Enhancing 3D Building Model Textures with Super-Resolution of Aerial Photographs

Satoko Hattori-Nagao[1], Zhiqian Zhu[1], Shungo Tsutsui[1], Satomi Kakuta[1], Yasuhito Niina[1], Kazuo Oda[1]

[1]Asia Air Survey Co. Ltd., 1-2-2 Manpukuji, Asao-ku, Kawasaki-shi, Kanagawa, Japan
- (stk.hattori, zhi.zhu, sng.tsutsui, stm.kakuta, ysh.niina, kz.oda)@ajiko.co.jp

Keywords: Super-Resolution, SwinIR, LOD2 building textures, Aerial photographs, PLATEAU

Abstract
We have applied a super-resolution technique to enhance the texture image quality of LOD2 building models. Specifically, we adopted SwinIR for upscaling low-resolution images. In order to achieve better results, several approaches for creating training data, consisting of pairs of low-resolution and high-resolution image were investigated. The results showed that training with low-resolution images created by downsampling high-resolution images by a factor of four and then applying blurring and noise improved the sharpness of building edge lines in super-resolution images. Training data with augmentation techniques, such as the use of random noise and random rotation, are proved to be effective in enhancing super-resolution images. Using the super-resolved images, LOD2 building models were created, and a subjective evaluation of the building roof texture quality was conducted. The results indicated that for the input images used in super-resolution, 87% of buildings from high-quality aerial photographs and 78% from lower-quality photographs were rated as having sharp edges without distortion. Even with limited training data, the developed method was able to achieve high-quality super-resolution, regardless of the input image quality, leading to improved texture quality in LOD2 building models.

1. INTRODUCTION

Recently, the development of City Digital Twins, which replicate urban spaces in virtual environments, has been advancing in various countries and cities (Lehtola et al., 2022). A City Digital Twin is composed of 3D city models, with 3D building models being particularly important components. Consequently, there is increasing interest in technologies for the automatic generation and updating of 3D building models.

PLATEAU is a project launched by Japan’s Ministry of Land, Infrastructure, Transport and Tourism (MLIT) in 2020 to promote the use of 3D digital twin models of cities. This project aims to advance the digital transformation of urban development, including the social implementation of smart cities, by establishing, utilizing, and opening to publish 3D city models as foundational data (Seto et al., 2023). As part of this project, efforts are being made to develop systems for the automatic generation and updating of 3D city models to support data development in various cities across Japan. Automatic LOD2 Building Model Generation Tool, which began development in 2022, is a system that automatically generates LOD2 building models using input data such as aerial photographs, DSM, and building footprint data, and outputs the models in CityGML format. LOD2 building models are representations that capture the roof structures of buildings (Figure 1).

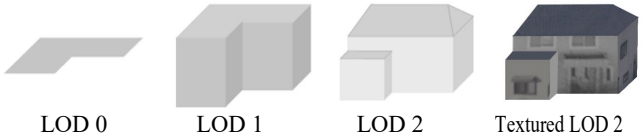

Figure 1. Conceptual diagram of building LOD

The generation of 3D city models in the PLATEAU project primarily utilizes survey data, such as aerial photographs, provided by local governments. The textures of the 3D models generated by Automatic LOD2 Building Model Generation Tool are based on these aerial photographs. As a result, depending on the input data, there may be cases where the resolution is low, leading to concerns about texture quality.
As the utilization of 3D city models expand, the demand for higher quality textures is increasing. Consequently, there is a need to explore methods for enhancing the resolution of these textures to support broader utilization. Enhancing the visibility of 3D building textures is expected to enable more realistic representations of urban landscapes and expand the application of 3D urban models beyond local governments and the mapping industry. Potential uses include visualization for events, interactive advertisements, and entertainment, promoting the development and utilization of 3D city models.

Moreover, increasing the resolution of aerial photographs used in LOD2 building models may contribute to improved accuracy in 3D shape reconstruction. Recent approaches to detecting building footprints and roof edge lines from aerial and satellite imagery increasingly rely on deep learning-based methods (Alidoost et al., 2019), (Yu et al., 2021) and (Jeong and Kim, 2021). Therefore, as the resolution of the input aerial images increases, the performance of automatic roof shape recognition is expected to improve, enabling the automatic generation of more detailed 3D models.

