Intelligent GAN-Based Framework for High-Resolution Satellite Imagery Enhancement

Victor Sineglazov[1], Artem Holovachov[2]

[1]State university "Kyiv aviation institute", 1 Lubomyra Huzara ave., Kyiv, 03058, Ukraine
[2]National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", 37 Prospect Beresteiskyi, Kyiv, 03056, Ukraine

Abstract

We propose a novel design that incorporates a dual-discriminator strategy: one focused on ensuring global image consistency, and the other on preserving local texture accuracy. Experiments conducted on benchmark datasets demonstrate that our approach surpasses prior solutions both quantitatively and in terms of perceptual quality. The proposed architecture is particularly effective at recovering fine textures and sharp edges while maintaining a natural appearance. We investigate the influence of different loss function combinations and training strategies on the fidelity-perceptual quality trade-off. Our findings contribute to the growing field of deep learning-based super-resolution, offering practical insights into architectural refinements that balance computational efficiency with output quality.

Keywords

super-resolution, generative adversarial networks, satellite imagery, deep learning, dual discriminator, aerial imagery

1. Introduction

The super-resolution (SR) problem implies the task of recovering low-resolution (LR) images to their high-resolution (HR) counterpart. The technology has been used in a wide spectrum of applications: medical imaging, satellite photography, and even surveillance systems. An especially important domain is unmanned aerial vehicle (UAV) navigation, where high-quality visual data directly influences the accuracy and reliability of flight control and decision-making. Previous research highlights the role of intelligent visual navigation systems of high accuracy [1], suppression of noise in visual navigation systems [2], landmarks-based navigation software [3], and integrated complexes for UAV detection and identification [4]. These studies emphasize the demand for enhanced image clarity and robustness, making SR techniques highly relevant for UAV applications. Despite its recent emergence, this field has already accumulated a variety of methods.

In this paper, we address limitations of existing methods by proposing a novel architecture that extends recent generative adversarial network (GAN) based image synthesis methods and incorporates extensions for super-resolution tasks. Our approach attempts to overcome the present limitations through a dual-discriminator architecture that significantly improves the quantitative and qualitative attributes of super-resolution outputs.

Overall, our main contributions: (1) We propose dual-discriminator architecture which simultaneously addresses two different aspects of image reconstruction: local features precision and global image coherence. With this design, our model is able to generate high-resolution outputs with both fine-grained texture details and general image consistency. (2) In advance, we propose to conduct an investigation of different loss functions to determine which one has the most impact on GAN performance. This analysis provides insightful information on the inherent trade-offs between conflicting optimization objectives and has practical relevance to architecture design decisions.

ProfIT AI'25: 5th International Workshop of IT-professionals on Artificial Intelligence, October 15–17, 2025, Liverpool, UK
svm@kai.edu.ua (V. Sineglazov); gas.ua.ztl@gmail.com (A. Holovachov)
0000-0002-3297-9060 (V. Sineglazov); 0009-0008-5847-6307 (A. Holovachov)

1.1. Traditional Super-Resolution Techniques

The traditional techniques for super-resolution are mostly consist of different interpolation techniques such as bicubic, bilinear, and nearest neighbor algorithms [5, 6]. While these techniques are computationally efficient, they tend to generate blurry images with insufficient texture detail. More advanced techniques employed example-based methods [7, 8] and sparse coding models[9, 10], which were superior but poor in reconstructing complex patterns.

1.2. Deep Learning for Super-Resolution

Implementation of deep learning has led to significant advancements in the field of image super-resolution [11]. The initial work was carried out by Dong et al. [12] using SRCNN, it demonstrated that a moderately deep convolutional neural network could outperform traditional methods. This was followed by huge number of CNN-based approaches. Kim et al. [13] presented VDSR, which employed residual learning and much deeper architectures to further improve the performance. Shi et al. [14] proposed ESPCN with efficient sub-pixel convolution layer for upscaling with less computational complexity by extracting features directly from low-resolution inputs.

With improvements in network architectures, Lai et al. [15] introduced LapSRN, a progressive reconstruction approach based on a Laplacian pyramid architecture. Lim et al. [16] created EDSR, which removed the redundant modules from base ResNet architectures and significantly expanded network capacity. Zhang et al. [17] created RDN, which combined residual and dense connections to allow feature extraction at multiple levels. These works demonstrated that neural network architectures could achieve significant improvements in reconstruction quality, particularly with fine-tuning for pixel-wise accuracy metrics such as PSNR and SSIM.

1.3. GAN-based Super-Resolution

Generative Adversarial Networks resulted in a huge leap in super-resolution research. Ledig et al. [18] introduced SRGAN, the first GAN-based approach to photo-realistic super-resolution, with perceptual loss functions along with adversarial training. ESRGAN [19] also refined this approach with residual in residual dense blocks and relativistic adversarial loss. Zhang et al. [20] proposed SFTGAN, which incorporated semantic segmentation information to promote structural consistency. Despite these refinements, GAN-based methods still have a tendency to suffer from loss of small details quality and are hard to train.

