Practical Advances in Artificial Intelligence and Computing
Frontier Research Publishing • https://www.frpubs.org
Vol. 1, No. 1, 2025

GAN-Based Image Enhancement and Restoration: A Multi-Loss Framework for Perceptual Consistency and Structural Preservation

Thayer Lockett
California State University, Chico, Chico, USA
t.lockett@csuchico.edu

Abstract:

This paper proposes an image quality reconstruction method based on generative adversarial networks to address common issues in image enhancement and restoration, such as detail loss, structural distortion, and visual inconsistency. The method is built upon an encoder-decoder framework, incorporating residual feature learning modules and multi-level structural modeling to enhance the joint representation of local textures and global structures. To improve perceptual consistency, a multi-objective optimization strategy is designed by combining perceptual loss and edge-preserving loss. This constrains the generation process from multiple dimensions and ensures that the reconstructed images align with both subjective perception and objective metrics. A deep convolutional discriminator is employed to guide the generator through adversarial training, encouraging outputs that better match real image distributions. For input degradation, the method simulates various scenarios including different levels of occlusion, noise interference, and structural damage. The robustness and restoration performance of the model are evaluated under complex conditions. Experiments conducted on multiple public image datasets demonstrate the effectiveness of the proposed approach, achieving superior performance on PSNR, SSIM, and LPIPS metrics. These results confirm the accuracy and practicality of the method for image enhancement and restoration tasks.

Keywords:

Image inpainting; Generative adversarial network; Perceptual loss; Edge constraint

1. Introduction

In the field of computer vision and image processing, enhancing image quality and effectively restoring damaged regions have become core challenges that support many application scenarios. With the widespread adoption of high-resolution imaging devices and the rapid growth of image data, recovering realistic and structurally consistent content from limited, blurry, corrupted, or low-quality images has become a pressing issue. Traditional image enhancement and restoration methods often rely on rule-based filters, interpolation techniques, or sparse coding strategies. Although these methods can achieve certain results under specific conditions, they usually fail to capture complex textures and global semantic relationships in images. They are also ineffective against the diverse and unstructured degradation types found in real-world scenarios[1].

In recent years, the rise of deep generative models has opened new technical avenues for image enhancement and restoration. Generative Adversarial Networks (GANs), in particular, have attracted significant attention due to their outstanding performance in data generation tasks[2]. These models introduce an adversarial mechanism that allows the generator to learn realistic and structurally plausible image distributions under continuous challenges from the discriminator. This end-to-end data-driven approach significantly reduces the reliance on prior knowledge. It gives the model stronger representational power and generalization ability, showing great potential in tasks such as image super-resolution, denoising, and inpainting. At the same time, the architectural evolution and training optimization of GANs provide strong support for texture preservation and detail reconstruction in the restoration process[3].

Image enhancement and restoration have technical value as well as broad application demands. In remote sensing, medical imaging, surveillance systems, and digital media restoration, images are often affected by noise, low resolution, occlusion, and damage. These factors lead to information loss or distortion. In such scenarios, improving image quality and recovering original content are no longer simple preprocessing tasks. They directly affect the reliability of subsequent intelligent decisions and the safety of system operations. Therefore, developing techniques with perceptual intelligence and structural reconstruction ability has become a key pathway for advancing visual understanding.

Against this background, GAN-based methods for image enhancement and restoration have become a research focus due to their capabilities in end-to-end learning, unsupervised reconstruction, and high-quality synthesis. These methods can model global structures and recover fine-grained details simultaneously. They are especially effective for highly degraded images. By incorporating semantic guidance, attention mechanisms, and conditional networks, GANs can better capture semantic dependencies and spatial structures in images. This enables effective reconstruction of complex degraded samples and pushes restoration algorithms toward more intelligent and robust directions[4].

In conclusion, GAN-based research on image enhancement and restoration reflects a major shift in modeling paradigms. It also represents a cutting-edge fusion of deep learning and image perception. Continued exploration of these technologies may significantly improve image reconstruction quality. It can also enable multi-source and multi-scale joint modeling, increasing the adaptability of vision systems to non-ideal inputs. This lays a solid foundation for building high-quality intelligent perception systems. In the context of data-intensive environments and increasingly complex vision tasks, further development of related theories and methods holds great scientific and practical value.

