Improve image repair efficiency and quality based on Diffusion-GANs model

Jingyi Huang[1*], *Xingsheng* Qin[2], and *Haopeng*Yang[3]

[1]School of Computer Science, Wuhan University, 430072, Wuhan, Hubei, China
[2]School of Computer and Communication Engineering, 410114, Changsha University of Science and Technology, Changsha, Hunan, China
[3]School of Computer and Information Science School of Software, 400715, Southwest University, Beibei, Chongqing, China

Abstract. With the rapid development of modern science and technology, deep learning and artificial intelligence have greatly promoted the progress of computer vision, among which image restoration is a key field of computer vision, aiming to repair images that are damaged or missing important parts of the main body. Although traditional interpolation and region filling techniques are effective in some environments, they often have difficulty handling complex scenes that require image restoration in today's world. In contrast, modern methods such as GANs and Diffusion models have significantly improved the quality and reliability of restoration. However, GANs are hindered by problems such as instability and mode collapse, and although diffusion models can generate high-quality images, they are computationally demanding. To address these challenges, this review explores the hybrid diffusion GANs framework and focuses on the basic conceptual restoration principles of the above three models and compares the performance of the GANs and Diffusion model and other traditional models by comparing indicators such as peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM). Furthermore, although Diffusion-GANs model is mainly applied to image generation, we discuss their great potential for image inpainting, providing new possibilities for future improvements in image restoration and generation.

1 Introduction

Image restoration technology has made remarkable progress in rapidly developing modern society. This progress is attributed to breakthroughs in deep learning and artificial intelligence in the field of computer vision [1]. Image restoration focuses on restoring images with blurred or missing data to enhance their visual integrity and visual effects. There are many ways to restore images, including noise reduction, distortion correction, and reconstruction of missing areas. All of these aim to produce a complete and clear image. After many years of development, the key methods in the image restoration field have already

* Corresponding author: 2019302110136@whu.edu.cn

transitioned from basic statistical methods to complex machine learning models, which marks a leap in technology.

In the early years, image restoration was mainly based on mathematical and signal processing techniques, such as the Wiener filter and total variation image denoising algorithm; although these methods are effective in some specific scenes, they usually cannot process complex realistic tasks because they are highly dependent on predefined assumptions about noise and the key part of the missing image. However, the emergence of deep learning has marked a shift in the models. It can learn directly from complex data so that the image can be better restored. The most influential of these models are the GANs model and the Diffusion model. Both have reshaped the landscape of image restoration.

Introduced by Goodfellow et al. in 2014, GANs use a competitive framework between generators and discriminators. This innovative method has achieved breakthrough success in image repair and in solving other visual tasks. However, the GANs model still has shortcomings, including the problem of collapse and convergence of the unstable training mode. These drawbacks make it difficult to deal with overly complicated image-restoration scenes. The Diffusion model proposed in 2015 by Jascha Sohl-Dickstein et al. works by destroying training data through the successive addition of Gaussian noise and then learning to recover the data by reversing this noising process. Because of its stability in the training process and high-quality and diversified output, it has attracted the attention of academic circles [2]. At the same time, the diffusion model also faces a series of challenges, which mainly focus on the high computational cost and slow generation speed.

To solve the disadvantages of the above two models, a new model has been proposed: the Diffusion-GANs model. By combining the efficient training speed of the GANs model with the superior image output quality of Diffusion models, the Diffusion-GANs framework achieves a perfect balance between the image quality and generation speed [3]. This paper provides a comprehensive review of Diffusion-GANs model and the other two models mentioned above, showing that they have the potential to overcome the limitations of traditional image restoration methods. This study provides valuable insights for the future development of this field and promotes the development of the image restoration field.

