Current Study on Image Restoration Leveraging CNNs and GANs

Feng Cai[1], Jingxu Peng[2*] and Peng Zhou[3]

[1]South China University of Technology, Panyu District, Guangzhou 510006, China
[2]Tianjin University of Technology, Xiqing District, Tianjin 300384, China
[3]Chengdu College of University of Electronic Science and Technology of China, Western High-Tech District, Chengdu 611731, China
[*]corresponding authors: lukasy@stud.tjut.edu.cn

Abstract. In recent years, advances in image restoration have prominently featured Convolutional Neural Networks (CNNs) and Generative Adversarial Networks (GANs). Specifically, in seismic image denoising, the DnCNN model, rooted in CNN, employs residual learning combined with batch normalization for a streamlined approach to denoising. Alternatively, SRGAN, leveraging GAN and super-resolution techniques, trains its generative model using a perceptual loss function, accentuating perceptual disparities between generated and authentic images to boost visual quality. Recognizing the importance of preserving salient seismic data, the Constrained-DnCNN model has been introduced, refining seismic data interpretation. Additionally, the CR-SRGAN model, targeting super-resolution of artifacts and color restoration, deviates from conventional training datasets acquired via high-resolution image interpolation and downsampling. Upon comparative analysis, the DCGAN, amalgamating features of CNN and GAN, stands out for harnessing CNN's robust feature extraction. This bolsters the model's capacity to adeptly fit real data distributions. DCGAN's potential spans aspects like multimodal fusion, model interpretability, trustworthiness, and image restoration quality, marking it a promising research avenue.

Keywords: CNNs, GANs, Image Restoration, DnCNN, SRGAN

1 Introduction

In the age of digital information, images have emerged as primary methods for accessing and utilizing informational resources. As image processing technology continually advances and its applications broaden, an increasing number of sectors are opting for images as their chief medium for information transmission. These applications span diverse domains. However, during the creation and spreading of images, quality deterioration is inevitably encountered due to various factors, such as device performance, environmental lighting, and interference during transmission, most prominently manifesting as image noise. Noise can be defined based on various probability density functions and can be categorized into the following types [1]:

B. H. Ahmad (ed.), *Proceedings of the 2023 International Conference on Data Science, Advanced Algorithm and Intelligent Computing (DAI 2023)*, Advances in Intelligent Systems Research 180,
https://doi.org/10.2991/978-94-6463-370-2_35

Gaussian noise, impulse noise, uniform noise, gamma noise, Poisson noise, and Rayleigh noise. Such noise negatively impacts image quality, directly or indirectly affecting our efficiency in information retrieval and utilization. Moreover, due to the constraints of modern computing and storage capabilities, images are often stored or transmitted in lower resolutions. Lower-resolution images encapsulate fewer details and information. Therefore, effective denoising and resolution enhancement during the image preprocessing phase have become pivotal topics in the field of computer vision. Denoising images holds not only profound theoretical significance but also vast application potential [2]. In earlier times, image restoration predominantly relied on conventional image processing techniques. Common traditional restoration methods utilized filters for denoising [3], with mean filtering and median filtering being the most widely adopted. The underlying premise of both filtering techniques is to harness the information from surrounding pixels to restore or estimate the genuine value of a focal pixel. By operating within the neighborhood window of a given pixel, the objective is to minimize the noise-induced distortions in the image. However, filtering techniques usually depend on predefined rules and assumptions. While effective in certain scenarios, these methods often fall short due to the intricacy and diversity of images, limiting their applicability and performance.

In recent years, with the rapid evolution of artificial intelligence, especially deep learning, image restoration methodologies grounded in deep learning have garnered extensive research and adoption. These approaches distance themselves from pre-established rules or hypotheses; instead, by training on extensive image datasets, they autonomously discern intricate image structures and distributions, thus facilitating image restoration. Notably, CNNs and GANs, owing to their potent representational learning and generative capabilities, have achieved remarkable results in the realm of image restoration. This review delves into the primary methodologies of CNNs and GANs in image restoration and their associated applications. We will assess the performance of both techniques across varied scenarios, provide a holistic appraisal of their merits and limitations, and further probe into potential enhancement strategies and future trends.