Efforts to apply super-resolution (SR) techniques using deep learning for the extraction of urban features have been actively pursued. Guo et al. (2019) employed SR techniques to generate high-resolution (HR) images from satellite data, improving the accuracy of building semantic segmentation. Panangian et al. (2024) addressed the SR of low-resolution (LR) DSMs, noting their applicability in 3D urban modeling and planning.

In this study, we developed a method to enhance the texture quality of LOD2 building models from aerial photographs by applying SR techniques to upscale HR images. Several cases were investigated regarding the creation of SR training data and training conditions, with a focus on achieving effective SR even with limited training data. Additionally, 3D building models were generated using SR images, and the quality of the building textures was evaluated.

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-G-2025-349-2025

2. METHODOLOGY

2.1 Selection of the SR Method

Single image SR is a technique for reconstructing HR images from a single HR image and is considered a classical problem in computer vision and image processing. As a fundamental issue in image restoration, SR has been applied to various scenarios.

The first SR method using deep learning was SRCNN (Dong et al., 2014) and (Dong et al., 2016), which employed CNNs for feature extraction while learning to minimize the mean squared error between the generated images and the ground truth. This method, however, struggled to restore fine details and often produced blurred images. Subsequently, SRGAN was introduced, achieving 4x SR of input images and generating more photorealistic images through the use of GANs (Ledig et al., 2017). However, one limitation of SRGAN was the appearance of unwanted artifacts when there was a significant difference between the training and testing data.

ESRGAN was proposed by Wang et al. (2018) to address the challenges of SRCNN and SRGAN. This approach improves upon the key components of SRGAN, particularly the network architecture and perceptual loss function. ESRGAN's network introduces the Residual in Residual Dense Block (RRDB) as its fundamental building block. The RRDB is deeper and more complex than SRGAN's residual blocks, based on experimental evidence that increasing layers and connections enhances performance.

SwinIR is an image restoration method based on the Swin Transformer architecture (Liang et al., 2021). The architecture of SwinIR is composed of the following three stages:

i) **Shallow Feature Extraction**: In the first stage, the features of the LR image are extracted using a simple convolutional layer (shallow network).
ii) **Deep Feature Extraction with RSTB**: The core of SwinIR's architecture lies in the deep feature extraction phase, which uses Residual Swin Transformer Blocks (RSTB). The features extracted in step i) are processed through multiple RSTBs, which consist of Swin Transformer layers augmented with residual connections. Each Swin Transformer layer uses shifted windows to perform self-attention, which allows capturing both local and long-range dependencies.
iii) **Image Reconstruction**: After deep feature extraction, SwinIR reconstructs the high-quality output image. A reconstruction layer, typically a series of convolution layers, is applied to the deep features to generate the restored image. This reconstruction layer uses the extracted and processed features from the RSTBs to produce the final output.

Compared to traditional CNN-based methods, SwinIR excels by leveraging local attention and long-range dependencies through shifted windows, enabling it to restore high-quality images with fewer parameters and enhanced efficiency.

While SwinIR demonstrated strong performance in image restoration tasks, its architecture was limited by the window-based self-attention mechanism, which primarily focused on local dependencies. As a result, it struggled to effectively capture long-range dependencies, which are crucial for comprehensive image restoration in HR tasks.

To address the limitations of SwinIR, Hybrid Attention Transformer (HAT) introduces a hybrid attention mechanism that combines channel attention and window-based self-attention (Chen et al., 2023). Additionally, it incorporates an overlapping cross-attention module to enhance the interaction between neighboring window features. This design enables the activation of more pixel information and allows the model to capture long-range dependencies more effectively. Furthermore, HAT leverages a same-task pre-training strategy to improve training efficiency. As a result, HAT outperforms SwinIR, particularly in HR tasks like SR and noise reduction, by utilizing a broader range of pixel information and producing higher-quality restoration results.

Diffusion models were first proposed in 2015, with a refined version introduced in 2020. SR techniques based on diffusion models reconstruct HR images through a denoising process, where noise is incrementally added to HR images and then removed in reverse to recover fine details. Since 2022, specialized methods utilizing diffusion models for SR tasks have emerged.
Notable methods include: Super-Resolution via Repeated Refinements (SR3, Saharia et al., 2022), Latent Diffusion Models (LDMs, Rombach et al., 2022), and DifferIR (Li et al., 2023). Compared to GAN-based techniques, these models have the capability to reduce artifacts and avoid issues such as mode collapse, which are common in GAN-based approaches.

