

Research Advanced in Image Inpainting based on Deep Learning

Shengkun Zhang*

International School, Beijing University of Posts and Telecommunications, Beijing, China

* Corresponding Author Email: 2022213255@bupt.cn

Abstract. Image inpainting has always been a hotspot in the computer vision community, which aims to restore damaged or missing parts of an image, ensuring that the restored image is physically reasonable, visually pleasing, and consistent in texture and structure with the original image. Early inpainting methods primarily depend on diverse image processing technologies, such as sample replication, partial differential equation solving, etc. As artificial intelligence rapidly progresses, image inpainting methods based on deep learning exploit robust feature extraction capabilities to extract useful information from vast amounts of data, contributing to more robust and natural inpainting effects, especially when dealing with complex and high-dimensional scenes. This paper provides a review of deep learning-based image inpainting technologies, categorized into four main types: autoencoders, generative adversarial networks (GANs), recurrent neural networks (RNNs), and convolutional neural networks (CNNs). It summarizes and analyzes the basic principles, advantages, disadvantages, and performance of these methods in image inpainting, as well as introduces commonly used datasets and quantitative evaluation metrics. Additionally, the paper discusses the principal challenges currently faced in this domain and provides insights into future research directions for deep learning-based image inpainting.

Keywords: Image inpainting, deep learning, autoencoders, generative adversarial networks.

1. Introduction

In the digital age, images have become an increasingly important medium for information transmission. Whether through the widespread use of social media or the preservation of digital archives and cultural heritage, high-quality images are indispensable. However, images often suffer damage due to various factors such as transmission errors, degradation of storage media, or external environmental factors. These damages affect the visual presentation and lead to the loss of critical information. Therefore, efficient and automatic image inpainting technologies are attracting increasing attention from both academics and industry.

During the previous decades, image inpainting technologies have evolved from traditional image processing techniques to advanced machine learning algorithms. Early image inpainting methods primarily focused on region-based filling and texture synthesis techniques. While these methods were effective in certain scenarios, they often required manual feature selection and relied on predefined algorithms, making them less effective when dealing with complex structures and textures in images.

With the advent of the 21st century, particularly over the last decade, the rise of deep learning has greatly advanced the field of image inpainting, opening up new pathways for automatic image restoration. Deep learning-based methods have become the mainstream approach in image inpainting research. By modeling high-level abstract features in image data, these techniques can restore image content more accurately and naturally, outperforming traditional methods in terms of detail recovery and consistency, while also reducing the need for manual intervention. Currently, Image inpainting methods driven by deep learning techniques can be primarily categorized into:

(1) CNN-based inpainting methods. CNN-based image restoration methods use the powerful feature extraction capabilities of convolutional neural networks (CNNs) to predict and fill in missing parts by learning the local features of the image. These methods can capture the spatial structural information of the image and are suitable for processing local damage in the image.

(2) RNN-based inpainting methods. RNN-based image restoration methods aim to use the sequence data processing capabilities of recurrent neural networks (RNNs) to model the dependencies

between image pixels. These methods are suitable for processing image restoration tasks with time series characteristics, such as filling in missing parts row by row or column by column.

(3) GAN-based inpainting methods. Methods of inpainting that utilize Generative Adversarial Networks (GANs) employ a dual-component framework comprising a generator and a discriminator to enhance image restoration outcomes. In this setup, the generator produces images aiming to replicate real-world visuals to mislead the discriminator. Conversely, the discriminator evaluates whether the images are generated or authentic. This method excels in handling intricate textures and preserving the coherence of global structures.

(4) VAE-based inpainting methods. VAE-based image restoration methods use variational autoencoders (VAEs) to directly master the hidden representation of the image requiring restoration, followed by reconstructing it into the actual image. By adjusting the latent variables to control the generated image style, these methods can generate diverse restoration results.

Focusing on the above four categories of mainstream methods, the objective of this work is to deliver a detailed analysis of contemporary deep learning methods applied to image inpainting, offering researchers and practitioners a clear technological framework through a detailed technical overview. Additionally, it explores the challenges faced in the field and potential future research directions, aiming to push image inpainting technology toward a more efficient and intelligent development stage.

