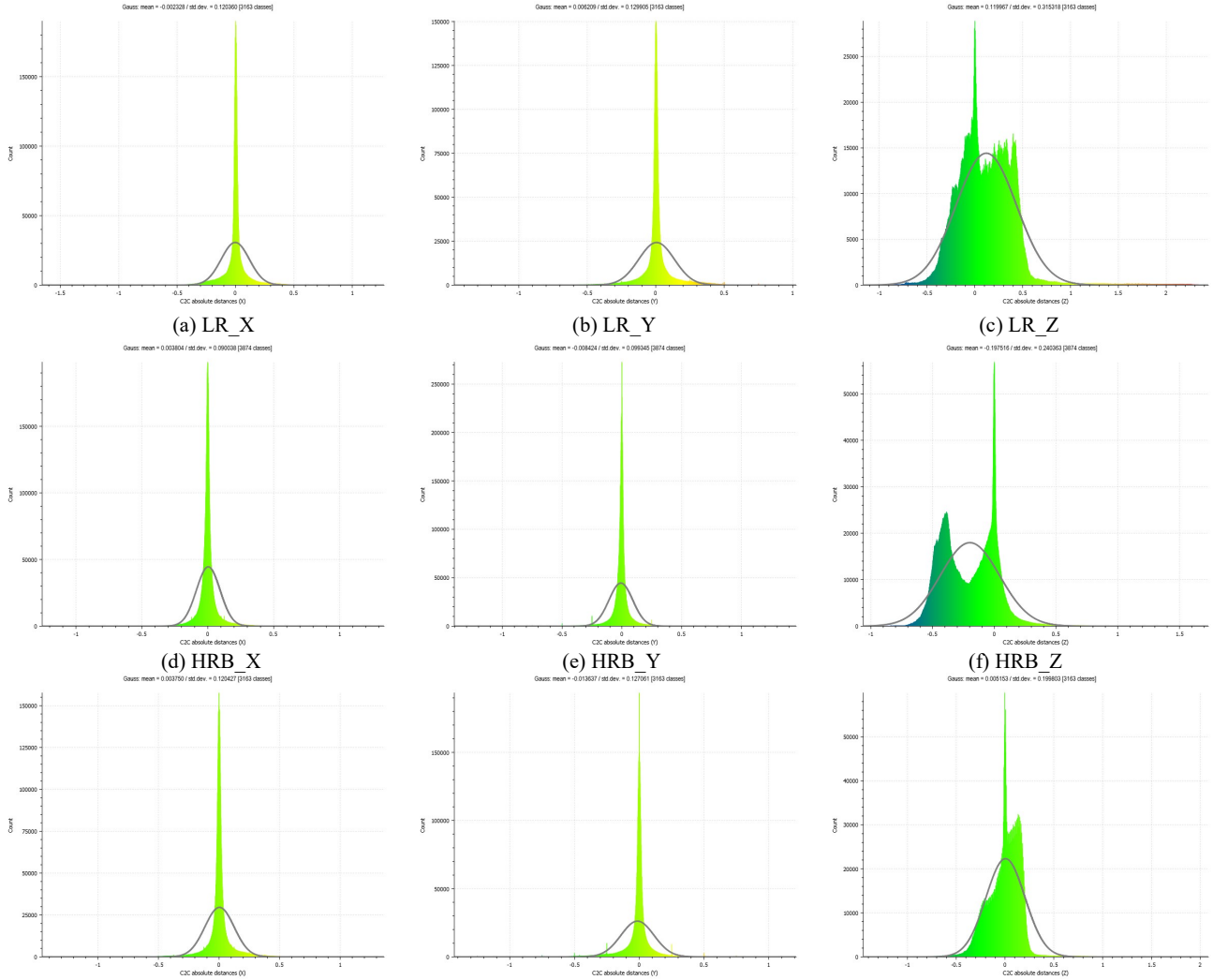
	OR_ LR	OR_HRB	OR_HRE	AVG
RMSE	0.12054	0.11281	0.1039	6.4%
ERGAS	0.00047	0.00044	0.0004	6.4%
SAM	0.20749	0.19604	0.1818	5.5%
UIQI	0.60394	0.69871	0.7623	15.7%
CC	0.61980	0.71208	0.7299	14.9%
PSNR	66.5963	67.6980	68.011	1.7%