In this study, SwinIR was adopted due to its excellent balance between computational efficiency and performance, making it widely used in industrial applications and various research fields. SwinIR is characterized by its ability to produce high-quality images with fewer artifacts compared to CNN-based methods. Additionally, compared to state-of-the-art approaches such as HAT and diffusion model-based SR, SwinIR offers higher computational efficiency, making it a more practical solution

2.2 Framework of the SR method for LOD2 building textures

Figure 2 illustrates the data flow for aerial image SR and LOD2 building model generation. For SR training, pairs of HR aerial images and their corresponding LR counterparts, obtained by downsampling the HR images by a factor of four, are used to train a SR model.

Next, HR aerial images captured for LOD2 building models are converted into SR images using the trained model. The SR images, along with DSM, building footprints, and interior/exterior orientation parameters, are input into the Automatic LOD2 Building Model Generation Tool to generate textured LOD2 building models.

This system consists of five functions: i) data input functionality, ii) model element generation, iii) topological consistency check and correction, iv) texture image mapping: the function of pasting aerial photographs to LOD2 building models, and v) CityGML output functionality. Each function operates independently, and the input data is processed sequentially from i) to v). The input data for this tool includes aerial photographs (central projection), interior orientation parameters, exterior orientation parameters, DSM, and building footprint from LOD1 CityGML. It is assumed that the DSM is pre-generated from aerial photographs using SfM/MVS software. The tool requires aerial photographs with the following specifications: a ground sampling distance (GSD) of 25 cm, overlap rate of

https://doi.org/10.5194/isprs-annals-X-G-2025-349-2025

≥60%, and side-lap rate of ≥30%. These imaging conditions are typically used by local governments for urban planning surveys. The source code is publicly available on GitHub (PLATEAU, 2022).

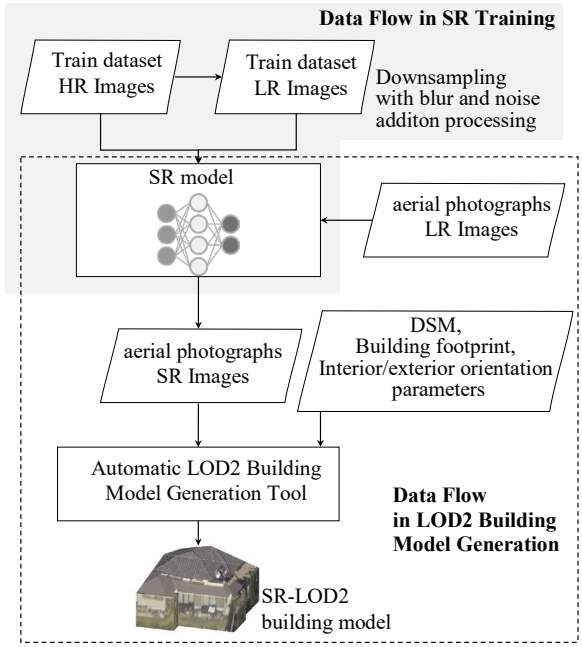

Figure 2. Framework of the SR method for LOD2 building textures.

3. EXPERIMENTS

3.1 Datasets and Preparation

The data used in the experiment consists of aerial photographs taken in six regions within Japan shown in Table 1. In the experiments described in the following Sections 4.1 and 4.2, datasets D1 and D2 were used for SR training, while datasets D1 through D4 were used for training in the LOD2 modeling visibility verification task described in Section 4.3. Data D1 to D4 are for training and validation, while D5 and D6 are for testing. One of the test datasets, D5 (Kawasaki region), has the best image quality (Figure 5a). On the other hand, the other test dataset, D6 (Mitaka B region), has relatively poor quality due to atmospheric conditions with high water vapor (Figure 5g).

#	Train/Val /Test	Region	GSD (cm)	Number of imeges
D1	Train/Val	Mitaka A	7.5	16,928
D2	Train/Val	Nara	7.5	16,928
D3	Train/Val	Itabashi	5.0	7,084
D4	Train/Val	Saitama	5.0	4,476
D5	Test	Kasawaki	25.0	33,793
D6	Test	Mitaka B	25.0	1,190

Table 1. Overview of the datasets used in the experiment

In general, training for SR requires HR images and corresponding HR images downsampled from the HR images by a factor of four. In this study, we downsampled aerial photos with a resolution of 5.0–7.5 cm by a factor of four to create LR images. The size of the LR images was 120 x 120 pixels, while the HR images were 480 x 480 pixels.

