# Comparative Study of Edge Losses for Remote Sensing Image Super-Resolution

Minkyung Chung, Youkyung Han

Dept. of Civil Engineering, Seoul National University of Science and Technology, 01811 Seoul, South Korea – mkjung4876@seoultech.ac.kr, han602@seoultech.ac.kr

**Keywords:** Image Super-Resolution, Remote Sensing Image, Deep Learning, Edge Loss, Structural Fidelity.

**Abstract**

Image super-resolution (SR) techniques have achieved significant performance improvements with the advancement of deep learning. Accordingly, deep learning-based SR methods have become the mainstream approach in SR research and are widely applied across various fields, including remote sensing. However, most state-of-the-art SR studies are primarily driven by computer vision research and tend to focus on generating visually realistic images rather than preserving structural fidelity with respect to the input images. In remote sensing applications, maintaining structural fidelity is particularly important because SR outputs are often used in downstream analytical tasks such as object detection.

In this study, we investigate the use of edge loss to enhance the structural fidelity of SR images for remote sensing imagery. The effectiveness of edge loss was evaluated using multiple benchmark datasets on both convolutional neural network (CNN)- and generative adversarial network (GAN)-based SR models. Several representative SR network architectures and GAN training frameworks were employed to assess the impact of integrating edge loss into the training objective. The experimental results demonstrate that incorporating edge loss improves both the structural fidelity and perceptual quality of SR images. Among the evaluated edge operators, the Prewitt-based edge loss showed the most consistent improvements compared with the Sobel- and Laplacian-based edge losses. These results indicate that edge loss is an effective and easily implementable strategy for improving SR reconstruction quality in remote sensing imagery. Furthermore, it can be combined with other edge-aware techniques to further enhance perceptual quality.

## 1. Introduction

Image super-resolution (SR) is an effective approach for reconstructing high-resolution (HR) images from low-resolution (LR) inputs. Recently, deep learning-based SR methods have become the mainstream approach, employing various network architectures such as convolutional neural networks (CNNs) and generative adversarial networks (GANs). However, most state-of-the-art SR studies, driven by computer vision research, focus on generating perceptually realistic images rather than preserving structural fidelity with respect to the input images. This limitation is particularly critical in remote sensing applications, where SR outputs are often used in downstream analytical tasks.

To preserve the structural fidelity of SR images with respect to their input LR images, edge information has been widely incorporated into SR models either explicitly through network architectures or implicitly through training objectives. One common approach involves constructing edge-guided or gradient-guided SR networks that enhance edge-preserving capability during the SR process. These architectures may integrate edge-processing modules within the SR network (Yang et al., 2017; Wang, 2022; Ren et al., 2023), employ parallel edge-processing branches alongside the SR branch (Ma et al., 2021; Shi et al., 2023), or apply edge-refinement subnetworks after SR reconstruction to enhance gradient or edge information in the output images (Rabbi et al., 2020).

Another approach involves incorporating edge- or gradient-related losses into the training objective. These losses are generally formulated to minimize the discrepancy between edge or gradient maps derived from SR images and those from the corresponding HR images (Ma et al., 2021). Such losses are typically combined with existing SR training losses using appropriate weighting, providing additional guidance to suppress structural distortions while maintaining broad applicability across models (Lepcha et al., 2023). A detailed description of the edge loss is provided in Section 2.1.

Although many studies employ edge-aware architectures together with edge-based losses (Yang et al., 2017; Rabbi et al., 2020; Ma et al., 2021; Ren et al., 2023), the individual contribution of edge loss has not been systematically evaluated. Compared with architectural modifications, edge loss can be easily integrated into existing SR models without altering network structures, making it applicable across different model architectures and image specifications, including remote sensing imagery. Therefore, this study focuses on isolating the effect of edge loss by examining losses derived from conventional edge operators to enhance the structural fidelity of SR images. To comprehensively evaluate their effectiveness, comparative experiments are conducted using both CNN- and GAN-based SR models.

## 2. Methodology

### 2.1 Edge Operator-Driven Edge Loss

Previous studies have attempted to enhance the structural fidelity of SR images by incorporating edge-related losses into the training objective. These approaches can be broadly categorized into two types: (1) conventional edge operator-based methods and (2) learning-based methods that utilize pre-trained edge detection networks.

The former approach has been explored in several studies using conventional edge operators such as Sobel, Laplacian, and Canny edge operators. Yang et al. (2017) proposed an edge-guided recurrent residual network that incorporates an edge loss based on the Sobel edge detector with horizontal and vertical gradient kernels. Seif and Androutsos (2018) adopted the Canny edge operator to obtain a binary edge map, where the edge loss was calculated as the mean of the product between the binary edge map and the reconstruction error. Moreover, the Laplacian operator has been used for edge extraction in edge loss computation in SR studies on remote sensing imagery (Rabbi et al., 2020; Ren et al., 2023).