2 Image Inpainting Methods

2.1 Image Inpainting Method Based on Diffusion-GANs Model

2.1.1 Concepts and Definitions

The Diffusion-GANs model innovatively integrates the high-quality generation capabilities of diffusion models with the efficient training features of the GANs model. The diffusion model progressively denoises images. GANs model uses the antagonistic interaction between the generator and discriminator to generate high-quality images quickly. When carrying out the task, the Diffusion-GANs model runs in two different stages. At first, the Diffusion model takes the lead and gradually denoises the image while filling in blurred or missing areas. This stage ensures that the image structure is preserved and aim to get high quality output. In the next stage, the GANs model generator intervenes to improve the restoration process, while the discriminator further enhances the realism of the image. This collaborative approach efficiently addresses the limitations of a single model.

2.1.2 Recent Advances in Diffusion-GANs Model for Image Restoration

Many researchers have made significant progress in the development of image restoration technology by innovatively utilizing Diffusion-GANs hybrid models. Luo et al. comprehensively reviewed the application of diffusion models in image restoration, covering tasks such as denoising, deblurring, and dehazing [4]. Their study emphasized the integration of diffusion models into low-level visual tasks and proposed methods to improve the synergy between diffusion models and GANs. Similarly, Xia et al. proposed the DiffIR model, which combines the diffusion process with components based on GAN models for super-resolution restoration tasks [5]. Through a two-stage training process, DiffIR not only improves the restoration performance but also reduces the computational cost, demonstrating the superiority of the fusion of diffusion models and GANs models. Based on these studies, Trinh et al. proposed a new method for performing the inverse diffusion process (denoising) in a low-dimensional latent space [6]. This method significantly reduces the inference time and computational overhead while improving the quality and diversity of the generated images. In addition, they introduced an autoencoder to compress the input data and adopted a weighted learning strategy to optimize the training process. In contrast, Kuznedelev et al. attempted to speed up the inference process by performing denoising in the wavelet space instead of the pixel space [7]. Their approach combines a generator fine-tuned for wavelet space with reconstruction loss, achieving faster convergence during training while minimizing computational effort. Overall, these studies emphasize the versatility of Diffusion-GANs models, providing new solutions that push the boundaries of image restoration in terms of efficiency and quality.

2.2 Image Inpainting Method Based on GANs Model

2.2.1 Concepts and Definitions

GANs, a class of generative models rooted in game theory, have achieved a significant breakthrough within the realm of Machine Learning applications. Leveraging the potency of competitive training methodologies and deep neural networks, GANs are adept at generating lifelike images and have demonstrated remarkable progress in numerous image generation and editing paradigms [8]. The objective function of a GAN is optimized through a process of minimization and maximization for both the generator and discriminator. The discriminator's objective is to enhance its capacity to discern between authentic and synthetic data, whereas the generator's aim is to mislead the discriminator by generating samples that are as lifelike as possible. The loss function for GANs can be articulated as Equation (1)

$$\min_{G}\max_{D} E_{x\in X}[\log D(x)] + E_{x\in X}\left[\log\left(1 - D\big(G(z)\big)\right)\right] \tag{1}$$

Notes the real samples originating from dataset X.G(z) is the fake sample generated by the generator, and zzz is the random noise. D(x) is the probability that the discriminator judges sample xxx to be real, as shown in equation (1).

Conditional Generative Adversarial Networks (cGANs) were proposed by Mirza. In the framework of cGANs, both the generative model (G) and discriminative model (D) are conditioned on supplementary information (c), which can include class labels, textual descriptions, or sketches. This conditioning element provides a means to guide the type of data produced, contributing to the widespread adoption of cGANs in numerous image-generation tasks [8]. Although GANs have made significant progress in image generation, they still have some limitations. For example, their resolution is limited, and in unconditional generative models, the people cannot control the patterns of the generated data. Although GANs can produce high-quality images, the instability of their training process and the problem of mode collapse limit their application in more complex image tasks. Next, this article introduces image inpainting methods based on improved LSGAN and other GAN models, taking Least Squares Generative Adversarial Networks (LSGAN) as an example.