In certain cases, such as the data from D6 (Mitaka B region), aerial photographs captured during periods of high atmospheric water vapor can result in blurred or smeared images. To address degraded image quality, we also prepared datasets with simulated blurring and noise effects.

The LR dataset consists of the following types of images:

- **Downsampled images**: HR images (Figure 3a) were downsampled by a factor of four to generate LR images (Figure 3b).
- **Blurred images**: HR images were first downsampled by a factor of 5-6 and then upsampled back to the original 4x scale. For example, an HR image with a resolution of 7.5 cm was first downsampled to 40 cm and then upsampled to 30 cm, creating a blurred effect (Figure 3c).
- **Blurred and noise-added images**: In addition to blurring, Gaussian noise was added. Two versions of the images were created with $\sigma = 1.0$ and $\sigma = 1.5$ noise levels in pixel count, and these were randomly selected for the dataset (Figure 3d).

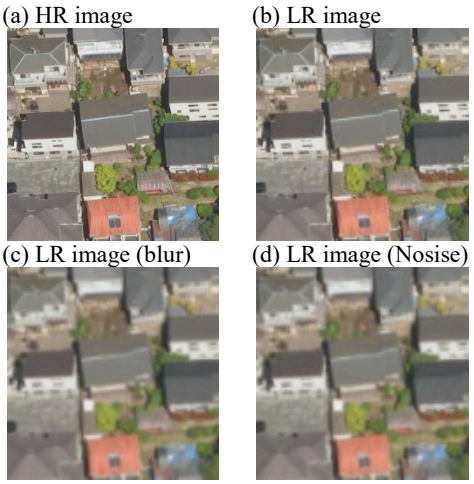

Figure 3. Example of training image: (a) HR image, (b) LR image with only downsampling processing, (c) LR image with blur processing, (d) LR image with blur and noise addition processing

3.2 Training Cases and Conditions

The HR images were paired with the LR images created in Section 3.1, and training was conducted using SwinIR. The training was performed with a batch size of 32, 32 workers, and 200,000 iterations. The Adam optimization algorithm was employed.

Data augmentation included paired random cropping, random horizontal and vertical flips, random transposition, and random color transformation, which were applied to all cases. As optional augmentations, random noise addition and random rotation were applied. The random color transformations were applied randomly to adjust brightness, saturation, hue, and gamma correction. Gaussian noise (σ=15) was used for random noise addition. Table 2 shows the five training cases conducted for comparison.

#	LR image processing		Data augmentation	
	Blur	Noise addition	Random noise	Random rotation
Case 1	-	-	✓	✓
Case 2	✓	-	✓	✓
Case 3	✓	✓	✓	✓
Case 4	✓	✓	✓	-
Case 5	✓	✓	-	✓

Table 2. Training cases

https://doi.org/10.5194/isprs-annals-X-G-2025-349-2025

3.3 Evaluation Method

The model's performance was evaluated both quantitatively and qualitatively with D5 (Kawasaki region) and D6 (Mitaka B region) in Table 1. For the quantitative evaluation, we used two widely adopted metrics in SR research: Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM, Wang et al., 2004), to assess image quality. PSNR measures the ratio between the maximum possible power of a signal and the noise that causes degradation. SSIM assumes that the similarity of image structures plays a significant role in how humans perceive image quality degradation. SSIM compares the luminance, contrast, and structure of the original and decoded images, resulting in a value between 0 and 1, with values closer to 1 indicating higher similarity.

For the qualitative evaluation, we assessed the clarity and resolution of the images, focusing on whether the images were enhanced, and whether the edges of building outlines and ridges were sharpened or distorted. Additionally, we evaluated the quality of the texture images of the LOD2 building models created from SR images (see Section 4.2) based on whether the edges were sharpened and free of distortion. The evaluation was conducted through human assessment using a three-tier ranking system: Rank A (improvement), Rank B (no improvement), and Rank C (deterioration).

The characteristics related to the fineness and sharpness of the edges in the texture of building models could be described as roughness. However, in this study, it refers to a different concept from that defined in ISO 25178, which mainly concerns to the three-dimensional evaluation of physical surface roughness. The assessment was conducted through subjective evaluation.