1.4. Dual-Discriminator Approaches

Recent research has shown that the utilization of more than one discriminator can enhance GAN performance on various tasks. For example, in the field of image-to-image translation, Isola et al. [21] employed a PatchGAN discriminator to successfully model local textures. Taking this concept further, Wang et al. [22] introduced multi-scale discriminators for high-resolution image synthesis. In the context of super-resolution, Sajjadi et al. [23] utilized a texture discriminator alongside a standard discriminator, but focused primarily on texture transfer rather than structural coherence.

Our contribution consist in implementation of dual discriminator, designed to capture both local texture coherence and global image consistency. The local discriminator operates on randomly sampled patches, ensuring realistic textures and preserving fine details, while the global discriminator evaluates the overall composition and perceptual quality of the entire image. By combining these complementary perspectives, the approach effectively overcomes the typical trade-off between sharp, realistic textures and coherent global structure.

1.5. Loss Functions

The choice of loss function plays crucial role in determining model performance. Conventional pixel-wise loss functions like L1 and L2 are effective for optimizing PSNR but often generate overly smooth and

less detailed outputs. Johnson et al. [24] initiated perceptual loss, which leverages feature activations from pre-trained network to capture semantic information beyond pixel-level accuracy. Wang et al. [25] proposed the LPIPS metric, which has good correlation with human judgment and was employed as an evaluation metric and training objective.

Adversarial losses have evolved from the initial GAN formulation to more stable formulations, such as WGAN [26], WGAN-GP [27], and relativistic GAN [28]. These formulations mitigate training instability often encountered in GAN-based super-resolution. Additionally, various approaches combine different content losses, adversarial losses, and regularization terms to achieve an optimal balance between reconstruction fidelity and perceptual quality.

Our proposed method advances these baseline approaches by addressing their limitations through a novel dual-discriminator architecture, tailored specifically for super-resolution task. We conduct a comprehensive analysis of various loss function combinations to identify training strategies that optimally balance reconstruction accuracy with perceptual quality.

2. Method

The primary objective is to reconstruct low-resolution input image I^{LR} and generate its super-resolved counterpart I^{SR}. During training, the high-resolution I^{HR}serves as groun truth reference. Bicubic interpolation is used to downsample images and create the low-resolution inputs.

2.1. Super Resolution Generative Adversarial Network

Following Goodfellow et al. [29], the GAN framework can be defined as a generator G and discriminator D. The primary goal is to optimize generator G together with discriminator D to address the adversarial min-max problem (1).

$$\min_G \max_D \mathbb{E}_{x\sim p_{\text{data}}(x)}[\log D(x)] + \mathbb{E}_{z\sim p_z(z)}[\log(1 - D(G(z)))] \tag{1}$$

Here $p_{data}(x)$ represents the distribution of real data samples, $p_z(z)$ is the prior distribution of input noise variables, $G(z)$ generates fake samples, and $D(x)$ outputs the probability that x came from the real data distribution rather than the generator's distribution. In the context of super-resolution, the generator G maps low-resolution images to high-resolution counterparts, while the discriminator D attempts to distinguish between real high-resolution images and the generator's super-resolved outputs.

SRGAN has a conventional GAN structure: a discriminator and a generator, without auxiliary models, but their structures are modified to address super-resolution task Fig. 1 and 2.

For our research, we will adopt SRGAN as the foundational model. This choice is driven by its proven ability to generate high-resolution images. By leveraging SRGAN's architecture, which combines adversarial training with perceptual loss, we can ensure that the resulting images not only have improved pixel density but also maintain visual realism. Its established success in super-resolution tasks provides a reliable starting point for further modifications and optimizations tailored to our specific objectives.

2.2. Dual Discriminator Structure

Building upon SRGAN's architecture, the modified model for our task retains the original generator to preserve its high-quality image reconstruction capabilities. The primary change lies in the discriminator, which is replaced by a dual-discriminator setup designed to evaluate different aspects of the generated images Fig. 3. This modification enables us to investigate how employing two complementary discriminators influences both the training dynamics and the final image quality.