2.2.2 Image Inpainting Method Based on Improved LSGAN

To address the two issues present in the original GANs — low image quality and unstable training process–Least Squares Generative Adversarial Networks (LSGAN) propose using a least squares loss function as a replacement for the sigmoid binary cross-entropy loss function in the original GAN. This improvement changes the opposing optimization directions of the generator and discriminator in the original GAN to a situation in which both are optimized in the same direction [9]. The objective of the generator within the framework of LSGAN is to deceive the discriminator into believing that the synthetic images it produces are "authentic," in other words, to drive the discriminator's output towards 1. The objective of the LSGAN discriminator is to discern the genuine image from the synthetic counterfeit one.

Yu introduced an image inpainting technique that leverages an enhanced version of the LSGAN. To bolster the stability of the model and its discriminator, they refined the discriminator architecture to incorporate a fully connected layer. To harness the texture details from the shallow layers and the semantic insights from the deeper layers of the network, without allowing the restoration process to be unduly influenced by artificial elements, they integrated a sub-pixel convolution block into the Feature Pyramid Network (FPN), thus forming the generator network. Concurrently, they devised a loss function to regulate the model's gradients, thereby enhancing the efficacy of image inpainting [9]. The model adopts an FPN structure and combines sub-pixel convolution blocks, effectively integrating information from shallow and deep networks, thus avoiding artifacts in the repaired areas and achieving improvements in the generator. In terms of the discriminator, the overall stability of the model is enhanced by introducing dense layers.

2.2.3 Various Image Inpainting Techniques Utilizing Generative Adversarial Networks

Yeh introduced an image inpainting technique that leverages the power of GAN networks to convert noisy inputs into clear image outputs post-training. This approach tackles the challenge of dealing with incomplete or corrupted images within the training dataset, which necessitates the availability of the corresponding pristine images. During the inference phase, it employs a hole-filling mechanism to produce high-quality pixel-perfect images. Nonetheless, this method is contingent upon a pretrained generative network model and is confined to converging on a specific dataset, thereby restricting its capacity to reconstruct intricate image structures [10]. Pathak introduced an innovative context encoder that integrates an encoder-decoder network architecture with a GAN. The convolutional neural network structure they devised, upon completion of training, is capable of generating content for any region requiring restoration, drawing upon the context of the neighboring areas in the compromised images [10].

2.3 Image Inpainting Method Based on Diffusion Model

2.3.1 Concepts and Definitions

Image inpainting based on diffusion models is a cutting-edge approach that leverages generative capabilities to reconstruct missing or occluded regions in images. These models excel at producing high-quality, coherent, and context-aware completions for a wide range of inpainting tasks.

Diffusion models are generative models that add noise to data during a forward process and learn to reverse this process (denoising) to generate new samples. Diffusion models are latent variable models of the formula (2).

$$\left(p_\theta(x_0) \coloneqq \int p_\theta(x_{0:T})d(x_{1:T})\right) \tag{2}$$

The random variables from this formula(from x0 to xT) are latent of the same dimensionality as the data x0 ∼ q(x0). It is considered a Markov chain that can be learned through Gaussian distribution transfer. The forward process begins by adding Gaussian noise to the input image in a fixed number of steps. During this phase, the known regions of the image remain unchanged, providing the context for the reconstruction task. The reverse denoising process, guided by a pre-trained diffusion model, then attempts to remove this noise iteratively [2]. The difference between diffusion models and other types of latent variable models is the posterior q (x1: T | x0), which is called a forward process or diffusion process and is fixed according to the variance table β 1, β T gradually adds noise to the Markov chain of the data.

2.3.2 Image Inpainting Method Based on Denoising Diffusion Probabilistic Model

Diffusion models have shown superior performance in producing high-resolution and diverse image outputs; therefore, they can also be used in image inpainting. For image inpainting, this process incorporates conditional inputs, such as a binary mask indicating the missing areas and the content of the known regions. By conditioning the reverse process on these inputs, the model generates plausible and coherent content for the occluded regions that seamlessly blend with the unmasked portions of the image.