4. RESULTS AND DISCUSSION

4.1 Quantitative evaluation

The results of the quantitative evaluation are shown in Table 3. The highest PSNR was observed in Case 5, while the highest SSIM was achieved in Case 1. The SR images of the validation data showed good HR reconstruction across all cases, with no significant differences in quality between them (Figure 4). Focusing on the buildings, the SR image quality of Case 1, 2, and 3 is satisfactory, whereas Case 4 shows edge distortions, and Case 5 exhibits artificial noise. These observations are not entirely consistent with the PSNR and SSIM results. The quantitative evaluation metrics used in this study are primarily designed to assess pixel-level similarity in images and do not fully reflect human visual perception. As a result, discrepancies between the quantitative evaluations and visual differences, such as edge distortions and texture details, are observed in the sample images.

#	PSNR	SSIM
Case 1	21.820	**0.691**
Case 2	22.118	0.603
Case 3	21.364	0.568
Case 4	21.488	0.586
Case 5	**23.716**	0.597

Table 3. Quantitative evaluation

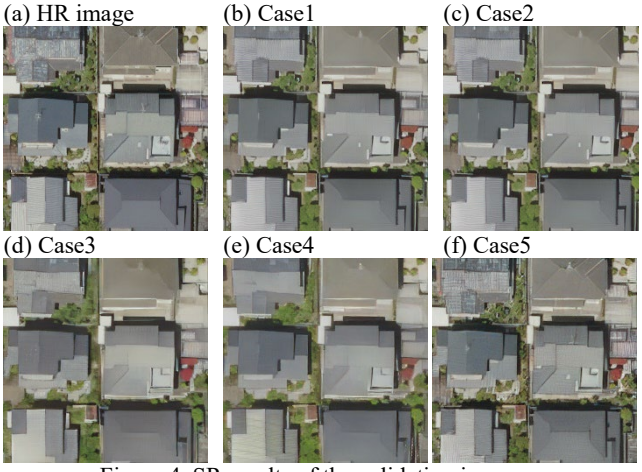

Figure 4. SR results of the validation images

4.2 Quality evaluation

4.2.1 Comparison of Results Based on Input Image Quality

The SR results for the test images are shown in Figure 5. Figure 5a to 4f represent the results for the D5 (Kawasaki region), where the input image quality is high, while Figure 5g to 4l represent D6 (Mitaka B region), where the input image quality is lower. In D5 (Kawasaki region), Case 1 shows successful SR (Figure 5b), and in Cases 2 and 3, the building ridgelines and other edges are sharper than in Case 1 (Figure 5c, d). Although the edges in Cases 4 and D5 are slightly blurred, SR was still achieved (Figure 5e, f). It was observed that when the input image quality is high, the variations in results attributable to different training conditions are minimal.

4.2.2 Comparison of Results Based on Training Conditions for Low-Quality Images

The results based on different training conditions for low-quality images were compared. First, the impact of different methods of creating LR data was examined. In Case 1, the SR image showed only a slight change in brightness, with the overall image remaining blurred and not adequately super-resolved (Figure 5h). In Case 2, where blurring was applied to the LR image, the SR result showed sharper edges (Figure 5i). In Case 3, where both blurring and noise were applied to the LR image, the edges became even sharper, and distortions were minimized (Figure 5j). These results suggest that when the input image is blurred, applying blurring and noise to the LR image can lead to better SR outcomes, with Case 3, which included noise, yielding the best results.

The results based on different data augmentation techniques were compared. In Case 4, where random rotation was not applied as part of the data augmentation, the building edges showed significant distortion, with artificial noise appearing (Figure 5k). Edge distortion tends to occur when the main ridgeline direction of the building does not align with the pixel direction of the image (Figure 6). Figure 6a shows edge distortion in Case 4 ranked in three levels: A, B, and C. Regions where the building roof edges align with the pixel direction (the upper area of Figure 6a, Figure 6b) received higher evaluations, while regions with an angular offset between the pixel direction and roof edges (the lower area of Figure 6a, Figure 6c) received lower evaluations. This can be attributed to insufficient variation in the training data with respect to the orientation of