Generator G takes a low-resolution image I^{LR} as an input and upscales it to produce super-resolution image I^{SR}. Two discriminators D_1 and D_2, then receive both generated super-resolution image I^{SR} and

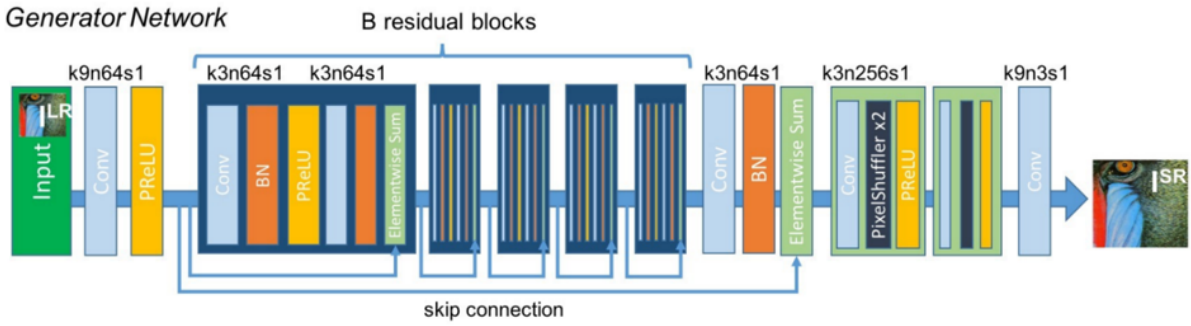

Figure 1: Generator model structure [18].

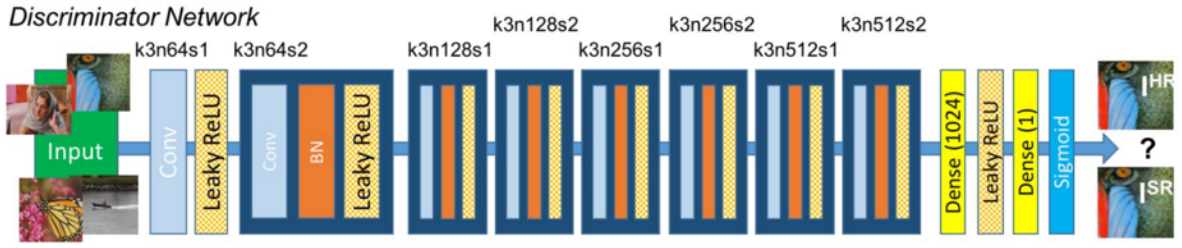

Figure 2: Discriminator model structure [18].

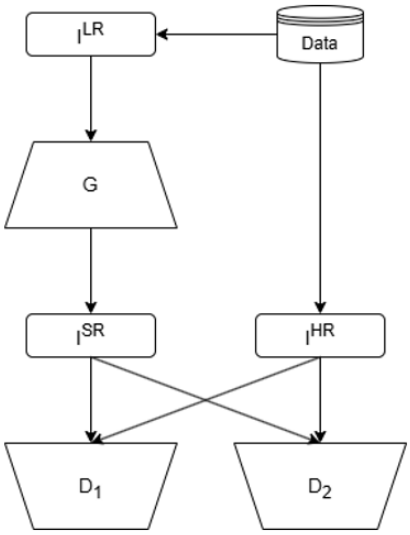

Figure 3: Structure of GAN.

original high-resolution image I^{HR} as inputs. Their task is to distinguish between real high-resolution and synthetically generated image

2.3. Structure of the global discriminator

The structure of the global discriminator consists of three main parts: a downsampling block, a bottleneck and a residual block. Their interaction allows for effective classification of the input images. The architecture of the global discriminator is illustrated in Fig. 4.

The network takes as input an image that passes through an initial convolutional layer with a kernel size of 4x4 and 64 feature maps, and then a Leaky ReLU layer. Next, let's take a closer look at the structure of each block and the task they perform when processing the input image.

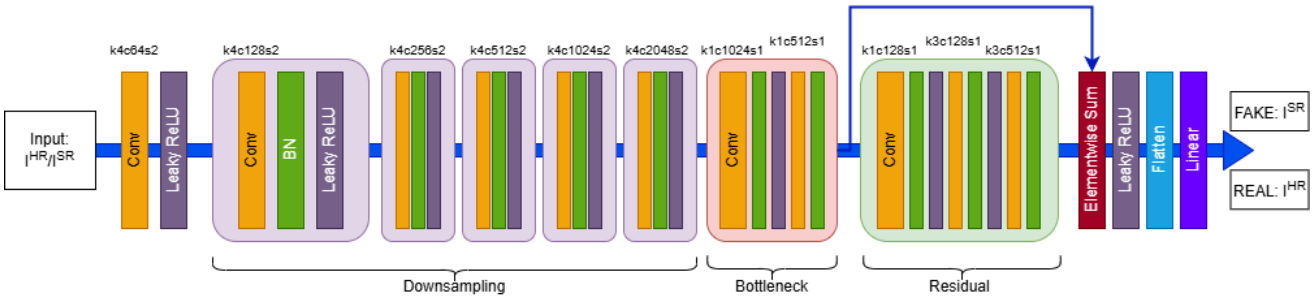

Figure 4: Architecture of the global discriminator network.