2 Image restoration methods based on CNNs

2.1 CNNs Architecture and Principles

CNNs are a pivotal deep learning framework inspired by human visual perception, predominantly utilized in the domain of image processing. The fundamental architecture of CNN encompasses: an input layer, convolutional layers, activation layers, pooling layers, and a fully connected layer.

The input layer initiates the CNN by receiving the raw image and typically undergoes preprocessing tasks such as normalization and mean subtraction to standardize data range. Subsequently, the convolutional layers employ a variety of filters to extract local features from the image, reducing network parameters through weight sharing. This convolution operation inherently involves each neuron forming local connections with its input, subsequently generating a weighted sum with the associated connection weights and adding a bias to derive the neuron's input value. Following the convolutional layer, an activation layer is usually appended, tasked with nonlinearly mapping the preceding inputs, thus enhancing the expressive capacity of the network. The pooling layers, placed after, perform feature down-sampling and data compression, bolstering the network's robustness against overfitting and further reducing parameter counts.

Finally, the fully connected layer amalgamates and transforms the various local features processed by prior convolutional and pooling layers. Through weight matrices, it aggregates these low-level features into more abstract, high-level representations. This layer projects the features from the convolutional, pooling, and activation layers further into the sample label space. To realize this, the fully connected layer typically employs a Softmax function, translating network outputs into class probabilities, ultimately guiding the final classification decision [3-5]. An example of a convolutional neural network Architecture is illustrated in the Figure 1.

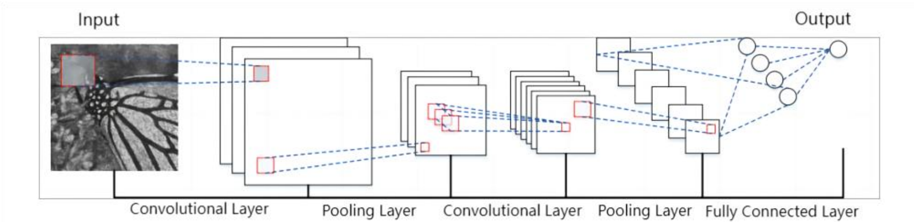

Fig. 1. Example of convolutional neural network architecture [2]

2.2 DnCNN Denoising Network

In recent years, CNNs have emerged as the cornerstone for image restoration. Various architectures tailored for specific tasks, such as VGG, ResNet, and U-Net, have been introduced with notable success in image processing. Among them, for image denoising tasks, the DnCNN framework has gained considerable attention and adoption due to its remarkable restoration capabilities.

DnCNN, proposed by Zhang et al. [6], is a denoising model that integrates residual learning with batch normalization. The essence of residual learning lies in learning the difference between the output and input, rather than focusing directly on the output. In a denoising context, this difference corresponds to the noise component. While traditional denoising models aim to map noisy images to clean ones, DnCNN endeavors to learn the noise itself. This learning does not directly rely on conventional residual blocks but employs a global residual learning strategy.

The DnCNN architecture encompasses multiple convolution layers, ReLU activation functions, and a batch normalization layer. Typically, the initial convolution layer captures basic image features, while intermediate layers target intricate noise characteristics. The network employs Mean Squared Error (MSE) as its loss function to gauge the discrepancy between the predicted and actual noise. During training, the objective is to minimize this MSE. For inference, the predicted noise is subtracted from the original noisy image, yielding the denoised result.

Batch Normalization (BN) serves as an optimization strategy for training deep networks. It computes feature mean and variance during each training batch, normalizing these features accordingly. Subsequently, learnable scale and shift parameters adjust the features. BN mitigates the 'internal covariate shift' phenomenon, enhancing training stability and allowing for higher learning rates, thus expediting the training process.

In sum, DnCNN amalgamates residual learning with batch normalization to address image denoising succinctly. While residual learning hones in on precise noise characteristics, batch normalization ensures training consistency and efficiency. These mechanisms synergistically constitute a potent image denoising framework. DnCNN Architecture was shown in the Figure 2.