On the other hand, some studies have utilized learning-based approaches that employ pre-trained edge detection networks. Li et al. (2021) fed the generated SR images and the corresponding HR images into the HED network (Xie and Tu, 2015) and measured feature discrepancies at a specific layer. Similarly, Tan et al. (2026) employed the EDTER model (Pu et al., 2022) to compute edge differences between SR outputs and HR images. Although these networks can produce accurate object boundaries by leveraging semantic information, objects in remote sensing imagery are often not characterized by clearly defined edges, which may limit their effectiveness in capturing fine structural details.

In this study, we adopt the former approach by focusing on the effects of conventional edge operators—Sobel, Prewitt, and Laplacian—on SR performance. These operators are well-established and widely used in remote sensing for edge extraction due to their simplicity, computational efficiency, and effectiveness in capturing local intensity gradients.

For the Sobel and Prewitt operators, we extended the conventional horizontal and vertical directions to include two additional diagonal directions, resulting in four directional filters: horizontal (*x*), vertical (*y*), and two diagonal directions (±45°), as illustrated in Figure 1(a) and (b). Unlike common implementations that solely rely on *x* and *y* gradients, this configuration enables a more comprehensive representation of edge structures, which is particularly advantageous for remote sensing imagery, where ground objects often appear in arbitrary orientations. For the Laplacian operator, a 3 × 3 kernel was employed with 8-neighbor pixel connectivity for gradient computation, as shown in Figure 1(c).

## 2.2 Loss Function

The extracted edge maps were used to compute an edge loss that measures the discrepancy between the edges of the SR images ($I_{SR}$) and reference HR images ($I_{HR}$). We examined the effect of these edge losses in both CNN- and GAN-based SR models.

For CNN-based SR models, an edge loss term was incorporated in addition to the pixel-wise L1 loss. The edge loss is computed using the L1 norm for the Sobel and Prewitt operators and the Charbonnier loss for the Laplacian operator, following Rabbi et al. (2020). The overall loss function for CNN-based SR models is formulated as:

$$L_{total}^{CNN} = L_{pix}\left(I_{SR}, I_{HR}\right) + \lambda L_{edge}\left(E\left(I_{SR}\right), E\left(I_{HR}\right)\right) \quad (1)$$

where $L_{pix}$ and $L_{edge}$ denote the pixel-wise and edge losses, respectively. $E(\cdot)$ represents the edge extraction operator and $\lambda$ is the weight of the edge loss, which was set to 0.1 for the Sobel and Prewitt operators and 0.5 for the Laplacian operator.

| | | |
|---|---|---|
| 1 | 2 | 1 |
| 0 | 0 | 0 |
| -1 | -2 | -1 |

$E_{Sobel}^{0°}$

| | | |
|---|---|---|
| 1 | 0 | -1 |
| 2 | 0 | -2 |
| 1 | 0 | -1 |

$E_{Sobel}^{90°}$

| | | |
|---|---|---|
| 2 | 1 | 0 |
| 1 | 0 | -1 |
| 0 | -1 | -2 |

$E_{Sobel}^{45°}$

| | | |
|---|---|---|
| 0 | 1 | 2 |
| -1 | 0 | 1 |
| -2 | -1 | 0 |

$E_{Sobel}^{-45°}$

(a)

| | | |
|---|---|---|
| 1 | 1 | 1 |
| 0 | 0 | 0 |
| -1 | -1 | -1 |

$E_{Prewitt}^{0°}$

| | | |
|---|---|---|
| 1 | 0 | -1 |
| 1 | 0 | -1 |
| 1 | 0 | -1 |

$E_{Prewitt}^{90°}$

| | | |
|---|---|---|
| 1 | 1 | 0 |
| 1 | 0 | -1 |
| 0 | -1 | -1 |

$E_{Prewitt}^{45°}$

| | | |
|---|---|---|
| 0 | 1 | 1 |
| -1 | 0 | 1 |
| -1 | -1 | 0 |

$E_{Prewitt}^{-45°}$

(b)

| | | |
|---|---|---|
| 1 | 1 | 1 |
| 1 | -8 | 1 |
| 1 | 1 | 1 |

$E_{Laplacian}$

(c)

Figure 1. Edge operator kernels used in this study: (a) Sobel kernels (four directions); (b) Prewitt kernels (four directions); (c) Laplacian kernel.