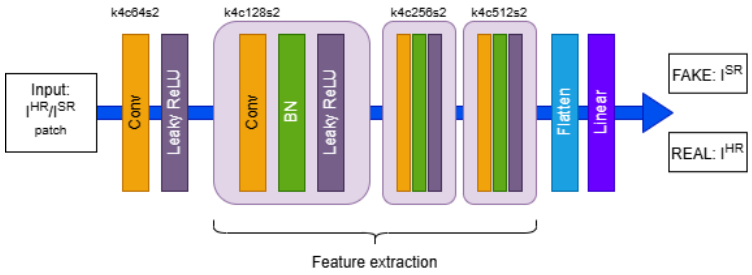

Figure 5: Architecture of the local discriminator network.

The first part is the downsampling block. It consists of 5 blocks of a similar structure, including a convolutional layer, batch normalization, Leaky ReLU, but in each subsequent block the number of feature maps doubles. Such structure allows to gradually reduce the image discretization, which allows the network to obtain image features at different sizes. Initial blocks receive basic patterns such as edges and textures, when deeper blocks can recognize higher-level semantic features.

Next part of the network is the sifting block. It also consists of a standard structure: a convolutional layer with a kernel dimension of 1x1 and 1024 feature maps, a batch normalization, LeakyReLU, a convolutional layer with a kernel dimension of 1x1 and 512 feature maps, and another batch normalization. Its task is to compress information and create cross-channel integration. It also reduces the number of parameters and thus the cost of calculations.

The last part is the residual block. It contains of a convolutional layer of 1x1 kernels and 128 feature maps, followed by a layer of batch normalization, Leaky ReLU, which are repeated twice with the only difference being the size of the 3x3 kernel of the convolutional layer. After, there is a convolutional layer of 3x3 kernels and 512 feature maps, batch normalization, and the elementwise sum. Its implementation helps prevent the problem of vanishing gradients.

In the final stage of the global discriminator, the output from the preceding blocks is passed through a LeakyReLU activation, followed by a flattening layer and a linear activation function. This produces a final prediction, indicating whether the analyzed input image is classified as real or fake.

2.4. Structure of the local discriminator

Unlike the global discriminator, which evaluates the image as a whole, the local discriminator focuses on analyzing smaller regions. This approach enables the model to examine fine-grained features in greater detail, such as textures, surface patterns, object edges, and other local structures.

By assessing these localized areas, the discriminator can provide the generator with precise feedback on the quality of each region. This, in turn, helps enhance local detail—particularly important when restoring images containing numerous small objects and intricate features that might otherwise be lost during resolution enhancement.

The local discriminator has more compact architecture compared to the global discriminator, consisting primarily of a single feature extraction block. It processes several cropped patches of fixed size, as specified in the network parameters. The overall structure of the local discriminator is illustrated in Fig. 5.

Sliding window with determined parameters picks parts of the images, which are used as input of the local discriminator. This parts of image pass through the following initial layers: convolutional with 4x4 kernel size and 64 feature maps and Leaky ReLU layer. Next, the main part begins, where features are extracted from the input data. It consists of three blocks of similar structure: a convolutional layer with 4x4 kernel size, batch normalization, and Leaky ReLU. The only difference between them is in the size of the feature maps, which, with increasing depth, have the following sizes: 128, 256, and 512. Finally, the obtained features are passed through a straightening layer and a linear activation function.

Similarly to the global discriminator, the 4x4 kernel size allows us to maintain a balance between the complexity of calculations and the capture of local features. This convolutional size also prevents any loss of features.

The combination of global and local discriminators improves the image analysis and thus increases the percentage of correctly identified images. As a result, the generator has much more feedback, and therefore better results.

A sliding window, configured with predefined parameters, extracts image patches that serve as inputs for the local discriminator. These patches first pass through an initial processing stage consisting of a convolutional layer with a 4x4 kernel, 64 feature maps, and a Leaky ReLU activation.

The main feature extraction stage follows, comprising three sequential blocks of identical structure. Each block contains a convolutional layer with a 4x4 kernel, batch normalization, and a Leaky ReLU activation. The primary difference between these blocks lies in the number of feature maps, which increase with network depth: 128, 256, and 512, respectively. The extracted features are then flattened and passed through a fully connected layer with a linear activation function.

By combining the outputs of both the global and local discriminators, the network benefits from additional perspectives: global structure evaluation and fine-grained local analysis. This synergy improves image assessment accuracy, providing the generator with richer and more precise feedback, that leads to higher-quality results.