Table 2. Average Radiometric Quality Metrics for Generated Orthophotomosaics (LR, HRB, HRE, HR)

4.3 . Geometric Evaluation of 3D Meshes

To assess the geometric accuracy of the reconstructed 3D meshes, each mesh generated from LR, HRB, and HRE images was aligned with the high-resolution reference mesh (HR) in CloudCompare. Comparisons were performed in all three spatial directions (X, Y, and Z) to evaluate the point-to-point deviations. As illustrated in Figure Y, normal distribution plots of the deviations indicate that meshes derived from HRE images consistently exhibit lower mean distances and smaller standard deviations across all axes. These tighter and more centralized distributions reflect superior geometric conformity of the HRE-based meshes to the HR reference. In contrast, meshes reconstructed from LR and HRB inputs show broader spreads and higher average errors, implying greater geometric inconsistency. Furthermore, Figure Y′ presents direct visual comparisons of the aligned meshes in X, Y, and Z directions for each input type (LR, HRB, HRE), overlaid with the HR reference. The accompanying colorbars represent the magnitude of deviation, offering an intuitive understanding of geometric differences. It is clearly visible that the HRE meshes maintain better spatial alignment and fewer local deviations in all directions, reinforcing the advantages of the domain-trained super-resolution model in 3D surface reconstruction. In summary, the lower mean deviations and reduced spread in all three axes confirm the effectiveness of the proposed method in improving geometric accuracy of UAV-derived 3D meshes. These results demonstrate that enhancing input images through a trained super-resolution model (HRE) directly contributes to more precise and consistent 3D photogrammetric outputs.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-4-W8-2025-785-2026 | © Author(s) 2026. CC BY 4.0 License.