For GAN-based SR models, we followed the adversarial training frameworks of ESRGAN (Wang et al., 2018) and Real-ESRGAN (Wang et al., 2021), in which the total loss consists of pixel-wise, perceptual, and adversarial losses. To investigate the influence of the edge loss, we integrated an edge loss into the total loss as:

$$\begin{aligned} L_{total}^{GAN} &= \alpha L_{pix}\left(I_{SR}, I_{HR}\right) + \beta L_{per}\left(I_{SR}, I_{HR}\right) \\ &+ \gamma L_{adv}\left(I_{SR}, I_{HR}\right) + \mu L_{edge}\left(E\left(I_{SR}\right), E\left(I_{HR}\right)\right) \end{aligned} \quad (2)$$


https://doi.org/10.5194/isprs-archives-XLIX-B3-2026-49-2026 |  

where $L_{per}$ and $L_{adv}$ denote the perceptual and adversarial losses. The weights $\alpha$, $\beta$, and $\gamma$ correspond to the pixel-wise, perceptual, and adversarial losses, respectively. Here, $\mu$ represents the weight for the edge loss, which was set to 0.5 for the Sobel and Prewitt operators and 2.5 for the Laplacian operator. In the ESRGAN configuration, $\alpha$, $\beta$, and $\gamma$ were set to 0.01, 1, and 0.005, respectively, whereas in the Real-ESRGAN configuration, they were set to 1, 1, and 0.1, following the original studies.

# 3. Experiments

## 3.1 Datasets

In this study, experiments were conducted using three publicly available remote sensing image datasets: WHU-RS19 (Dai and Yang, 2010), RSSCN7 (Zou et al., 2015), and AID (Xia et al., 2017), as summarized in Table 1. These benchmark datasets were originally developed for scene classification and contain 19, 7, and 30 classes, respectively. Although they were not specifically designed for SR tasks, they provide HR images covering diverse land-cover types and ground objects, making them suitable for SR training and evaluation.

All images were collected from Google Earth with spatial resolutions ranging from 0.5 m to several meters depending on the scene source. The image sizes are 600 × 600 pixels for WHU-RS19 and AID, and 400 × 400 pixels for RSSCN7.

Since SR performance is strongly influenced by structural complexity in the training data, scenes with limited edge or texture information may bias reconstruction toward overly smooth results and weaken discriminative learning. Therefore, classes containing minimal structural content were excluded prior to dataset splitting. Specifically, the *beach*, *desert*, and *meadow* classes were removed from WHU-RS19 and the *bare land*, *beach*, *desert*, and *meadow* classes were excluded from AID, whereas no classes were removed from RSSCN7. After exclusion, 844 images from WHU-RS19 and 8,710 images from AID were used for experiments. Each dataset was then randomly divided into training, validation, and test sets with a ratio of 6:2:2.

| Dataset | Image Source | Image Size | Total Images |
|---|---|---|---|
| WHU-RS19 | Google Earth | 600 × 600 | 1,005 |
| RSSCN7 | Google Earth | 400 × 400 | 2,800 |
| AID | Google Earth | 600 × 600 | 10,000 |

Table 1. Summary of datasets used in this study.

## 3.2 Implementation Details

For training and evaluation, LR images were synthetically generated by applying bicubic downsampling with a scale factor of 4 to the HR images in all experiments. During CNN-based SR training, LR image patches of size 64 × 64 pixels were randomly cropped in each iteration, with a batch size of 4, and augmented using horizontal and vertical flips as well as rotations of 90°, 180°, and 270°. The networks were trained using the Adam optimizer with parameters $\beta_1$ = 0.9, $\beta_2$ = 0.999, and $\varepsilon = 10^{-8}$. The initial learning rate was set to $10^{-4}$ and reduced by half at the midpoint of the total training epochs.

For GAN-based SR models, LR image patches of size 32 × 32 pixels were used with a batch size of 8. The generator networks were initialized with pre-trained weights using pixel-wise loss. Both the generator and discriminator were trained using the Adam optimizer with the same parameters as in CNN-based SR training and the learning rate was set to $10^{-4}$.

All SR models were trained from scratch on each dataset to ensure a fair comparison. The experiments were implemented in PyTorch 2.6.0 with CUDA Toolkit 12.6 and conducted on an NVIDIA GeForce RTX 4080 GPU.

## 3.3 Evaluation Metrics

The SR performance was evaluated using three widely used image quality assessment metrics: peak signal-to-noise ratio (PSNR), structural similarity index measure (SSIM) (Wang et al., 2004), and learned perceptual image patch similarity (LPIPS) (Zhang et al., 2018a).

PSNR and SSIM are conventional full-reference metrics that measure image fidelity between the reconstructed image ($I_{SR}$) and the ground-truth HR image ($I_{HR}$). PSNR is defined as

$$\mathrm{PSNR}(I_{SR}, I_{HR}) = 10\log_{10}\frac{MAX^2}{MSE(I_{SR}, I_{HR})} \tag{3}$$

where $MAX$ denotes the maximum pixel value of the image and $MSE$ is the mean squared error between $I_{SR}$ and $I_{HR}$.