2.5. Loss function

Defining the loss function is a crucial factor for the quality of the network. In many similar works, a common option is to use the mean square error (MSE) loss function [30, 31]. In this work, we use a modified version of MSE (2), which incorporates additional image analysis through a separate VGG19 network [32].

$$\mathcal{L}^{SR} = \mathcal{L}^{MSE} + \mathcal{L}^{VGG} + \mathcal{L}^{ADV} \tag{2}$$

Here $\mathcal{L}^{MSE}$ is the common MSE loss function, $\mathcal{L}^{VGG}$ is the loss function associated with the VGG19 network and $\mathcal{L}^{ADV}$ is a adversarial loss.

2.5.1. MSE loss

The pixel-by-pixel MSE loss (3) is well known function which measures the average squared difference between corresponding pixels in predicted $G(I^{LR})$ and target I^{HR} images of width W and height H.

$$\mathcal{L}^{MSE} = \frac{1}{WH} \sum_{x=1}^{W} \sum_{y=1}^{H} \left(I_{x,y}^{HR} - \mathrm{G}\left(I^{LR}\right)_{x,y} \right)^2 \tag{3}$$

Since this loss function cannot provide enough information to support the recovery of small image details, resulting in smoothing of textures and poor appearance of the generated image, we introduce

an additional perceptual loss function based on VGG19 [33], which will bring the results closer to generating realistic images.

2.5.2. Perceptual loss function

$$\mathcal{L}^{VGG} = 2 * 10^{-6} \cdot \left(\left\| \phi \left(I^{HR} \right) - \phi \left(I^{SR} \right) \right\|_2 \right)^2 \tag{4}$$

As previously noted, the perceptual loss function (4) uses a pre-trained VGG19 neural network [32] to extract high-level feature representations of images. It is defined as the Euclidean distance between the feature maps produced by VGG network $\phi(\cdot)$ when applied to real image (I^{HR} and generated image I^{SR}.

2.5.3. Adversarial loss

$$\mathcal{L}^{ADV} = \mathcal{L}^{LOC} + \mathcal{L}^{GLOB} \tag{5}$$

$$\mathcal{L}^{LOC} = 10^{-3} \mathcal{L}^{BCE} \left(D_{\text{Local}} \left(I^{SR} \right) \right. \tag{6}$$

$$\mathcal{L}^{GLOB} = 10^{-3} \mathcal{L}^{BCE} \left(D_{\text{Global}} \left(I^{SR} \right) \right. \tag{7}$$

Here $D_{\text{Global}}\left(I^{SR}\right)$ and $D_{\text{Local}}\left(I^{SR}\right)$ are results of the local and global discriminators, respectively and $\mathcal{L}^{BCE}$ is Binary Cross-Entropy (BCE) [34].

In addition we also utilize the adversarial part of our GAN network. The associated loss function drives the generator G tooward producing more natural-looking images generation capable of deceiving the network's dual discriminator. Given that our model utilizes two discriminators, the loss function is shown in (5), (6), and (7).

The overall loss function effectively improves high-resolution image generation by integrating three components: MSE, VGG, and adversarial loss. This integration achieves an optimal balance between reconstruction accuracy, fine-detail preservation, and perceptual realism, making it a highly effective approach for generating high-quality images.

2.6. Dataset

The dataset used for training the network is the DOTA dataset [35]. This dataset contains 2806 aerial images collected from various sensors and platforms, each of them is approximately 4000×4000 pixels in size, although the dataset includes images ranging from 800×800 to 20,000×20,000 pixels.

The images are taken from several sources, including Google Earth, GF-2 and JL-1 satellites, and others. The dataset contains both RGB and grayscale images, providing a variety of spectral information for analysis.

DOTA covers a wide range of geographic locations, such as airports, agricultural fields, urban areas, etc. and objects - airplanes, ships, cars, buildings, etc. Wide range of scales, orientations, and shapes enriches training data, reducing overfitting and supporting effective image reconstruction across both fine and coarse details.

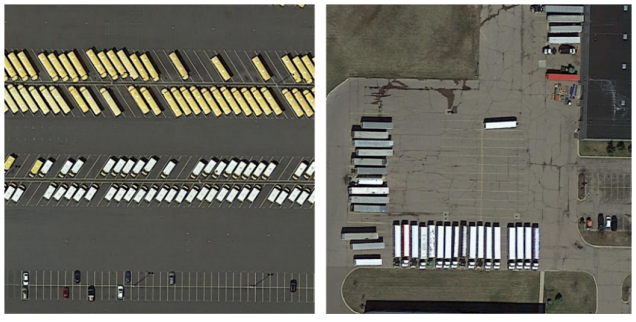

Figure 6: Examples of images, used for training.

However, original dataset does not fully meet the requirments for network training must be preprocessed. Each image will be cropped to a fixed size of 1024x1024 pixels. Additionally, since some images contain black borders, further actions will be applied to remove these artifacts and avoid potential training issues. Examples of processed images are illustrated in Fig. 6.