The implementation of diffusion models for inpainting typically involves several stages. The first step is to generate a mask to indicate the missing or damaged areas in the image. Next, a diffusion model is trained using paired data comprising original images and their masked versions. This training involves a loss function that not only minimizes reconstruction errors but also ensures consistency between the generated content and the existing context. Architectures like UNet are commonly used, as they effectively process both global and local features, making them suitable for tasks requiring detailed reconstruction. During inference, the model performs iterative denoising starting from a noisy initial state and progressively refining the image until the final reconstruction is achieved [11].

A strength of diffusion models in inpainting tasks is their ability to produce high quality and diverse outputs. Unlike traditional GANs, which often suffer from mode collapse or artifacts in challenging scenarios, diffusion models generate results that are more consistent with the statistical distribution of natural images. They are particularly well suited for handling irregularly shaped masks or a picture with a large missing area, producing outputs that maintain semantic coherence with the known areas. In addition, their stochastic nature allows users to generate multiple plausible solutions for a given input by altering the random seed used during sampling [12]. The defects of the diffusion model are also apparent: it typically requires extensive iterations based on its calculation process to estimate the entire image or feature map, resulting in lower efficiency.

Several techniques are commonly used to optimize the performance of diffusion models for inpainting. Pre-trained diffusion models, such as Stable Diffusion or DALL-E 2, are often fine-tuned on domain-specific datasets to adapt them to specialized inpainting tasks. Loss functions incorporating perceptual and adversarial components were used to improve the visual realism of the generated content. Furthermore, methods like DDIM (Denoising Diffusion Implicit Models) are employed to accelerate the reverse denoising process, reducing the computational cost of sampling while preserving the quality of results.

3 Image Restoration Evaluation Indicators

3.1 Inception Score

IS (Inception Score) is a commonly used evaluation metric to measure the quality and diversity of images generated by the model. The calculation formula is shown in formula (3).

$$IIS = \exp\left(Ex{\sim}pg\left[KL\left(p(y|x)\middle|\middle|p(y)\right)\right]\right) \tag{3}$$

The term refers to the image category distribution generated by pre-training the Inception v3 model, which outputs a softmax probability. In contrast, represents the marginal distribution of the image generation process, calculated as the average of the category distributions across all generated images. To quantify the difference between these two distributions, this formula uses the Kullback-Leibler (KL) divergence, denoted as. This metric measures the divergence between p and q. Therefore, a higher Inception Score (IS) suggests that the generated images demonstrate both strong recognizability and greater diversity. As shown in Table 1, the Denoising Diffusion GAN achieves a higher IS value than the other models, indicating that it generates images with superior quality and greater variety.

Table 1. A numerical comparison of the Diffusion-GANs, Diffusion, and GANs models based on the Inception Score (IS) index using the CIFAR-10 dataset [8, 13, 14].

Model	IS
Denoising Diffusion GAN	9.63
Gated PixelCNN	4.60
PixelIQN	5.29
EBM	6.78
NCSN	8.87
SNGAN	8.82±0.05
SNGAN-DDLS	9.09±0.10
DDPM	9.46
Score SDE(VE)	9.89
Score SDE(VP)	9.68
Probability Flow(VP)	9.83
LSGM	9.87
DDIM,T=50	8.87
FastDDPM,T=50	8.98
SNGAN	8.22
SNGAN+DGflow	9.35
AutoGAN	8.60
TransGAN	9.02
StyleGAN2 w/o ADA	9.18
StyleGAN2 w/ ADA	9.83
StyleGAN2 w/ Diffaug	9.40

3.2 Peak Signal-to-Noise Ratio

Peak Signal-to-Noise Ratio(PSNR) is a widely recognized image quality evaluation metric in the field of computer science. It is primarily used to assess the reconstruction quality of images and plays a crucial role in tasks such as image restoration, compression, and denoising. PSNR works by comparing the difference between the original image and its reconstructed counterpart, providing a quantitative measure of how closely the reconstructed image approximates the original image in terms of visual fidelity. The calculation formula is shown in formula (4).