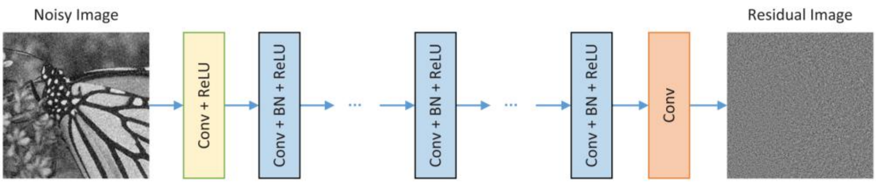

Fig. 2. DnCNN architecture [6]

2.3 Application of DnCNN in Image Restoration

Zhang et al. advanced the DnCNN architecture, introducing an enhanced model named Constrained-DnCNN specifically for denoising seismic data [7]. This model is attuned to the nuances of seismic information, striving to precisely retain valuable details and features inherent in the data during the denoising process. The Constrained-DnCNN integrates a constrained convolution layer, originally devised for image tampering detection. This layer discerns tampering traces by recognizing alterations in local pixel relationships. It simulates prediction error filtering, creating feature maps of the prediction error domain as low-level tampering markers. These are subsequently provided to the CNN to extract high-level detection features. Within this model, the feature maps generated by the constrained convolution layer are added to the input seismic data, thereby accentuating the data's intricate features. The model's design also draws inspiration from the deep learning residual networks, particularly utilizing skip connections. This ensures the capture and preservation of pivotal information from the original data, rather than merely discarding potentially noisy sections, guaranteeing post-denoising data integrity.

To attest to the model's utility and efficacy, Zhang and colleagues conducted extensive experiments. When benchmarked against DnCNN and other conventional denoising techniques under varied noise intensities, Constrained-DnCNN not only demonstrated superior denoising outcomes but also adeptly retained crucial details and information within the seismic data. This optimized iteration of DnCNN furnishes a more accurate foundation for seismic data interpretation, holding significant application potential in the domain of seismic data denoising. The comparison images for ‘Original Data’, ‘Noisy Data (SNR5)’, and ‘DnCNN Denoised’ are displayed in Figure 3. Comparative images of ‘Constrained-DnCNN Denoised’, ‘Curvelet Transform Denoised’, and ‘Numerical Anisotropic Diffusion Filtering Denoised’ are presented in Figure 4.

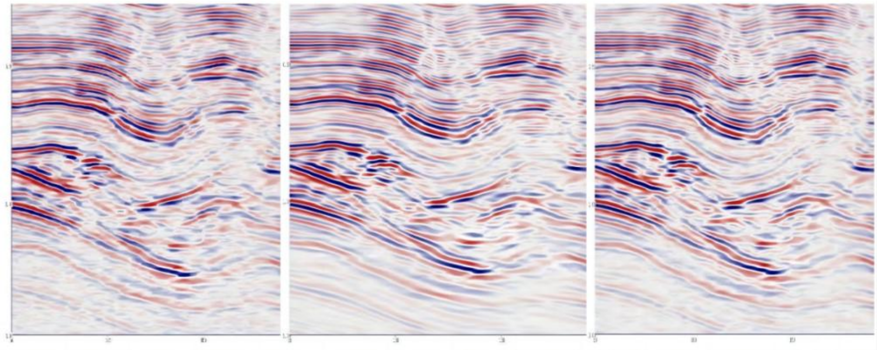

Fig. 3. Original data image noisy data image (SNR5); DnCNN denoised image [7]

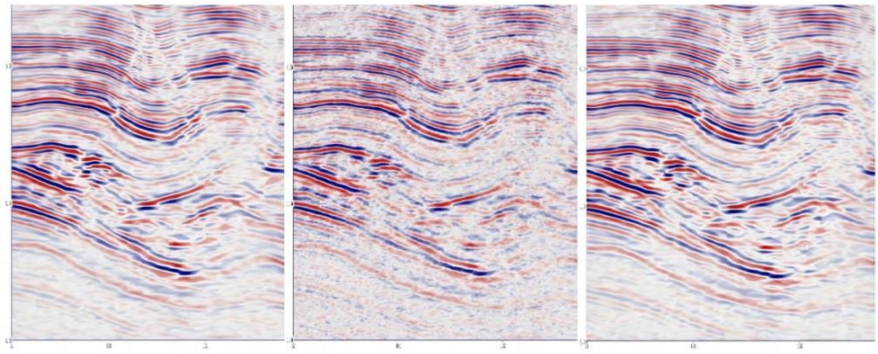

Fig. 4. Constrained-DnCNN denoised image; Curvelet transform denoised image numerical anisotropic; Diffusion filtering denoised image [7]

3 Image Restoration Methods based on GANs

3.1 GANs Architecture and Principles

GANs [8] was proposed in 2014 by Goodfellow et al. A GAN network consists of at least two parts - Generative Network G and Discriminative Network D, where the Discriminative Network comes to determine the authenticity of the content generated by the Generative Network [9]. During training, the purpose of the generative network is to generate content that the discriminative network considers true, while the purpose of the discriminative network is to discriminate as much as possible the false outputs generated by the generative network, in order to promote and inhibit each other [9]. An example of GANs was shown in the Figure 5.