SSIM evaluates structural similarity by comparing luminance, contrast, and structural properties:

$$\mathrm{SSIM}(I_{SR}, I_{HR}) = \frac{\left(2\mu_{I_{SR}}\mu_{I_{HR}} + c_1\right)\left(2\sigma_{I_{SR}I_{HR}} + c_2\right)}{\left(\mu_{I_{SR}}^2 + \mu_{I_{HR}}^2 + c_1\right)\left(\sigma_{I_{SR}}^2 + \sigma_{I_{HR}}^2 + c_2\right)} \tag{4}$$

where $\mu$ and $\sigma$ denote the mean and standard deviation of the images, respectively, and $\sigma_{I_{SR}I_{HR}}$ represents the cross-covariance between $I_{SR}$ and $I_{HR}$. The constants $c_1$ and $c_2$ are included to prevent division by zero.

While PSNR and SSIM primarily measure image fidelity and structural similarity, they do not fully reflect the perceptual quality of SR results. To complement these metrics, LPIPS was employed to assess perceptual similarity between $I_{SR}$ and $I_{HR}$. LPIPS is computed by measuring the distance between deep feature representations extracted from a pre-trained network:

$$\mathrm{LPIPS}(I_{SR}, I_{HR}) = \sum_l \frac{1}{H_l W_l C_l}\sum_{h,w}\left\| w_i \odot \left( f_{h,w,c}^l - f_{0,h,w,c}^l \right)\right\|_2^2 \tag{5}$$

where $f^l_{h,w,c}$ and $f^l_{0,h,w,c}$ denote the features extracted from the $l$-th layer at locations ($h$, $w$, $c$) in the images $I_{SR}$ and $I_{HR}$, respectively. $H_l$, $W_l$, and $C_l$ represent the height, width, and number of channels of the feature map at layer $l$. The vector $w_l$ is a learned weight and ⊙ denotes the element-wise product. In this study, a pre-trained VGG-16 network was adopted to compute LPIPS. Higher PSNR and SSIM values indicate better image and structural fidelity, whereas lower LPIPS values correspond to improved perceptual quality.

# 4. Results & Discussion

## 4.1 Edge Loss Effect Analysis in the CNN-Based SR Model

To examine the effect of edge loss on SR training, three representative CNN-based SR architectures were employed: RRDBNet (Wang et al., 2018), RCAN (Zhang et al., 2018b),

| Dataset | Edge Operator for $L_{edge}$ | CNN-Based SR Model | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | RRDBNet | | | RDN | | | RCAN | | |
| | | PSNR (dB) | SSIM | LPIPS | PSNR (dB) | SSIM | LPIPS | PSNR (dB) | SSIM | LPIPS |
| WHU-RS19 | (Baseline) | 25.6461 | 0.7378 | 0.3346 | 26.0178 | 0.7526 | 0.3071 | 25.9480 | 0.7494 | 0.3086 |
| | Sobel | **25.7676** | **0.7426** | **0.3238** | 25.9926 | 0.7542 | **0.3030** | 25.9622 | **0.7519** | 0.3056 |
| | Prewitt | 25.7307 | 0.7407 | 0.3284 | 26.0356 | **0.7546** | 0.3034 | 25.9671 | 0.7511 | **0.3043** |
| | Laplacian | 25.7141 | 0.7384 | 0.3319 | **26.0540** | 0.7533 | 0.3078 | **25.9672** | 0.7498 | 0.3094 |
| RSSCN7 | (Baseline) | 26.0167 | 0.6704 | 0.4004 | 26.3217 | 0.6860 | 0.3642 | 26.2758 | 0.6822 | 0.3691 |
| | Sobel | 26.0514 | 0.6739 | **0.3815** | 26.3147 | 0.6882 | **0.3535** | 26.2824 | **0.6854** | **0.3587** |
| | Prewitt | **26.0933** | **0.6752** | 0.3883 | 26.3550 | **0.6886** | 0.3546 | **26.2910** | 0.6843 | 0.3615 |
| | Laplacian | 26.0844 | 0.6709 | 0.3976 | **26.3709** | 0.6864 | 0.3650 | 26.2876 | 0.6820 | 0.3681 |
| AID | (Baseline) | 26.2828 | 0.7281 | 0.3616 | 26.6237 | 0.7438 | 0.3315 | 26.5612 | 0.7389 | 0.3318 |
| | Sobel | 26.2936 | 0.7309 | 0.3528 | 26.6599 | 0.7458 | **0.3257** | 26.6013 | 0.7431 | **0.3272** |
| | Prewitt | **26.3361** | **0.7319** | **0.3510** | **26.6611** | **0.7459** | 0.3273 | 26.6064 | **0.7432** | 0.3276 |
| | Laplacian | 26.3094 | 0.7298 | 0.3608 | 26.6582 | 0.7448 | 0.3317 | **26.6525** | 0.7426 | 0.3306 |

Table 2. Effect of edge loss with different edge operators on CNN-based SR models. The best results for each dataset and model are highlighted in bold. For models trained with edge loss, performance degradations relative to the baseline are marked in red.

and RDN (Zhang et al., 2018c). RRDBNet utilizes residual-in-residual dense blocks, which integrate dense connections into multi-level residual structures to enhance network capacity. Similarly, RDN employs residual dense blocks to effectively extract hierarchical features through dense connections and local residual learning. RCAN incorporates a channel attention mechanism within residual groups to emphasize informative channel-wise features. These architectures remain widely used backbone models in SR research and were therefore selected to evaluate the impact of edge loss across different SR networks. All models were trained using either the conventional L1 loss (baseline) or L1 loss combined with an additional edge loss, as described in (1).