After preprocessing, the dataset contains 1,900 images, which are split into training (70%), validation (20%), and test (10%) subsets. This refined dataset supports effective model training and reliable performance evaluation.

2.7. Results

Initial training process for the generator lasted 50 epochs, after which adversarial training for double discriminator GAN was conducted witrh a length of 1200 epochs with training parameters:

- Batch size: 16.
- Learning rate decay: 0.1.
- Optimizer: Adam with parameters $weightdecay = 10^{-4}; \beta_1 = 0.9; \beta_2 = 0.999$.

Model evaluated its performance every 5 epochs on a validation subset. According to visual assessment, the model demonstrated quite good results. Fig. 7 demonstrates the results of increasing image resolution.

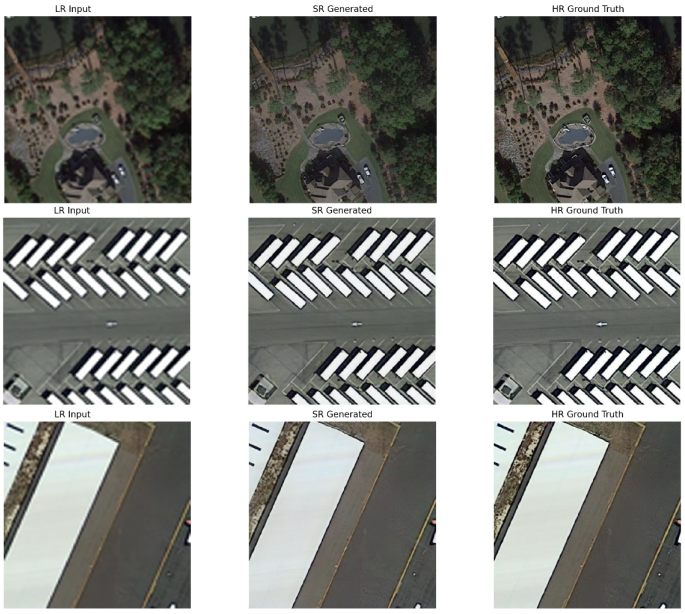

Figure 7: Examples of super resolution results.

In this image, the left column shows the original low-resolution images from the dataset, the center column presents the results produced by the modified generative adversarial network, and the right column displays the original high-resolution images.

To evaluate objective quality of super-resolution results, we computed numerical metrics including Mean Squared Error (MSE), Peak Signal-to-Noise Ratio (PSNR), and Structural Similarity Index Measure (SSIM). Table 1 compares these values between initial and final training epochs.

The results confirm the effectiveness of the developed modified generative adversarial network for satellite image super-resolution. Similar metric values between training and test sets indicate good model generalization without overfitting. Despite limitations in recovering fine details in complex images, the model shows stable performance and is suitable for practical satellite image enhancement with 4x resolution upscaling.

Table 1
Comparison of metric values between epochs

Epoch	MSE	PSNR	SSIM
10	0.0142	24.72 dB	0.6491
200	0.0105	25.56 dB	0.6978
1200	0.0082	27.06 dB	0.7696

Table 2
Metric values for test dataset

MSE	PSNR	SSIM
0.0089	26.95 dB	0.7663

Table 3
Comparison of metric values with base model

Model	MSE	PSNR	SSIM
Dual discriminator	0.0082	27.06 dB	0.7696
Base model	0.0240	22.24 dB	0.4656

2.8. Comparison with base model

To further assess effectiveness of the dual discriminator model, we conducted an additional comparison with a baseline version of the network containing only global discriminator. For a fair evaluation, both models were configured identically and trained for the same number of epochs. After training the baseline model for 1200 epochs, the results presented in Table 3 were obtained.

The results clearly indicate that proposed double discriminator model outperforms base model in all metrics. Moreover, the difference in the obtained values is significant.

2.9. Comparison with RealESRGAN

To evaluate the performance of our dual discriminator approach against state-of-the-art super-resolution methods, we conducted a comparative analysis with RealESRGAN, a widely recognized model for real-world image super-resolution. RealESRGAN employs a U-Net generator with skip connections and spectral normalization, trained with a combination of L1 loss, perceptual loss, and adversarial loss using a single discriminator architecture.

Table 4
Comparison with RealESRGAN

Method	**MSE**	**PSNR**	**SSIM**	**Parameters**
Dual discriminator	0.0082	27.06 dB	0.7696	39 million
RealESRGAN	0.0119	26.41 dB	0.7148	43 million

For fair comparison, both models were evaluated on the same test dataset containing satellite images with 4x upscaling factor. The RealESRGAN model was fine-tuned on our satellite image dataset to ensure optimal performance for this specific task.

The quantitative results demonstrate that our dual discriminator architecture achieves superior performance across all evaluated metrics. Our model consistently outperforms RealESRGAN across all metrics.