$$PSNR = 10 * log_{10}\left(\frac{Imax^2}{MSE}\right) \tag{4}$$

Where is the maximum value of the pixel value in the image, and (Mean Square Error) represents the average of the pixel differences between the original and reconstructed images. Therefore, a higher PSNR value indicates a smaller difference between the generated and original images, resulting in better quality. Ideally, an infinite PSNR value indicates that the generated image is identical to the original image. As shown in Table 2, ESRGAN, its improved version, and SRDiff achieve higher PSNR values, demonstrating that images repaired using GANs and Diffusion model exhibit better quality and more realistic visual effects.

 https://doi.org/10.1051/itmconf/20257302032

Table 2. A comparison of PSNR values for the GANs, Diffusion, and Flow models on the DIV2K dataset [15, 16].

Model	PSNR↑
SRGAN	27.16
ESRGAN-PI	27.80
ESRGAN	28.16
RankSRGAN	28.01
SR3	25.90
SRDiff	27.41
IR-SDE	25.90
RRDB	29.44
ESRGAN	26.22
SRFlow	27.09
HCFlow	26.61

3.3 Structural Similarity Index

Structural Similarity Index (SSIM) is a widely employed image quality evaluation metric in the field of computer vision. It measures the similarity between two images, focusing particularly on structural information, brightness, contrast, and texture. Unlike traditional metrics such as Mean Squared Error (MSE) and PSNR, SSIM offers a more perceptually accurate evaluation by better simulating the way the human visual system interprets image quality. This makes SSIM particularly effective in capturing nuances that traditional metrics might overlook, delivering superior performance in tasks involving image quality assessment. The calculation formula is shown in formula (5).

$$SSIM_{(x,y)} = \frac{(2\mu_x\mu_y + C1)(2\sigma_{xy} + C2)}{(\mu_x^2 + \mu_y^2 + C1)(\sigma_x^2 + \sigma_y^2 + C2)} \tag{5}$$

where and are the mean values (brightness) of the two images x and y, respectively; are the standard deviations (contrast) of the images x and y, respectively; and is the covariance (structure) of the two images. C1 and C2 are constants used for stable calculations. The closer the SSIM value is to 1, the more similar the repaired image is to the original in terms of structure, brightness, and contrast, indicating better repair quality. As shown in Table 3, the GANs model values were above average, with the improved version showing minimal fluctuation. This indicates that the repair algorithm effectively restores the overall image quality while preserving the structure and details of the original image.

Table 3. A comparison of SSIM values for the GANs, RRDB, Flow, and Diffusion model on the DIV2K dataset [15, 16].

Model	SSIM↑
SRGAN	0.7600
ESRGAN-PI	0.7653
ESRGAN	0.7752
RankSRGAN	0.7652
SR3	0.75
SRDiff	0.79
IR-SDE	0.66
RRDB	0.84
ESRGAN	0.75
SRFlow	0.76
HCFlow	0.74

4 Prospect

This study seeks to analyze and contrast the methodologies employed by various models in the realm of image repair. The objective is to find approaches that foster the advancement of image restoration technology, thereby contributing to its broader development and enhancing the broader development of image restoration technologies, enabling more efficient and high-quality applications in diverse domains. At present, traditional image repair techniques based on adversarial generation networks and Diffusion Models still have some limitations. However, the advent of Diffusion-GANs model has brought new techniques and ideas to the field of image repair. By combining the efficient training ability of the GANs model and the high quality generation ability of the Diffusion model in the framework of innovative multitask learning hybrid architecture, can realize the efficiency and quality in the process of image repair and generation balance, has the potential to overcome the existing limitations of image recovery technology, This article also proves that the combined model has some advantages in the efficiency and effect of image processing.