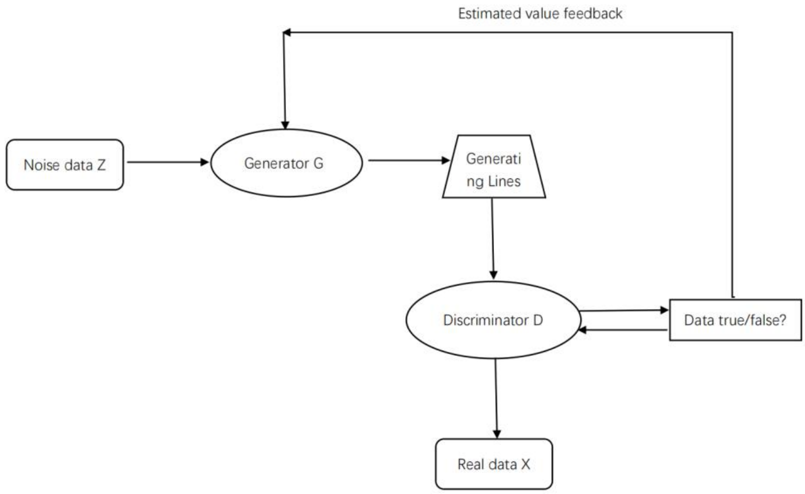

Fig. 5. Example of Generative Adversarial Networks [5]

3.2 SRGAN

Ledig et al who used GAN networks in the field of image super-segmentation reconstruction [10], proposed an SRGAN image super-segmentation reconstruction model based on GANs, which encourages the pursuit of a realistic sense of the image instead of pursuing a high peak signal-to-noise ratio, which makes the generated image closer to the detailed texture of the real image.

This model consists of two parts, generator network (generator) and discriminator network (discriminator) [8]. The quality of the generated images is continuously improved by means of adversarial learning. The SRGAN model accepts low resolution images as input and generates high resolution images through deep residual network. The discriminator model is responsible for discriminating the real image and introduces perceptual loss to reconstruct the detailed features of the image. SRGAN Architecture was shown in the Figure 6.

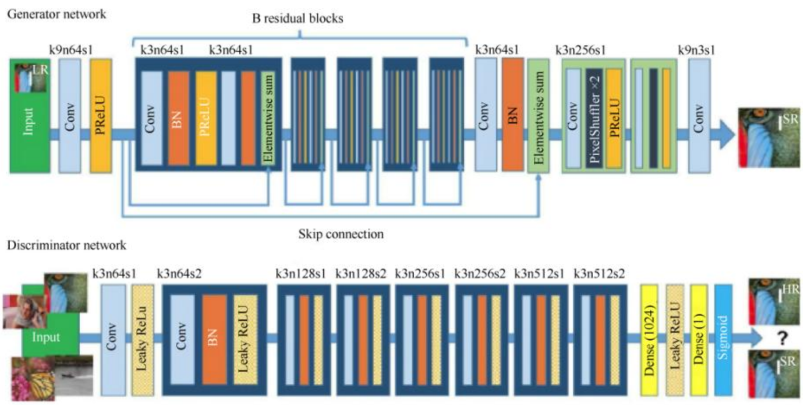

Fig. 6. SRGAN Architecture [11]

In order to improve the realism of the generated images, traditional single-frame image super-segmentation algorithms use MSE as a loss function, but high-frequency texture details are lost, which results in less realistic subjective visualization of the generated images. To solve this problem, SRGAN uses a perceptual loss function to train the generative model. This loss function focuses on the perceptual difference between the generated image and the real image to enhance the visual quality of the generated image. The loss function expression for the generative model is as follows:

$$l^{SR} = l_X^{SR} + 10^{-3} l_{Gen}^{SR} \tag{1}$$

l_X^{SR} is the content loss and $10^{-3} l_{Gen}^{SR}$ is the adversarial loss. The two loss functions add up to the perceptual loss [13].