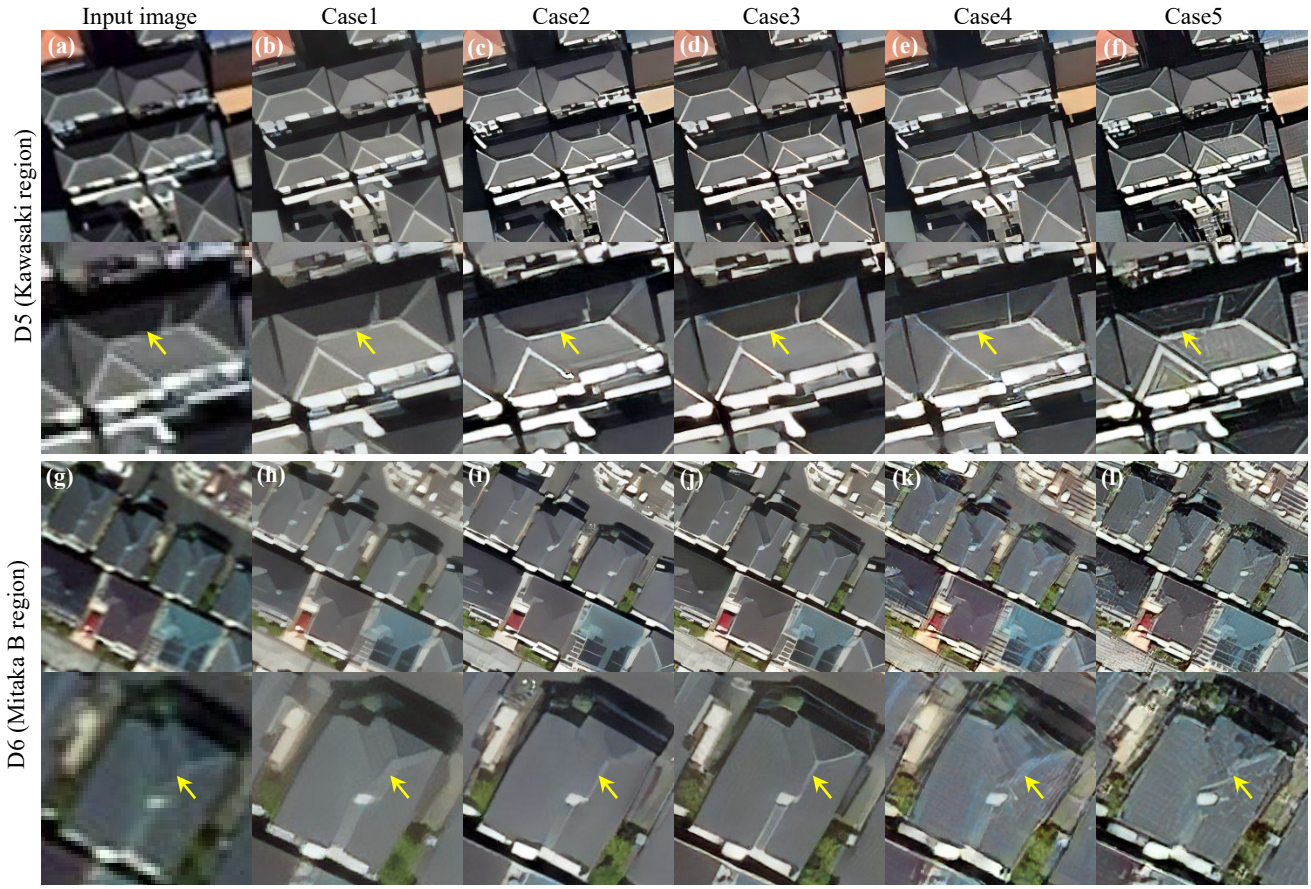

Figure 5. SR results of the test images

building contours and ridgelines. In contrast, Case 3, which included data augmentation with random rotation, showed improved performance, highlighting the importance of incorporating rotation in data augmentation to cover a wide distribution of orientations.

In Case 5, where random noise data augmentation was not applied, significant distortion of the building edges occurred, and artificial noise became more pronounced compared to Case 4 (Figure 5l). It was found that incorporating random noise in data augmentation was highly effective in suppressing noise.