5 Conclusion

This article examines GANs and diffusion Models, and introduces the hybrid model Diffusion-GANs for image inpainting. Specifically, we focus on the image inpainting and release methodology of the improved LSGAN and other image inpainting techniques that leverage the power of generative adversarial networks. In the section on Diffusion Models, this article primarily focuses on the characteristics of the model and the methodologies for enhancing its performance. Diffusion models are generative models that iteratively add noise to data during a forward process and learn to reverse this process (denoising) to generate new samples. This article also explores the current techniques and various models employed in the realm of image painting through Diffusion.

This article also discusses in detail an innovative combination model—of diffusion—gans that has emerged in recent years, which brings new ideas and methods to improve the field of image inpainting technology. In terms of the evaluation index of image inpainting, by comparing the performance of GANS, Diffusion Models in the PSNR, and SSIM, they reveal the performance of different models on differences between old and new images, maintenance of image structural information, and differences in image pixel values. It is noteworthy that, given that Diffusion-GANs models are predominantly utilized for image generation at present, this study has primarily compiled metrics for this model's image generation capabilities. However, Diffusion-GANs also hold substantial promise for image inpainting and may offer more efficient methodologies for this purpose in the future. Simultaneously, the combined model provides an excellent perspective and possibility for solving problems such as image repair and image generation.

Authors Contribution

The names of the authors were presented alphabetically, and each author made an equal contribution.

References

1. J. Kim, et al., Combining GANs and Diffusion Models for Image Restoration, Comput. Vis. Image Underst. **165**, 15-28 (2022)

2. J. Ho, A. Jain, P. Abbeel, Denoising diffusion probabilistic models, Adv. Neural Inf. Process. Syst. **33**, 6840-6851 (2020)
3. J. Smith, X. Zhang, Advances in Deep Learning for Image Restoration, J. Comput. Vis. **35**(4), 123-140 (2022)
4. Z. Luo, F.K. Gustafsson, Z. Zhao, et al., Taming diffusion models for image restoration: A review, arXiv preprint arXiv:2409.10353 (2024)
5. B. Xia, Y. Zhang, S. Wang, et al., Diffir: Efficient diffusion model for image restoration, in Proceedings of the IEEE/CVF International Conference on Computer Vision, (2023), 13095-13105
6. L.T. Trinh, T. Hamagami, Latent Denoising Diffusion GAN: Faster Sampling, Higher Image Quality, IEEE Access. **12**, 78161–78172 (2024)
7. D. Kuznedelev, V. Startsev, D. Shlenskii, et al., Does Diffusion Beat GAN in Image Super Resolution?, arXiv preprint arXiv:2405.17261 (2024)
8. S. Nasr Esfahani, S. Latifi, Image Generation with GANs-based Techniques: A Survey, Int. J. Comput. Sci. Inf. Technol. (2019)
9. Y. Chen, Research on Image Inpainting Method Based on Improved LSGAN, East China Jiaotong University (2022)
10. J. Wei, Research on Image Inpainting Algorithm Based on GAN, North China University of Technology (2023)
11. R. Rombach, A. Blattmann, D. Lorenz, et al., High-resolution image synthesis with latent diffusion models, in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, (2022), 10684-10695
12. Y. Song, J. Sohl-Dickstein, D.P. Kingma, et al., Score-based generative modeling through stochastic differential equations, arXiv preprint arXiv:2011.13456 (2020)
13. H. Xiao, X. Wang, J. Wang, et al., Single Image Super-Resolution with Denoising Diffusion GANs, Sci. Rep. **14**, 4272 (2024)
14. Z. Xiao, K. Kreis, A. Vahdat, Tackling the generative learning trilemma with denoising diffusion GANs, arXiv preprint arXiv:2112.07804 (2021)
15. W. Li, K. Zhou, L. Qi, L. Lu, J. Lu, Best-Buddy GANs for Highly Detailed Image Super-resolution, Proc. AAAI Conf. Artif. Intell. **36**(2), 1412-1420 (2022)
16. Y. Yuan, C. Yuan, Efficient Conditional Diffusion Model with Probability Flow Sampling for Image Super-resolution, Proc. AAAI Conf. Artif. Intell. **38**(7), 6862-6870 (2024)