3.3 Application of SRGAN in Image Restoration

Zhu [13] and others optimized for SRGAN and proposed a CR-SRGAN model for super-resolution reconstruction of cultural relics as well as color restoration. The model is different from the low-resolution dataset traditionally obtained by interpolation and down sampling of high-resolution images to train the model, according to the characteristics of fading and dark old images of cultural relics to deal with the degradation of the original dataset in the interpolation of the basis of the dark old and noise processing to obtain the corresponding low-resolution images, and finally through the training of the network model, to obtain the color restoration of super-resolution cultural relics images.

The process of color and texture restoration of dark old artifact images by CR-SRGAN model is mainly divided into 4 steps：

1) Preprocess the data set to obtain the data set corresponding to the high and low resolutions according to the characteristics of cultural relics images.

2) Construct generative and discriminative networks, and train the image super-resolution model with the high- and low-resolution datasets obtained by ordinary down sampling method only.

3) using the generative network and discriminative network constructed in step 2), train the artifact image super-resolution model CR-SRGAN with the high- and low-resolution dataset obtained by noise addition and dark old processing in step 1); and

4) Reconstruct high-resolution cultural relics images using the generated model trained in step 3). CR-SRGAN Architecture was illustrated in the Figure 7.

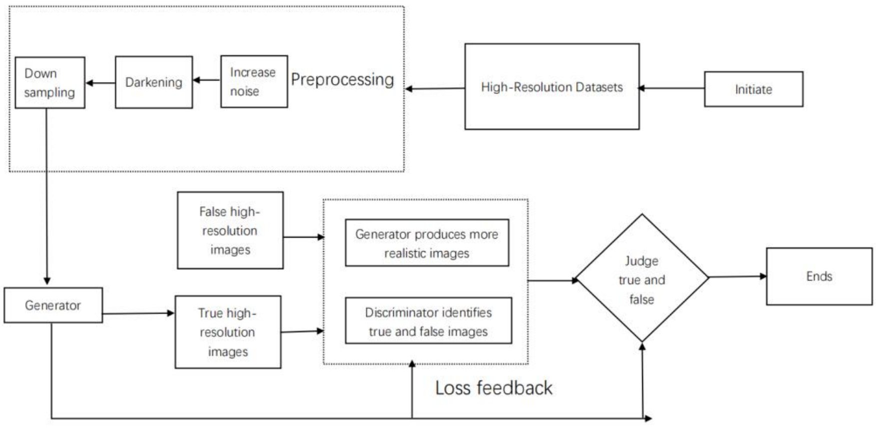

Fig. 7. CR-SRGAN Architecture [13]

Figure 8 and Figure 9 respectively showcase the comparisons of high-resolution images restored using 'Bicubic', 'SRGAN', 'CR-SRGAN', and manual techniques.

Fig. 8. High-resolution restoration images of murals using Bicubic; SRGAN; CR-SRGAN; Manual techniques [10]

Fig. 9. High-resolution restoration images of bronze artifact using Bicubic; SRGAN; CR-SRGAN; Manual techniques [10]

4 Discussion

The image restoration method based on deep learning will be the main research direction to solve the problem of image restoration in the future. The existing methods based on CNN and GAN have achieved some results in theoretical research and practical application, and each has its own advantages and disadvantages.

PSNR and SSIM are used to evaluate the quality of the image. PSNR is to evaluate the image quality from the perspective of the machine, and to evaluate the image quality is to calculate the difference between the pixels of the two images. SSIM index is to stand in the human perspective and evaluate the overall image quality through brightness, contrast, and structure. Higher values of PSNR and SSIM indicate higher image quality, where the maximum value of SSIM is 1 [1].

4.1 Discussion on CNNs-based Image Restoration

Table 1 shows the performance index comparison table of different algorithms in part of 2.3 seismic image restoration, that is, SSIM and PSNR of different denoising algorithms under different SNR are compared. SNR is the input SNR after adding Gaussian noise to the seismic data. In the case of SNR changes, SSIM and PSNR of each group increase with the increase of SNR, showing a positive correlation. These methods can obtain better denoising effect.