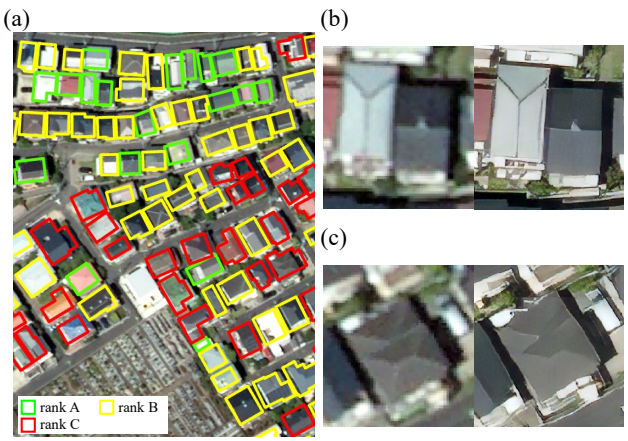

Figure 6. SR results and evaluation of Case 4 (D6): (a) Building edge quality assessment results, (b) Results without edge distortion (left:input, right:SR image), (c) Results with distorted edges (left:input, right:SR image)

4.3 Quality Evaluation of Roof Textures in LOD2 Building Models

LOD2 building models were created using SR images, and the texture quality of the building roofs was evaluated. The 3D models were generated using the Automatic LOD2 Building Model Generation Tool. Case 3, which produced the sharpest edges with minimal distortion, was adopted for training. The evaluation was conducted using two test datasets: D5 (Kawasaki region) with higher image quality, and D6 (Mitaka B region) with lower image quality. The number of buildings evaluated was 544 in D5 and 1133 in D6. Figure 7 shows the models from the D5 (Kawasaki region) validation region, while Figure 8 presents examples of the LOD2 models generated from the original images and the SR images. Compared to LR-LOD2 building models, SR-LOD2 building models demonstrated improved visibility of rooftop features and sharper edges.

Figure 7. Overview of LOD2 Building Models in D5 (Kawasaki region)

https://doi.org/10.5194/isprs-annals-X-G-2025-349-2025

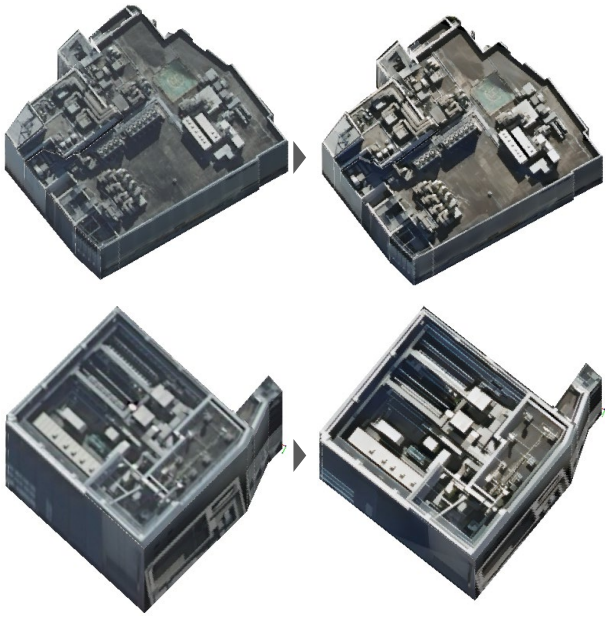

Figure 8. Example of LOD2 building model with SR textures (left: LR-LOD2 building models, right: SR-LOD2 building models)

As a result of the evaluation (Figure 9) with heigher image quality, 87% of the buildings were ranked as A. In D6 (Mitaka B region) with lower image quality, 78% were ranked as A. Although there is a 9-point difference between the two, high-quality SR images with sharp, distortion-free edges were generated at a high rate, regardless of the image quality. Focusing on Rank C, less than 1% of images in D5 (Kawasaki region) fell into this category, compared to 2% in D6 (Mitaka B region). In D6 (Mitaka B region), edge distortions were observed in some buildings.

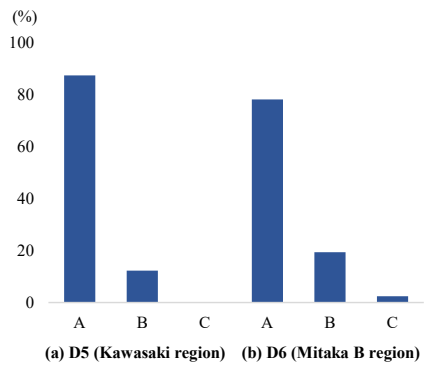

Figure 9. Quality evaluation results of LOD2 building model with SR textures

5. CONCLUSION

In this study, a SR model using SwinIR, specialized for aerial images, was developed to enhance the texture quality of LOD2 building models. We successfully achieved satisfactory SR results, even for low-quality datasets. To accomplish this achievement, we examined different methods for creating LR images used for training and explored various data augmentation techniques, comparing the results. It was observed that applying blurring and noise after downsampling HR images by a factor of four led to sharper results in SR images. Among the data augmentation methods, both random rotation and random noise were found to be particularly effective.