As shown in Table 2, incorporating edge loss generally improved SR performance across all datasets, producing higher PSNR and SSIM values and lower LPIPS scores compared with the baseline models. For RRDBNet, the integration of edge loss consistently enhanced all evaluation metrics. Similar performance improvements were observed for RDN and RCAN, indicating that the benefits of edge loss are not restricted to a specific network architecture. Among the evaluated CNN-based SR models, RDN generally achieved the highest structural fidelity and perceptual quality, indicating stable SR performance across the datasets.

Among the edge operators used for edge loss, the Sobel and Prewitt operators produced relatively stronger improvements in SSIM and LPIPS, whereas the Laplacian operator showed comparatively smaller gains. These trends were observed across the evaluated CNN-based SR models. In particular, the Prewitt-based edge loss demonstrated consistent improvements across datasets and evaluation metrics, suggesting stable behavior during SR training.

These differences can be explained by the gradient characteristics of the edge operators. The Sobel and Prewitt operators estimate directional gradients using first-order derivatives, which provide relatively stable edge representations and help preserve structural information during training. In contrast, the Laplacian operator relies on a second-order derivative, making it more sensitive to high-frequency intensity variations and noise. As a result, Laplacian-based edge loss may emphasize local intensity changes, which can increase pixel-level sharpness and lead to higher PSNR values in some cases, while not necessarily improving perceptual quality, as reflected by higher LPIPS scores. In addition, the Sobel operator occasionally produced slight decreases in PSNR, which may be attributed to its stronger gradient weighting that emphasizes edge transitions during optimization. Overall, the quantitative results demonstrate that incorporating edge loss can enhance the structural fidelity and perceptual quality of SR outputs in CNN-based SR models, with the Prewitt operator providing the most consistent improvements across different architectures and datasets.

### 4.2 Edge Loss Effect Analysis in the GAN-Based SR Model

In this section, the CNN-based SR networks used in Section 4.1 were employed as generators for GAN-based SR training. The effect of edge loss was evaluated under two widely adopted GAN-based SR frameworks: ESRGAN (Wang et al., 2018) and Real-ESRGAN (Wang et al., 2021).

As one of the pioneering GAN-based SR methods, ESRGAN improves the adversarial training of SRGAN (Ledig et al., 2017) by adopting the relativistic average GAN (RaGAN) (Jolicoeur-Martineau, 2019), which estimates the relative probability between real and generated images. In addition, ESRGAN enhances the SRGAN generator architecture by introducing residual-in-residual dense blocks to increase network capacity.

Building upon ESRGAN, Real-ESRGAN introduces several improvements, including the replacement of the VGG-style discriminator with a U-Net-based architecture that incorporates skip connections to provide more detailed pixel-wise feedback to the generator. Moreover, spectral normalization is applied to stabilize adversarial training and suppress artifacts commonly introduced during GAN training. In this study, the high-order degradation model proposed in Real-ESRGAN was not adopted to isolate the effect of edge loss under a controlled and consistent degradation setting.

Based on the results in Section 4.1, the Prewitt operator was selected to compute the edge loss due to its consistent performance across datasets and evaluation metrics. To evaluate its effectiveness, models trained with edge loss were compared

| Dataset | Edge Operator for $L_{edge}$ | Generator Architecture | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | RRDBNet | | | RDN | | | RCAN | | |
| | | PSNR (dB) | SSIM | LPIPS | PSNR (dB) | SSIM | LPIPS | PSNR (dB) | SSIM | LPIPS |
| WHU-RS19 | (Baseline) | 23.6654 | 0.6397 | 0.3160 | 24.4380 | 0.6791 | 0.2961 | 24.2303 | 0.6763 | 0.2974 |
| | Prewitt | 23.8442 | 0.6595 | 0.3128 | 24.5131 | 0.6864 | 0.2869 | 24.3156 | 0.6718 | 0.2921 |
| RSSCN7 | (Baseline) | 23.7776 | 0.5643 | 0.3669 | 24.3858 | 0.5866 | 0.3423 | 24.3359 | 0.5773 | 0.3482 |
| | Prewitt | 24.4757 | 0.5909 | 0.3620 | 24.4914 | 0.5902 | 0.3363 | 24.4803 | 0.5871 | 0.3422 |
| AID | (Baseline) | 24.0255 | 0.6386 | 0.3381 | 24.7203 | 0.6597 | 0.3147 | 24.8734 | 0.6622 | 0.3119 |
| | Prewitt | 24.5486 | 0.6595 | 0.3362 | 25.0592 | 0.6738 | 0.3006 | 24.9759 | 0.6704 | 0.3135 |