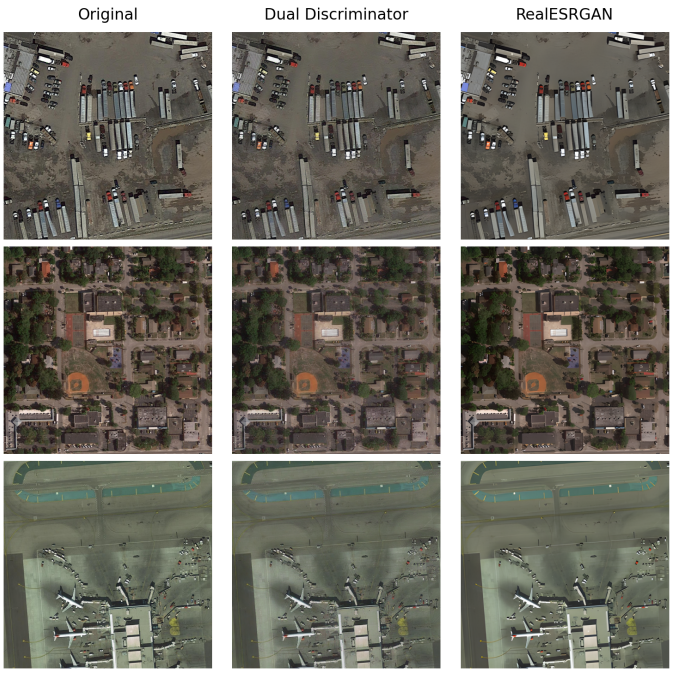

Figure 8: Visual comparison with RealESRGAN results.

Visual comparison reveals that our approach produces sharper edges and better preserves fine-grained details in satellite imagery, particularly in areas with complex textures such as urban regions and vegetation boundaries. RealESRGAN produces smoothed results that lose important structural information crucial for satellite image analysis applications. Fig. 8 presents a qualitative comparison between the two methods.

The improved performance can be attributed to the complementary roles of global and local discriminators in our architecture. While the global discriminator ensures overall image coherence similar to RealESRGAN's approach, the local discriminator specifically focuses on enhancing fine details and texture quality, preventing smoothing and leading to more realistic and detailed super-resolution results.

3. Conclusion

We propose a novel approach for enhancing the resolution of satellite images using a GAN architecture with dual discriminators. Each discriminator targets different quality features, addressing existing limitations in image enhancement and enabling more precise visual improvements.

The dual discriminator approach improves training stability and mitigates mode collapse, while the use of a diverse training dataset increases model's adaptability in different satellite conditions, such as changing lighting conditions, seasonal fluctuations, and geographical terrain.

Experimental evaluations on both validation and test datasets show strong performance. For the test set, the model achieved MSE = 0.0089, PSNR = 26.95 dB, and SSIM = 0.7663. The generated images demonstrate high visual fidelity and photorealistic detail.

This research holds significant potential for practical applications in environmental monitoring, urban planning, emergency response, and crop assessment and other fields where high-resolution imagery is important for decision-making.

In conclusion, the proposed dual-discriminator GAN offers an effective and stable solution for satellite image super-resolution, combining computational efficiency with substantial improvements in image quality.

Declaration on Generative AI

During the preparation of this work, the author(s) used X-GPT-4 in order to: Grammar and spelling check. After using these tool(s)/service(s), the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

References

[1] Sineglazov, V.M., Ischenko, V.S. Intelligent Visual Navigation System of High Accuracy 2019 IEEE 5th International Conference Actual Problems of Unmanned Aerial Vehicles Developments Apuavd 2019 Proceedings, 2019, pp. 123–127, 8943916

[2] Lutsky, M.G., Sineglazov, V.M., Ishchenko, V.S.Suppression of Noise in Visual Navigation Systems2021 IEEE 6th International Conference on Actual Problems of Unmanned Aerial Vehicles Development Apuavd 2021 Proceedings, 2021, pp. 7–10

[3] Sineglazov, V. Landmarks navigation system software2014 IEEE 3rd International Conference on Methods and Systems of Navigation and Motion Control Msnmc 2014 Proceedings, 2014, pp. 62–65, 6979731

[4] Sineglazov, V.M. Multi-functional integrated complex of detection and identification of UAVs2015 IEEE 3rd International Conference Actual Problems of Unmanned Aerial Vehicles Developments Apuavd 2015 Proceedings, 2015, pp. 320–323, 7346631

[5] Keys, R. (1981). Cubic convolution interpolation for digital image processing. IEEE transactions on acoustics, speech, and signal processing, 29(6), 1153-1160.

[6] Li, X., & Orchard, M. T. (2001). New edge-directed interpolation. IEEE transactions on image processing, 10(10), 1521-1527.