2. Related work

Image enhancement and restoration techniques have evolved from traditional image processing methods to deep learning-based approaches. Early methods relied mainly on statistical properties and handcrafted rules, such as bilateral filtering, median filtering, and wavelet transforms, to enhance or repair images. These techniques achieved certain success in tasks like noise removal and edge enhancement. However, they often failed when dealing with structurally complex or severely corrupted images. They struggled to restore realistic content. To address these shortcomings, some studies introduced sparse representation and dictionary learning to guide the restoration process using prior models. Yet, their generalization and expressive power remained limited in large-scale and diverse degradation scenarios[5].

With the development of deep learning, convolutional neural networks have been widely adopted in image enhancement and restoration tasks. These methods can automatically learn underlying structure and texture patterns from large-scale data, significantly improving reconstruction quality. Many studies have proposed end-to-end architectures for denoising, deblurring, super-resolution, and image inpainting. These models showed better adaptability and performance than traditional methods. However, due to the limitation of optimizing a single loss function, they often failed to balance local texture and global structure. As a result, the restored images may still appear unnatural. This issue becomes more prominent when handling large missing regions, often leading to blur or artifacts[6].

The introduction of Generative Adversarial Networks has brought a new research paradigm to image enhancement and restoration. Through adversarial training between generator and discriminator, GANs can approximate real image distributions without relying on explicit labels. This leads to more natural and realistic image generation. Many studies have shown that GANs are effective in tasks such as super-resolution, denoising, dehazing, and occlusion inpainting. The key lies in the adversarial loss, which complements traditional reconstruction losses by modeling high-frequency details. This improves both the subjective quality and perceptual consistency of the generated images. Furthermore, extensions such as conditional GANs, multi-scale discriminators, and structure-aware losses have continuously enhanced the accuracy and stability of image restoration[7].

In recent years, to address instability and divergence during GAN training, researchers have proposed various improvements in model architecture and training strategies. Attention mechanisms and semantic guidance modules help the generator focus on key regions, improving structural preservation and content coherence[8]. At the same time, combining perceptual loss, style loss, and structural consistency loss provides richer supervision signals. This enhances the model's ability to jointly capture semantic and visual details. To meet the requirements of diversity and contextual understanding in restoration tasks, existing work has also explored dual-branch networks, residual connections, and spatial transformation modules. These advances increase network depth and nonlinear capacity. They further expand the application boundaries and performance of GAN-based image enhancement and restoration.

3. Method

This paper proposes an image enhancement and restoration method based on generative adversarial networks. The overall architecture consists of a generator and a discriminator, where the generator is responsible for reconstructing low-quality or damaged input images, and the discriminator is used to determine the difference between the generated image and the real image. The generator adopts an encoder-decoder structure, extracts multi-scale features through layer-by-layer convolution, and combines the jump connection mechanism in the decoding stage to enhance the ability to reconstruct local details. The model architecture is shown in Figure 1.

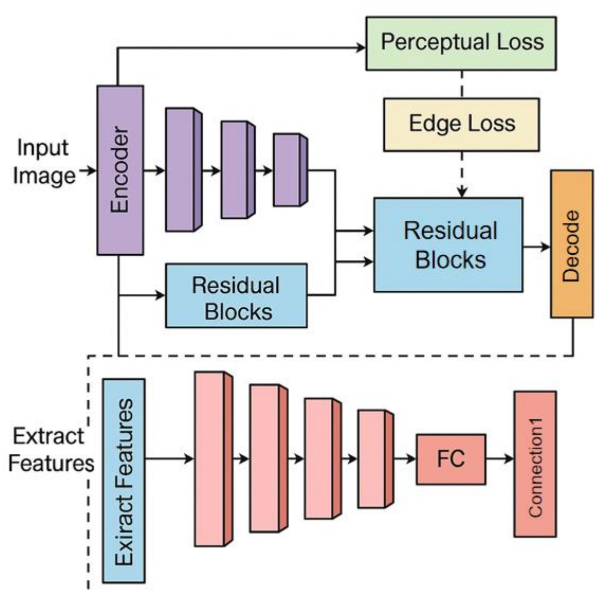

Figure 1. Overall model architecture diagram

The discriminator is constructed using a multi-layer convolutional network to perform binary classification predictions on the authenticity of the image, thereby providing a reverse constraint signal for the generator. During the training process, the generator and the discriminator are optimized alternately in an adversarial manner, so that the model learns a more realistic image distribution in a continuous game. The overall training objective function can be expressed as:

$$\min_G \max_D L_{GAN}(G,D) = E_{x\sim p_{data}}(x)[\log D(x)] \\ + E_{z\sim p_z(z)}[\log(1-D(G(z)))]$$

In order to improve the structural consistency and semantic rationality of the generated images, this paper introduces perceptual loss and reconstruction loss as auxiliary constraints in addition to the basic adversarial loss. The perceptual loss is calculated through the intermediate layer features of the pre-trained convolutional neural network and is used to measure the similarity between the generated image and the original image in the high-level semantic space, thereby ensuring the consistency of the generated image with the real image in visual perception. The formula is as follows:

$$L_{perceptual} = \sum_{i=1}^{N} \lambda_i \| \phi_i(G(z)) - \phi_i(x) \|_2^2$$

Where ϕ_i represents the feature extracted by the i-th layer of the pre-trained network, and λ_i is the corresponding weight factor. In order to enhance the pixel accuracy of the image, the reconstruction loss defined by the L1 norm is used to constrain the pixel difference between the generated image and the original image. Its expression is as follows:

$$L_{recon} = \| G(z) - x \|_1$$

In addition, in order to enhance the model's ability to model the boundaries of image structures, an edge preservation loss function is introduced. This loss is calculated based on the image gradient and aims to preserve the edge information of the image. The expression is:

$$L_{edge} = \| \nabla G(z) - \nabla x \|_1$$

Finally, the overall loss function proposed in this paper combines the above multiple components to jointly optimize the generator parameters:

$$L_{total} = L_{GAN} + \alpha L_{perceptual} + \beta L_{recon} + \gamma L_{edge}$$

Where α, β, γ is the weight coefficient of each loss, which is used to balance the relationship between adversarial resistance, perceptual quality, pixel accuracy and structure preservation. Through the above multi-loss collaborative mechanism, the generator can achieve high-quality restoration of local details and global structures while maintaining the naturalness of the image.

4. Dataset

This study uses the CelebA-HQ dataset as the primary source of experimental data. The dataset is constructed based on the original CelebA dataset by applying high-quality image generation and face alignment techniques. It contains 30,000 high-resolution facial images at 1024 × 1024 pixels. The images cover a wide range of gender, age, facial expressions, and background conditions. Compared with traditional facial image datasets, CelebA-HQ provides significant advantages in texture quality and image consistency. It is well-suited for evaluating texture and structure reconstruction in image enhancement and restoration tasks.

During data preprocessing, all images are cropped and resized to a fixed resolution of 256 × 256 to improve training robustness and generalization. Degraded inputs are created by applying occlusion simulation, noise injection, and block removal. These steps are designed to mimic image corruption in real-world scenarios. The original image serves as the ground truth. The degraded version is used as input to train the generative adversarial network under complex conditions.

To evaluate the model's generality and adaptability to different input distributions, a subset of images from the FFHQ (Flickr-Faces-HQ) dataset is used for extended testing. This dataset also features high resolution, high quality, and strong diversity. It provides richer structural and semantic information. These additional samples help to further assess model performance in complex image restoration scenarios.

5. Experimental Results

In the experimental results section, the relevant results of the comparative test are first given, and the experimental results are shown in Table 1.

Table 1. Comparative experimental results

Method	PSNR	SSIM	LPIPS
EdgeConnect[9]	26.42	0.891	0.112
HiFill[10]	25.76	0.873	0.127
ICT[11]	27.33	0.905	0.098
CoMod-GAN[12]	28.91	0.918	0.085
Ours	30.24	0.935	0.073

From the overall experimental results, the proposed method for image enhancement and restoration outperforms several mainstream models across multiple key evaluation metrics. It shows stronger comprehensive performance. Notably, the method achieves 30.24 in PSNR and 0.935 in SSIM, which are significantly higher than recent advanced models such as EdgeConnect, HiFill, ICT, and CoMod-GAN. These results indicate that the method has clear advantages in pixel-level reconstruction accuracy and structural preservation. It can effectively recover detailed information and local textures in the images.