Table 3. Effect of the Prewitt-based edge loss on GAN-based SR models with different generator architectures under the ESRGAN training framework. For models trained with edge loss, performance degradations relative to the baseline are marked in red.

| Dataset | Edge Operator for $L_{edge}$ | Generator Architecture | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | RRDBNet | | | RDN | | | RCAN | | |
| | | PSNR (dB) | SSIM | LPIPS | PSNR (dB) | SSIM | LPIPS | PSNR (dB) | SSIM | LPIPS |
| WHU-RS19 | (Baseline) | 24.5311 | 0.6808 | 0.2925 | 24.9868 | 0.6920 | 0.2720 | 24.7002 | 0.6799 | 0.2805 |
| | Prewitt | 24.6196 | 0.6837 | 0.2871 | 25.0467 | 0.6975 | 0.2711 | 24.7655 | 0.6851 | 0.2773 |
| RSSCN7 | (Baseline) | 24.8393 | 0.6031 | 0.3348 | 25.1885 | 0.6167 | 0.3138 | 24.9035 | 0.6060 | 0.3211 |
| | Prewitt | 25.0141 | 0.6054 | 0.3349 | 25.2558 | 0.6233 | 0.3119 | 25.2403 | 0.6163 | 0.3177 |
| AID | (Baseline) | 25.3412 | 0.6749 | 0.3070 | 25.6137 | 0.6932 | 0.2864 | 25.6568 | 0.6911 | 0.2925 |
| | Prewitt | 25.4778 | 0.6811 | 0.3078 | 25.6606 | 0.6903 | 0.2849 | 25.7572 | 0.6922 | 0.2916 |

Table 4. Effect of the Prewitt-based edge loss on GAN-based SR models with different generator architectures under the Real-ESRGAN training framework. For models trained with edge loss, performance degradations relative to the baseline are marked in red.

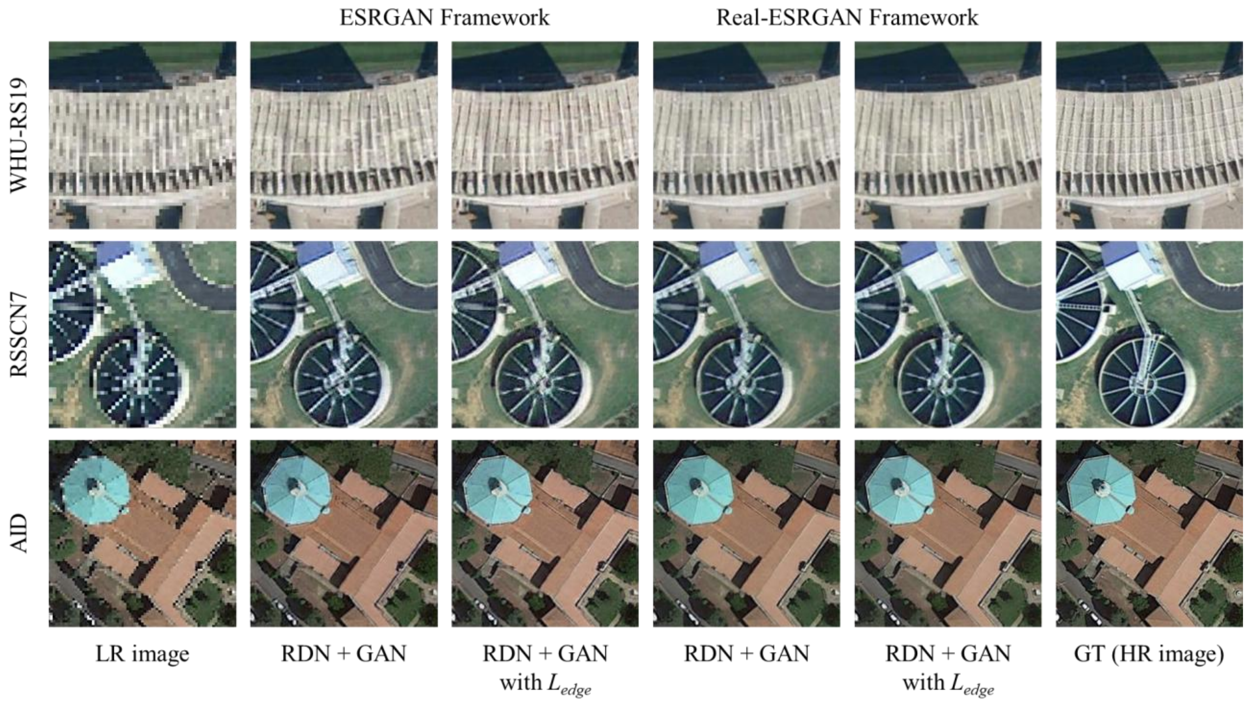


Figure 2. Visual comparison of SR outputs generated by GAN-based SR models with the RDN generator architecture under the ESRGAN and Real-ESRGAN frameworks. Results from baseline models (trained without edge loss) and models trained with the Prewitt-based edge loss are shown for the WHU-RS19, RSSCN7, and AID datasets, arranged from top to bottom by row.

with baseline models trained without edge loss under the same training settings.