Table 1. different algorithms in seismic image restoration

restoration method	SNR_IN	PSNR(db)	SSIM
Curvelet transform	3	32.1949	0.9663
	5	33.7859	0.9700
	10	34.5855	0.9715
Complex domain anisotropic diffusion filtering	3	31.7441	0.9513
	5	32.7786	0.9623
	10	34.3329	0.9759
DnCNN	3	31.2622	0.9660
	5	32.8707	0.9751
	10	36.5017	0.9877
Constrained-DnCNN	3	31.4575	0.9703
	5	33.0928	0.9784
	10	36.8037	0.9896

The traditional methods of curvelet transform and complex domain anisotropic diffusion filtering have their own advantages and disadvantages in denoising effect, and have different denoising performance for different degrees of SNR. However, DnCNN and Constraind-DNCNN, two deep neural network denoising methods, can be processed independently without too much human intervention, and have better performance in several methods.

Through data analysis, comparing the Constraind-DNCNN discussed in this paper with DnCNN, we can see that in the evaluation of SSIM and PSNR, Constraind-DNCNN is about 0.33% and 0.71% higher than DnCNN on average. Therefore, the effect of Constraind-DNCNN is better, which quantitatively shows that this method has certain performance improvement.

4.2 Discussion on GANs-based Image Restoration

In Figures 8 (a), (b) and (c) show the color restoration and super-resolution reconstruction images of dark old murals using different methods, and (d) show the high-resolution images manually restored.In Figure 9 (e), (f), and (g) are the color restoration and super-resolution reconstruction images of faded bronzes using different methods, and (h) is the high-resolution image restored by hand. In Figure (a) and (b), there is almost no change in color during super-resolution. In figure (c), the repaired image becomes brighter as a whole, and the colors of the clothes of the four princesses change significantly, which is consistent with the clothes colors of the manually restored image (d). Compared with (e) and (f), the color of (g) is improved.

Table 2 shows the performance index comparison table of different methods in mural and bronze image restoration, that is, the PSNR and SSIM of different denoising algorithms under different signal-to-noise ratios are compared. The data show that the PSNR and SSIM values of the faded color restoration and super-resolution reconstruction images obtained by the CR-SRGAN model in this paper are improved compared with the comparison methods. Compared SRGAN model CR - SRGAN average PSNR improved 0.135 dB, SSIM increase by 0.01 on average. Compared with the Bicubic model, the PSNR of CR-SRGAN model is increased by 0.86dB and the SSIM is increased by 0.04. The results show that the CR - SRGAN model is better than the contrast.

Table 2. different algorithm in super-resolution reconstruction images of old murals

restoration method	PSNR(db)	SSIM	PSNR(db)	SSIM
Bicubic	11.48	0.53	15.91	0.81
SRGAN	11.64	0.56	17.20	0.84
CR-SRGAN	11.71	0.58	17.40	0.84
	Mural		Bronze ware	

4.3 Comparative and Evaluative Analysis

CNN can effectively capture the spatial features of images in image restoration tasks, and extract local and global information through convolution and pooling operations. Widely used in the field of image processing; In the trained CNN model, the pre-trained weights can be used for transfer learning to achieve better performance on the specific image restoration task. But in image restoration, CNN may generate some relatively conservative repair as a result, the lack of diversity. Image restoration based on GAN method can generate lifelike images, can be used to create high quality repair results; Can generate images of the diversity, help to create different styles and contents of the repair. But GAN training can suffer from mode collapse, mode collapse, etc., which requires careful hyperparameter tuning and tricks to solve.

DCGAN (Deep Convolutional GAN), or Deep Convolutional generative network model, is a combination of convolutional neural Network (CNN) and Generative Adversarial Network (GAN). Compared with the simple GAN network model, the basic principle of DCGAN is similar, but the generative network and discriminative network are improved, and a CNN network is used to implement them. In order to overcome the problems of simple GAN network to a certain extent, the CNN network is improved as follows: a fully convolutional network is used, that is, a convolutional layer is used to replace the pooling layer, and the network is trained to learn the upsampling and downsampling methods. Cancel the convolution full connection layer after layer and improve the convergence speed of the network; A BN layer was added to the other layers except the output layer of the generative network and the discriminative network to ensure the strength of the gradient in the network. Tanh function is used as the activation function of the output layer of the generative network, and leakyReLU function is used as the activation function of the discriminative network.