[7] Freeman, W. T., Jones, T. R., & Pasztor, E. C. (2002). Example-based super-resolution. IEEE Computer graphics and Applications, 22(2), 56-65.

[8] Glasner, D., Bagon, S., & Irani, M. (2009, September). Super-resolution from a single image. In 2009 IEEE 12th international conference on computer vision (pp. 349-356). IEEE.

[9] Yang, J., Wright, J., Huang, T. S., & Ma, Y. (2010). Image super-resolution via sparse representation. IEEE transactions on image processing, 19(11), 2861-2873.

[10] Zeyde, R., Elad, M., & Protter, M. (2010, June). On single image scale-up using sparse-representations. In International conference on curves and surfaces (pp. 711-730). Berlin, Heidelberg: Springer Berlin Heidelberg.

[11] Zgurovsky, M., Sineglazov, V., Chumachenko, E.Classification and Analysis Topologies Known Artificial Neurons and Neural Networks Studies in Computational Intelligence, 2021, 904, pp. 1–58

[12] Dong, C., Loy, C. C., He, K., & Tang, X. (2014). Learning a deep convolutional network for image super-resolution. In Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part IV 13 (pp. 184-199). Springer International Publishing.

[13] Kim, J., Lee, J. K.,& Lee, K. M. (2016). Accurate image super-resolution using very deep convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1646-1654).

[14] Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A. P., Bishop, R., ... & Wang, Z. (2016). Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1874-1883).

[15] Lai, W. S., Huang, J. B., Ahuja, N., & Yang, M. H. (2017). Deep laplacian pyramid networks for fast and accurate super-resolution. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 624-632).

[16] Lim, B., Son, S., Kim, H., Nah, S., & Mu Lee, K. (2017). Enhanced deep residual networks for single image super-resolution. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 136-144).

[17] Zhang, Y., Tian, Y., Kong, Y., Zhong, B., & Fu, Y. (2018). Residual dense network for image super-

resolution. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2472-2481).

[18] Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., ... & Shi, W. (2017). Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4681-4690).

[19] Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., ... & Change Loy, C. (2018). Esrgan: Enhanced super-resolution generative adversarial networks. In Proceedings of the European conference on computer vision (ECCV) workshops (pp. 0-0).

[20] Zhang, Y., Li, X., & Zhou, J. (2019). SFTGAN: a generative adversarial network for pan-sharpening equipped with spatial feature transform layers. Journal of Applied Remote Sensing, 13(2), 026507-026507.

[21] Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1125-1134).

[22] Wang, T. C., Liu, M. Y., Zhu, J. Y., Tao, A., Kautz, J., & Catanzaro, B. (2018). High-resolution image synthesis and semantic manipulation with conditional gans. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 8798-8807).

[23] Sajjadi, M. S., Scholkopf, B., & Hirsch, M. (2017). Enhancenet: Single image super-resolution through automated texture synthesis. In Proceedings of the IEEE international conference on computer vision (pp. 4491-4500).

[24] Johnson, J., Alahi, A., & Fei-Fei, L. (2016). Perceptual losses for real-time style transfer and super-resolution. In Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14 (pp. 694-711). Springer International Publishing.

[25] Wang, Z., Simoncelli, E. P., & Bovik, A. C. (2003, November). Multiscale structural similarity for image quality assessment. In The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003 (Vol. 2, pp. 1398-1402). Ieee.

[26] Arjovsky, M., Chintala, S., & Bottou, L. (2017, July). Wasserstein generative adversarial networks. In International conference on machine learning (pp. 214-223). PMLR.

[27] Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., & Courville, A. C. (2017). Improved training of wasserstein gans. Advances in neural information processing systems, 30.

[28] Jolicoeur-Martineau, A. (2018). The relativistic discriminator: a key element missing from standard GAN. arXiv preprint arXiv:1807.00734.

[29] Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. Advances in neural information processing systems, 27.

[30] Dong, C., Loy, C. C., He, K., & Tang, X. (2015). Image super-resolution using deep convolutional networks. IEEE transactions on pattern analysis and machine intelligence, 38(2), 295-307.

[31] Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A. P., Bishop, R., ... & Wang, Z. (2016). Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1874-1883).

[32] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

[33] Gatys, L., Ecker, A. S., & Bethge, M. (2015). Texture synthesis using convolutional neural networks. Advances in neural information processing systems, 28.

[34] Ruby, U., & Yendapalli, V. (2020). Binary cross entropy with deep learning technique for image classification. Int. J. Adv. Trends Comput. Sci. Eng, 9(10).

[35] Xia, G. S., Bai, X., Ding, J., Zhu, Z., Belongie, S., Luo, J., ... & Zhang, L. (2018). DOTA: A large-scale dataset for object detection in aerial images. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3974-3983).