The quantitative results for the ESRGAN and Real-ESRGAN frameworks are presented in Tables 3 and 4, respectively. The results indicate that incorporating the Prewitt-based edge loss improves structural fidelity and perceptual quality across all datasets in both GAN frameworks, demonstrating that the effectiveness of edge loss extends to both CNN- and GAN-based SR training. Similar tendencies to those observed in the

CNN-based SR models were also found in the GAN-based SR models. In most cases, the RDN generator achieved superior performance across the evaluation metrics compared with RRDBNet and RCAN. Furthermore, the introduction of edge loss noticeably increases PSNR and SSIM values in GAN-based SR models, mitigating the degradation in image and structural fidelity that commonly arises from incorporating adversarial learning into SR training.

Under the ESRGAN framework (Table 3), the integration of edge loss resulted in noticeable PSNR increases for the RRDBNet generator, particularly for the RSSCN7 and AID datasets, where the gains exceeded 0.5 dB. While the PSNR improvements for RDN and RCAN were relatively smaller than those of RRDBNet, RDN showed the largest reductions in LPIPS scores across the datasets when trained with edge loss. Overall, the RDN generator achieved the best performance in terms of structural fidelity and perceptual quality.

A similar tendency was observed under the Real-ESRGAN framework (Table 4), where the SR results with edge loss also showed improvements across all evaluation metrics. Compared with the ESRGAN framework, the Real-ESRGAN framework achieved higher overall performance, with average gains of 0.75 dB in PSNR and 0.024 in SSIM, as well as an average reduction of 0.024 in LPIPS. Although the Real-ESRGAN configuration already provides stronger baseline SR performance, additional benefits from the incorporation of edge loss were still observed, albeit with smaller margins.

Figure 2 presents a visual comparison of SR outputs generated by GAN-based SR models with the RDN generator architecture under both frameworks. The first row shows examples from the WHU-RS19 dataset, illustrating a zoomed-in view of a stadium structure, where SR images produced by models trained with the Prewitt-based edge loss more clearly delineate structural details. The second row presents examples from the RSSCN7 dataset, depicting industrial facilities, while the third row shows examples from the *church* class of the AID dataset. In the RSSCN7 and AID examples, the visual improvements resulting from the integration of edge loss are more noticeable under the ESRGAN framework, particularly in terms of reduced artifacts and clearer edge structures, which is consistent with the larger quantitative improvements observed in Tables 3 and 4.

In summary, both quantitative and qualitative results confirm that incorporating edge loss into the total optimization objective improves GAN-based SR training and enhances the quality of reconstructed images. However, the effectiveness of edge loss appears to be influenced by the baseline performance of the GAN framework, with stronger frameworks exhibiting relatively smaller performance gains, particularly in perceptual quality. Therefore, as the benefits of edge loss are primarily associated with enhanced image and structural fidelity, it can be combined with complementary strategies, such as architectural approaches that explicitly incorporate edge-processing modules or auxiliary edge-supervised branches, to further improve perceptual quality.

## 5. Conclusion

This study investigated the impact of edge losses on SR models for remote sensing imagery using multiple benchmark datasets. The experimental results demonstrate that edge losses derived from conventional edge operators enhance both the structural fidelity and perceptual quality of SR images. In particular, the Prewitt operator provided the most robust and consistent improvements across both CNN- and GAN-based SR models. Moreover, these edge losses can be easily integrated into state-of-the-art SR models without introducing significant computational overhead. These findings indicate that incorporating edge-aware constraints into the SR optimization objective is an effective strategy for improving reconstruction quality in remote sensing imagery. Since edge loss primarily improves structural fidelity, it can be further combined with explicit edge-processing strategies, such as architectural approaches, to enhance the perceptual quality of SR images.

## Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2025-00523183).

## References

Dai, D., Yang, W., 2010. Satellite image classification via two-layer sparse coding with biased image representation. *IEEE Geosci. Remote Sens. Lett.*, 8(1), 173–176. doi.org/10.1109/LGRS.2010.2055033.

Jolicoeur-Martineau, A., 2019. The relativistic discriminator: A key element missing from standard GAN. *Proc. Int. Conf. Learn. Represent. (ICLR)*, New Orleans, LA, USA, 6–9 May 2019.

Ledig, C., Theis, L., Huszár, 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. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, 21–26 July 2017, 105–114. doi.org/10.1109/CVPR.2017.19.