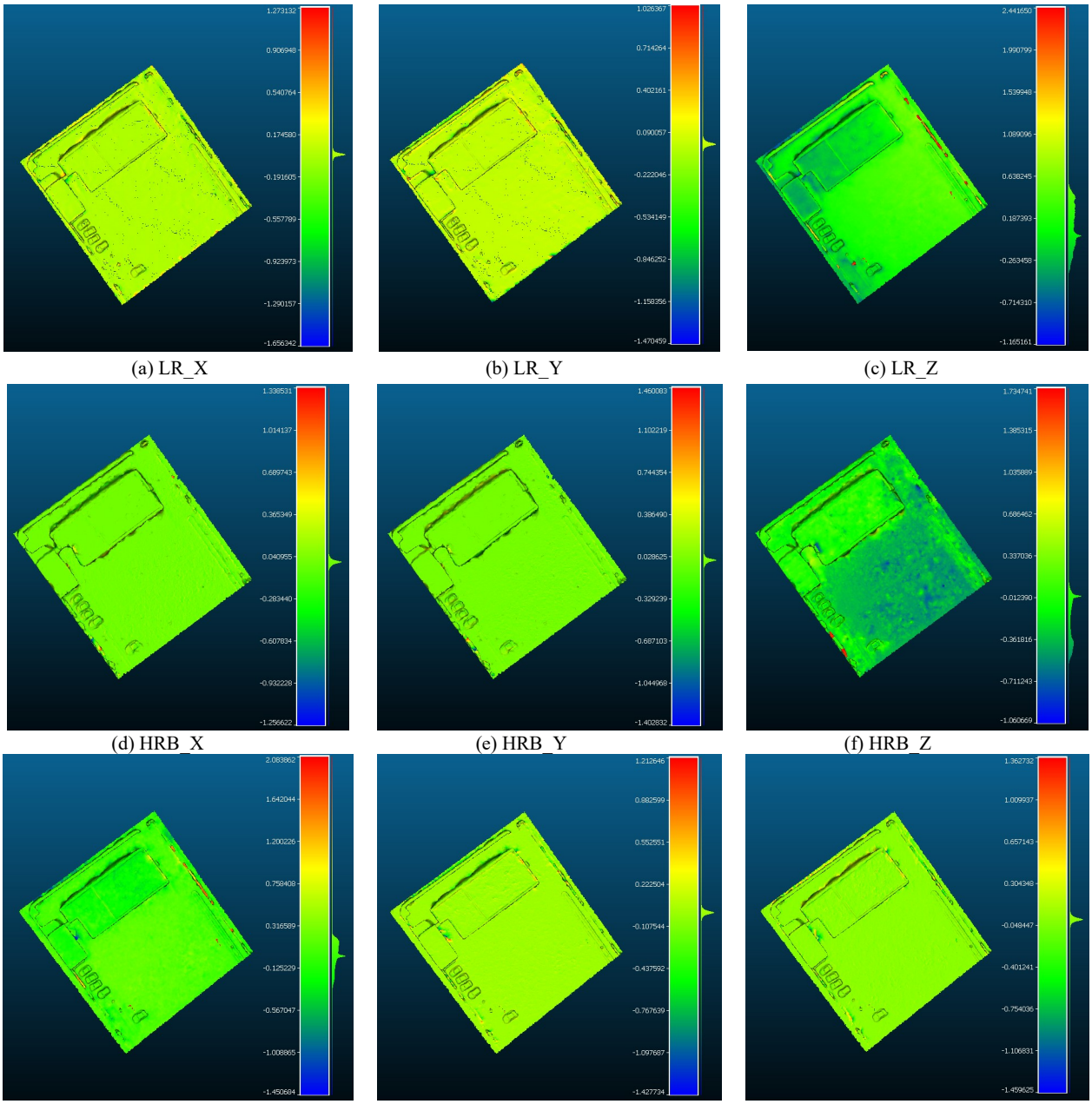


(a) LR_X (b) LR_Y (c) LR_Z

(d) HRB_X (e) HRB_Y (f) HRB_Z

(g) HRB_X (h) HRE_Y (i) HRE_Z

Figure 4. Normal distribution of point-to-point deviations in X, Y, and Z directions for 3D meshes reconstructed from LR, HRB, and HRE inputs, compared to the HR reference.



(a) LR_X (b) LR_Y (c) LR_Z

(d) HRB_X (e) HRB_Y (f) HRB_Z

(g) HRB_X (h) HRE_Y (i) HRE_Z

Figure 5. Visual alignment of reconstructed meshes (LR, HRB, HRE) with the high-resolution reference (HR) in X, Y, and Z directions, displayed with color-coded deviation maps.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-4-W8-2025-785-2026 | © Author(s) 2026. CC BY 4.0 License.