This study successfully improved texture quality through the use of SR images. This study primarily focused on qualitative assessments of image quality. Future work should examine the contribution of SR images to the performance of automatic roofline extraction for creating LOD2 building models. To confirm the generalization capability of the SwinIR model, aerial images from a broader range of regions and images captured under different shooting conditions should be used for validation. In this study, a direct comparison with other advanced SR techniques, such as diffusion-based models, was not conducted. Future research will address this by evaluating the performance and computational efficiency of SwinIR against state-of-the-art methods.

Acknowledgements

This research was conducted as part of Project PLATEAU, promoted by MLIT. We would like to express our deep gratitude to the staff of the City Bureau in MLIT for their valuable advice and guidance during the development of this project.

References

Lehtola, V.V., Koeva, M., Oude Elberink, S., Raposo, P., Virtanen, J.P., Vahdatikhaki, F., Borsci, S., 2022: Digital twin of a city: Review of technology serving city needs. *Int. J. Appl. Earth Obs. Geoinf.*, 114, 102915.

Seto, T., Furuhashi, T., Uchiyama, Y., 2023: ROLE OF 3D CITY MODEL DATA AS OPEN DIGITAL COMMONS: A CASE STUDY OF OPENNESS IN JAPAN'S DIGITAL TWIN "PROJECT PLATEAU". *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLVIII-4/W7, 201-208. doi.org/10.5194/isprs-archives-XLVIII-4-W7-201-2023.

Alidoost, F., Arefi, H., Tombari, F., 2019: 2D Image-To-3D Model_Knowledge-Based 3D Building Reconstruction (3DBR) Using Single Aerial Images and Convolutional Neural Networks (CNNs). *Remote Sens.*, 11(19), 2219.

Yu, D., Ji, S., Liu, J., Wei, S., 2021: Automatic 3D Building Reconstruction from Multi-view Aerial Images with Deep Learning. *ISPRS J. Photogramm. Remote Sens.*, 171, 155-170.

Jeong, D., Kim, Y., 2021: Keypoint-based Deep Learning Approach for Building Footprint Extraction Using Aerial Images. *Korean J. Remote Sens.*, 37(1), 111-122.

Guo, Z., Wu, G., Song, X., Yuan, W., Chen, Q., Zhang, H., 2019: Super-Resolution Integrated Building Semantic Segmentation for Multi-Source Remote Sensing Imagery. *IEEE Access*, 7, 99381-99397. doi.org/10.1109/ACCESS.2019.2928646.

Panangian, D., Bittner, K., 2024: Real-GDSR: Real-World Guided DSM Super-Resolution via Edge-Enhancing Residual Network. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, X-2, 185–192.

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

Dong, C., Loy, C.C., He, K., Tang, X., 2016: Image Super-Resolution Using Deep Convolutional Networks. *IEEE Trans. Pattern Anal. Mach. Intell*., 38(2), 295-307.

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., et al., 2017: Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 4681-4690.

Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Qiao, Y., Loy, C.C., 2018: ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks. *Proc. European Conference on Computer Vision (ECCV) Workshops*, 701-710.

Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., Timofte, R., 2021: SwinIR: Image Restoration Using Swin Transformer. *Proc. IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, 1833-1844.

Chen, Y., Wang, T., Zhou, Z., Qiao, Y., Dong, C., 2023: Activating More Pixels in Image Super-Resolution Transformer. *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 22367-22377.

Saharia, C., Ho, J., Chan, W., Salimans, T., Fleet, D.J., Norouzi, M., 2022: SR3: Image Super-Resolution via Iterative Refinement.
IEEE Trans. Pattern Anal. Mach. Intell., 45(4), 4713-4726.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B., 2022: High-Resolution Image Synthesis with Latent Diffusion Models. *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 10684-10695.

Li, C., Liu, C., Xu, X., Liu, X., 2023: DiffIR: Efficient Diffusion Model for Image Restoration. *Proc. IEEE/CVF International Conference on Computer Vision (ICCV)*, 13095-13105.

Project Plateau, 2024. GitHub Repositories. github.com/Project-PLATEAU/Auto-Create-bldg-lod2-tool (1 Oct 2024).

Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P., 2004: Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Trans. Image Process*., 13(4), 600-612.