Lepcha, D.C., Goyal, B., Dogra, A., Goyal, V., 2023. Image super-resolution: A comprehensive review, recent trends, challenges and applications. *Inf. Fusion*, 91, 230–260. doi.org/10.1016/j.inffus.2022.10.007.

Li, J., Liu, W., Zhang, K., Liu, B., 2021. Edge loss for remote sensing image super-resolution. In: *Proc. CECNet 2021*, 262–267. IOS Press. doi.org/10.3233/FAIA210411.

Ma, C., Rao, Y., Lu, J., Zhou, J., 2021. Structure-preserving image super-resolution. *IEEE Trans. Pattern Anal. Mach. Intell.*, 44(11), 7898–7911. doi.org/10.1109/TPAMI.2021.3114428.

Pu, M., Huang, Y., Liu, Y., Guan, Q., Ling, H., 2022. EDTER: Edge detection with transformer. *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, New Orleans, LA, USA, 18–24 June 2022, 1392–1402. doi.org/10.1109/CVPR52688.2022.00146.

Rabbi, J., Ray, N., Schubert, M., Chowdhury, S., Chao, D., 2020. Small-object detection in remote sensing images with end-to-end edge-enhanced GAN and object detector network. *Remote Sens.*, 12(9), 1432. doi.org/10.3390/rs12091432.

Ren, Z., He, L., Lu, J., 2023. Context aware edge-enhanced GAN for remote sensing image super-resolution. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, 17, 1363–1376. doi.org/10.1109/JSTARS.2023.3333271.

Seif, G., Androutsos, D., 2018. Edge-based loss function for single image super-resolution. *Proc. IEEE Int. Conf. Acoustics,* *Speech and Signal Process. (ICASSP)*, 1468–1472. doi.org/10.1109/ICASSP.2018.8461664.

Shi, W., Tao, F., Wen, Y., 2023. Structure-aware deep networks and pixel-level generative adversarial training for single image super-resolution. *IEEE Trans. Instrum. Meas.*, 72, 1–14. doi.org/10.1109/TIM.2023.3246523.

Tan, X., Wu, F., Li, Y., Ye, W., 2026. Progressive hybrid transformer of image super-resolution for precision measurement. *Measurement*, 257(A), 118642. doi.org/10.1016/j.measurement.2025.118642.

Wang, X., Xie, L., Dong, C., Shan, Y., 2021. Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data. *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, Montreal, BC, Canada, 11–17 Oct. 2021, 1905–1914. doi.org/10.1109/ICCVW54120.2021.00217.

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. In: Leal-Taixé, L., Roth, S. (Eds.), *Computer Vision – ECCV 2018 Workshops*, Vol. 11133, 63–79. Springer, Cham. doi.org/10.1007/978-3-030-11021-5_5.

Wang, Y., 2022. Edge-enhanced feature distillation network for efficient super-resolution. *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Worksh. (CVPRW)*, New Orleans, LA, USA, 19–20 June 2022, 776–784. doi.org/10.1109/CVPRW56347.2022.00093.

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. doi.org/10.1109/TIP.2003.819861.

Xia, G.-S., Hu, J., Hu, F., Shi, B., Bai, X., Zhong, Y., 2017. AID: A benchmark dataset for performance evaluation of aerial scene classification. *IEEE Trans. Geosci. Remote Sens.*, 55(7), 3965–3981. doi.org/10.1109/TGRS.2017.2685945.

Xie, S., Tu, Z., 2015. Holistically-nested edge detection. *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Santiago, Chile, 7–13 Dec. 2015, 1395–1403. doi.org/10.1109/ICCV.2015.164.

Yang, W., Feng, J., Yang, J., Zhao, F., Liu, J., Guo, Z., Yan, S., 2017. Deep edge guided recurrent residual learning for image super-resolution. *IEEE Trans. Image Process.*, 26(12), 5895–5907. doi.org/10.1109/TIP.2017.2750403.

Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O., 2018a. The unreasonable effectiveness of deep features as a perceptual metric. *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Salt Lake City, UT, USA, 18–23 June 2018, 586–595. doi.org/10.1109/CVPR.2018.00068.

Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., Fu, Y., 2018b. Image super-resolution using very deep residual channel attention networks. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (Eds.), *Computer Vision – ECCV 2018*, Vol. 11211, 294–310. Springer, Cham. doi.org/10.1007/978-3-030-01234-2_18.

Zhang, Y., Tian, Y., Kong, Y., Zhong, B., Fu, Y., 2018c. Residual dense network for image super-resolution. *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Salt Lake City, UT, USA, 18–23 June 2018, 2472–2481. doi.org/10.1109/CVPR.2018.00262.

Zou, Q., Ni, L., Zhang, T., Wang, Q., 2015. Deep learning based feature selection for remote sensing scene classification. *IEEE Geosci. Remote Sens. Lett.*, 12(11), 2321–2325. doi.org/10.1109/LGRS.2015.2475299.