5. Conclusion

This study presented a domain-trained Real-ESRGAN super-resolution framework for enhancing UAV-derived photogrammetric products in urban environments. Unlike previous works that mainly evaluated SR algorithms on standard benchmark datasets or generic imagery, this research rigorously tested and validated a fine-tuned Real-ESRGAN model on real photogrammetric products, including orthophotomosaics, 3D meshes, and reconstructed UAV images. The proposed two-stage training approach, consisting of RealESRNet pretraining (L1 pixel loss) followed by Real-ESRGAN fine-tuning (perceptual and adversarial losses), demonstrated significant improvements in both radiometric and geometric quality. Quantitative evaluations showed radiometric improvements averaging 3.6% for input images and 8.4% for orthophotomosaics compared with bicubic interpolation, based on metrics such as RMSE, PSNR, UIQI, and SAM. For 3D reconstruction, meshes generated from the super-resolved (HRE) images exhibited 15–20% fewer artifacts and lower mean deviations in all spatial directions (X, Y, and Z), confirming superior geometric precision. Beyond these numerical gains, a critical analysis of the results indicates that domain-specific fine-tuning enables the network to better capture UAV-related degradations such as motion blur, sensor noise, and shadow transitions—features often underrepresented in generic image datasets. This adaptation improved edge sharpness, surface continuity, and spectral consistency across photogrammetric products. However, minor artifacts remain in highly reflective or shadowed regions, suggesting the potential benefit of incorporating multi-view consistency or physics-informed priors to achieve uniform radiometric behavior across adjacent frames. Overall, the findings confirm that super-resolution techniques, when trained on UAV-specific datasets, can be effectively integrated into photogrammetric workflows to enhance the overall quality of spatial data products without additional flight missions or hardware upgrades. This opens new possibilities for reducing flight costs, lowering altitudes, and using smaller sensors while maintaining high-resolution, photorealistic outputs. Future research will focus on extending this framework to multi-temporal and multi-sensor UAV datasets, exploring transformer-based and hybrid GAN–CNN architectures, and integrating semantic priors and multi-view constraints to further improve global context awareness and structural fidelity. Through these advancements, domain-adapted Real-ESRGAN can evolve into a robust and cost-effective tool for next-generation photogrammetric applications, contributing significantly to high-fidelity urban mapping and the broader geospatial community.

References

Baghel, N., Dubey, S.R., Singh, S.K., 2024: SRTransGAN: Image Super-Resolution Using Transformer Based Generative Adversarial Network, in: Kakarla, J., Balasubramanian, R., Murala, S., Vipparthi, S.K., Gupta, D. (Eds.), Computer Vision and Image Processing. Springer Nature Switzerland, Cham, pp. 308–322. https://doi.org/10.1007/978-3-031-93694-4_22

Chen, H., He, X., Qing, L., Wu, Y., Ren, C., Sheriff, R.E., Zhu, C., 2022: Real-world single image super-resolution: A brief review. Information Fusion 79, 124–145. https://doi.org/10.1016/j.inffus.2021.09.005.

Ledig, C., Theis, L., Huszar, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., Shi, W., 2017: Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. Presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4681–4690.

Li, J., Pei, Z., Li, W., Gao, G., Wang, L., Wang, Y., Zeng, T., 2024: A Systematic Survey of Deep Learning-Based Single-Image Super-Resolution. ACM Comput. Surv. 56, 249:1-249:40. https://doi.org/10.1145/3659100

Lim, B., Son, S., Kim, H., Nah, S., Mu Lee, K., 2017: Enhanced Deep Residual Networks for Single Image Super-Resolution. Presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 136–144.

Mei, Y., Fan, Y., Zhou, Y., Huang, L., Huang, T.S., Shi, H., 2020: Image Super-Resolution With Cross-Scale Non-Local Attention and Exhaustive Self-Exemplars Mining. Presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5690–5699.

Sharifi, A., Shah-Hosseini, R., 2024: Super-Resolution of Sentinel-2 RGB Images with VENµS Reference Images Using SRResNet CNNs. Environmental Sciences Proceedings 29. https://doi.org/10.3390/ECRS2023-16863

Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Qiao, Y., Change Loy, C., 2018: ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks. Presented at the Proceedings of the European Conference on Computer Vision (ECCV) Workshops, pp. 0–0.

Yang, W., Zhang, X., Tian, Y., Wang, W., Xue, J.-H., Liao, Q., 2019: Deep Learning for Single Image Super-Resolution: A Brief Review. IEEE Transactions on Multimedia 21, 3106–3121. https://doi.org/10.1109/TMM.2019.2919431

Yuhas, R.H., Goetz, A.F.H., Boardman, J.W., 1992: Discrimination among semi-arid landscape endmembers using the Spectral Angle Mapper (SAM) algorithm.