Further analysis shows that the proposed method achieves a score of 0.073 on the LPIPS metric, outperforming all comparison models. This suggests that the perceptual difference between the generated images and the original images is the smallest, leading to higher visual naturalness and subjective quality. In contrast, although HiFill and EdgeConnect maintain structural integrity to some extent, they suffer from poor visual effects. These methods often produce blurry outputs or discontinuities at edges. This difference highlights the potential of generative adversarial networks in perceptual optimization. When combined with perceptual loss and edge structure constraints, the model can significantly improve its ability to capture fine textures.

It is worth noting that while CoMod-GAN shows competitive performance on SSIM and LPIPS, it still falls short compared with the proposed method. This indicates that large-scale modeling and conditional fusion alone are not sufficient to comprehensively improve restoration quality. The proposed method incorporates residual learning and jointly optimized multi-loss strategies. It maintains the realism of the generated images while effectively balancing global structural consistency and local detail recovery. This results in more stable and superior performance across multiple dimensions.

In summary, the experimental results confirm the effectiveness and robustness of the proposed method in image enhancement and restoration tasks. By integrating adversarial training, structure-aware mechanisms, and perceptual constraints, the model improves both objective reconstruction quality and adaptability to complex textures and irregular corrupted regions. It demonstrates strong potential for practical application and research value. The method is especially suitable for real-world visual systems that require high image quality and semantic consistency.

This paper also gives an evaluation of the model robustness under changes in the input image occlusion ratio, and the experimental results are shown in Figure 2.

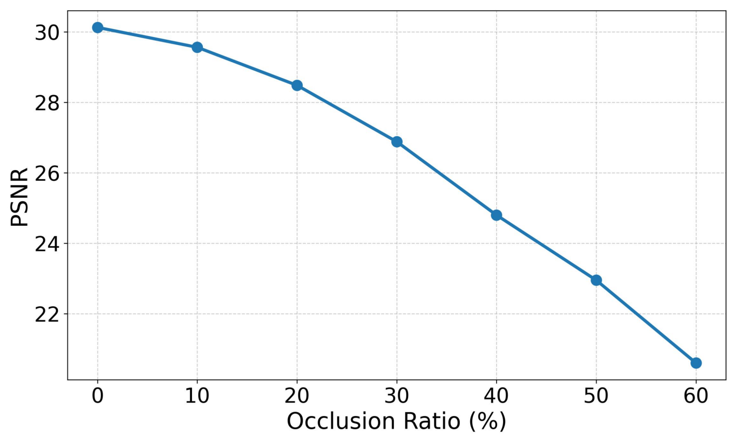

Figure 2. Model robustness evaluation under input image occlusion ratio changes

From the figure, it can be observed that as the occlusion ratio of the input image gradually increases, the PSNR value of the model shows a steady downward trend. This indicates that the expansion of the occluded region significantly affects image reconstruction quality. This phenomenon is consistent with the general principle of image inpainting. The larger the missing area, the more difficult it becomes to reconstruct the image. This is especially true in scenarios that require high texture continuity and structural consistency, where occlusion severely interferes with semantic perception and detail recovery.

Although the overall trend declines, the model maintains relatively high PSNR values at low occlusion ratios (such as 0% to 20%). This shows that it has a certain degree of capability in modeling local structures and understanding image content. Under light occlusion, the model can effectively infer missing information by leveraging the surrounding context. This confirms the effectiveness of the structure-aware mechanism and residual feature modeling used in this study under small-scale corruption.

When the occlusion ratio exceeds 30%, the performance decline accelerates. In particular, when occlusion reaches 50% to 60%, the PSNR drops significantly. This result indicates that in high-occlusion scenarios, the model faces greater uncertainty. It becomes difficult to accurately recover global structure and semantic information from limited regions. At this stage, traditional context-based completion strategies are insufficient, and deeper structural constraints and prior guidance are required to improve stability.