DCGAN takes advantage of the powerful feature extraction ability of convolutional networks to enhance the learning ability of the model, so that it can more effectively fit the real data distribution. Moreover, through the unique structure of DCGAN, it can also solve the problems of training instability, gradient descent disappearance, model collapse and other problems of simple GAN to a certain extent. However, DCGAN may require more parameters and computational resources than a simple CNN. Although DCGAN is relatively stable, the hyperparameters still need to be carefully tuned to obtain the best generation results.

Image restoration has become a hot research topic. Based on the understanding of the application of deep learning to image inpainting, this paper puts forward several prospects for development.

1) multimodal fusion. DCGAN generating capacity can be applied to multimodal image restoration. By fusion of different modal information, such as visible light image, such as thermal infrared image, DCGAN can generate much more comprehensive, more accurate modal repair results.

2) can be interpreted and trust: with the gradual application of deep learning model, the interpretability of the model and the demand of the trust property is becoming more and more big. Future research will look at how to make the model's decisions more interpretable and how to improve the user's sense of trust in the model.

3) image restoration quality control: DCGAN generated results affected by network structure and parameters of. Future research can explore how to control the quality of the resulting image, make repair results more controllable and adapt to the actual demand.

5 Conclusion

In this paper, n this paper, a survey of image restoration methods based on deep learning is presented, introducing three main categories of methods. Among them, the image restoration method based on CNN can effectively capture the spatial features of images by convolution and pooling operations of local and global information extraction. Gan-based image restoration can effectively capture high-level semantic information and keep the global texture and structure semantically consistent, but there are problems such as difficult network training and gradient explosion. DCGAN network is a combination of CNN and GAN, which uses the powerful feature extraction ability of convolutional network to enhance the learning ability of the model, and through the unique structure of DCGAN, it can also solve the problems of training instability, gradient descent disappearance, model collapse and other problems of simple GAN to a certain extent. It is a research scheme worth exploring and has great potential.

Authors Contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

1. Ren, X.: Improvement of DnCNN Image Denoising Network Based on Deep Learning. Nanjing University (2020).
2. Liu, Z.: Research on Transform Domain Image Restoration Technology Based on Deep Learning. Xi'an University of Electronic Science and Technology (2020).
3. Hu, S.: Discussion on Image Denoising Technology with Improved Adaptive Weighted Mean Filtering. Journal of Pu'er University, 38(06), 41-43 (2022).
4. Zhou, F., Jin, L., Dong, J.: A Review of Convolutional Neural Network Research. Journal of Computer Science, 40(6), 1229-1251 (2017).
5. Liu, Y., Peng, H., Dai, Y., et al.: Research Progress on Image Restoration Based on Deep Learning. Software Guide, 22(07), 220-226 (2023).
6. Zhang, K., Zuo, W., Chen, Y., Meng, D., Zhang, L.: Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising. IEEE Trans Image Process, 26(7), 3142-3155 (2017).
7. Zhang, C., Wen, X., Zhang, X., et al.: Seismic Data Denoising Method Based on DnCNN and Constraint Convolution [J/OL]. Advances in Geophysics, 1-18 (2023).
8. Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., et al.: Generative adversarial nets. In: Proceedings of Annual Conference on Neural Information Processing Systems, New York: NIPS Press, 2672-2680 (2014).
9. Wang, F.: Research on Image Quality Evaluation and Improvement Algorithms and Applications.
10. Ledig, C., Theis, L., Huszar, F., et al.: Photo-realistic Single Image Super-resolution Using a Generative Adversarial Network [EB/OL], (2016).
11. Li, H., Zheng, Q., Tao, R., et al.: Review of Image Super-Resolution Research Based on Deep Learning. Journal of Graphics, 44(01), 1-15 (2023).
12. Hu, Y.: Research on Mural Image Restoration and Super-resolution Reconstruction Method Based on GAN. North University of China (2023).
13. Zhu, X., Lei, Q., Wu, X.: Super-Resolution Reconstruction and Color Restoration of Cultural Relic Images Based on Generative Adversarial Networks. Journal of Xi'an Engineering University, 35(03), 86-92 (2021).