Overall, this experiment evaluates the robustness and adaptability of the proposed model under varying levels of occlusion, which is a key source of disturbance. The results show that although performance degradation still occurs under extreme conditions, the model demonstrates strong reconstruction and structural preservation abilities in most practical scenarios. This validates the practical potential and broad applicability of the proposed method in handling partial loss and unstructured corruption in images.

6. Conclusion

This paper addresses key challenges in image enhancement and restoration, including structural loss, texture blur, and perceptual inconsistency. A comprehensive solution based on generative adversarial networks is proposed. The method integrates an encoder-decoder structure with multi-layer residual modules. It also combines perceptual loss and edge-preserving mechanisms to enhance the model's ability to reconstruct semantic content and local textures. The overall design balances global structure and fine details, improving the naturalness and consistency of image reconstruction. This provides an efficient and stable modeling path for restoring low-quality images.

In various typical image enhancement and restoration tasks, the proposed model demonstrates significant advantages in visual quality and structural preservation. It achieves high-quality recovery of diverse corrupted samples without relying on large amounts of labeled data. Experimental results show that the model performs robustly under different types of interference and varying occlusion ratios. This indicates strong generalization ability and practical value. Compared with traditional rule-based or prior-dependent methods, the proposed approach offers higher flexibility and better perceptual adaptability. It effectively meets the diverse needs of real-world scenarios.

The proposed method has theoretical value in foundational areas such as visual perception and image understanding. It also shows strong application potential in various fields. These include remote sensing image restoration, medical image enhancement, target recovery in surveillance, and restoration of cultural heritage images. The model's end-to-end structure and unsupervised nature allow for easy deployment across scenarios. This offers intelligent and efficient support for image processing systems and lays a solid foundation for high-level visual tasks such as detection and recognition.

Future work may focus on further improving generation quality and training stability. Cross-modal perception mechanisms or graph-based representations could enhance the model's adaptability to semantic gaps and complex scenes. In addition, integrating self-supervised learning and multi-task joint optimization may improve the model's representation and restoration capabilities under sparse annotations. As computational resources advance and data acquisition methods diversify, generative models for image enhancement and restoration are expected to show broader impact and application in real-world tasks.

References

[1] Zhang X, Zhai D, Li T, et al. Image inpainting based on deep learning: A review[J]. Information Fusion, 2023, 90: 74-94.

[2] Qin Z, Zeng Q, Zong Y, et al. Image inpainting based on deep learning: A review[J]. Displays, 2021, 69: 102028.

[3] Xu Z, Zhang X, Chen W, et al. A review of image inpainting methods based on deep learning[J]. Applied Sciences, 2023, 13(20): 11189.

[4] Yu T, Feng R, Feng R, et al. Inpaint anything: Segment anything meets image inpainting[J]. arXiv preprint arXiv:2304.06790, 2023.

[5] Wang S, Saharia C, Montgomery C, et al. Imagen editor and editbench: Advancing and evaluating text-guided image inpainting[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023: 18359-18369.

[6] Quan W, Zhang R, Zhang Y, et al. Image inpainting with local and global refinement[J]. IEEE Transactions on Image Processing, 2022, 31: 2405-2420.

[7] Chen Y, Xia R, Zou K, et al. FFTI: Image inpainting algorithm via features fusion and two-steps inpainting[J]. Journal of Visual Communication and Image Representation, 2023, 91: 103776.

[8] Yu Y, Zhan F, Lu S, et al. Wavefill: A wavelet-based generation network for image inpainting[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2021: 14114-14123.

[9] Nazeri K, Ng E, Joseph T, et al. Edgeconnect: Generative image inpainting with adversarial edge learning[J]. arXiv preprint arXiv:1901.00212, 2019.

[10] Moskalenko A, Erofeev M, Vatolin D. Deep two-stage high-resolution image inpainting[J]. arXiv preprint arXiv:2104.13464, 2021.

[11] Liu Q, Tan Z, Chen D, et al. Reduce information loss in transformers for pluralistic image inpainting[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 11347-11357.

[12] Alkobi N, Shaham T R, Michaeli T. Internal diverse image completion[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 648-658.