2. Method

2.1. VAE-based Image Inpainting Methods

Variational autoencoder provides a robust and effective framework for image inpainting tasks, demonstrating excellent performance in various image inpainting scenarios. Pathak et al. [1] proposed the Context Encoders network structure in 2016, combining the concepts of autoencoders with Generative Adversarial Networks (GANs). Context Encoders process features in an unsupervised learning manner, allowing the network to understand the content of the entire image and generate a reasonable hypothesis for the missing parts. In this model, a fully connected layer connects the encoder and decoder, enabling each unit in the decoder to infer the overall image content while significantly reducing the number of model parameters. Additionally, Context Encoders are trained using an adversarial learning strategy, jointly optimizing the reconstruction loss and adversarial loss. This helps minimize the differences between the reference and inpainted images, resulting in more natural and realistic inpainted regions. However, the adversarial loss constraint is applied only to the inpainted area, which limits the model's understanding of texture details and makes it challenging to maintain local consistency with the surrounding regions and ensure sharp image boundaries.

To address this issue, Iizuka et al. [2] replaced the fully connected layer with dilated convolutions and introduced a global context discriminator, which helps achieve inpainting with arbitrary resolutions and shapes while better maintaining global structural consistency. Liu et al. [3] further enhanced this approach by replacing standard convolutions with Partial Convolutions and proposed an inpainting model with an automatic mask update mechanism to restore images with irregular missing regions. They also introduced Coherent Semantic Attention (CSA), an attention mechanism that links damaged areas with known regions. The method performs convolutions only on pixels within known image regions (valid pixels), fully considering the internal correlation of the areas to be restored. This approach effectively avoids the impact of invalid pixels on feature extraction and aids in removing artifacts, including color mismatches and image blur. However, during the mask update process, convolving regions with different numbers of valid pixels can lead to the disappearance of edge information, and the invalid pixels in the missing areas may gradually fade away in deeper layers.

Yu et al. [4] improved this by proposing learnable gated convolutions in 2019, where the mask is updated through convolutions with feature maps as input. This method enables the network to automatically learn the mask's weights, making the boundaries of irregular inpainting more natural

and realistic. Nazeri et al. [5] developed a two-part adversarial system termed EdgeConnect, which is composed of an edge generator and a network dedicated to image completion. In this model, edge detection is used to generate an edge map first, which is then used as a guide for the image completion network to fill and restore the missing areas. This model made significant progress in restoring global structural information and effectively enhanced the consistency of details in the restored images, making texture transitions smoother and more natural.

2.2. GAN-based Image Inpainting Methods

Generative Adversarial Networks (GANs) have become an important technology in recent developments in image inpainting. GANs consist of two neural network models: the generator and the discriminator. These two networks undergo adversarial training, in which the generator works to create data resembling the true data distribution, while the discriminator aims to identify the differences between the artificial data and the original data. In this game-like process, the generator continuously improves the quality of the generated data until the discriminator can no longer differentiate between the generated and real data. By leveraging this adversarial training between the generator and discriminator, GANs are capable of producing highly realistic image inpainting results. Compared to traditional autoencoder methods, GANs demonstrate significant advantages in handling complex textures and maintaining global structural consistency. Overall, GANs replicate the distribution of real images through the generator and distinguish between generated and real images through the discriminator, greatly enhancing the naturalness and visual quality of the inpainted images. As a result, GANs have gradually become one of the most widely applied network models in image inpainting.

One of the early applications of GANs in image inpainting was the Context Encoders method mentioned above. This approach laid the foundation for subsequent research, opening new avenues for the application of GANs in image inpainting. In 2018, Yu et al. [6] proposed the DeepFill v1, which introduced local and global attention mechanisms, enabling GANs to more effectively concentrate on the intricate textures and details within images. The local attention mechanism in this method helps handle fine detail restoration in small areas of the image, while the global attention mechanism ensures the network maintains the overall structural consistency of the image. This combination results in more natural and realistic inpainting outcomes, particularly when restoring complex textures and irregularly shaped missing regions.

Following the work of Yu et al., researchers continued to explore more sophisticated and intelligent GAN architectures to tackle even more complex image inpainting tasks. DE-GAN [7] introduced a Variational Autoencoder (VAE) to enhance the naturalness and fidelity of facial image inpainting by embedding facial domain knowledge in the latent variable space of the VAE. DE-GAN incorporates specific prior knowledge about facial structures during the inpainting process, resulting in harmonious and natural-looking facial regions, particularly excelling in details such as eyes and mouths. Pyramid-VAE-GAN [8] further combined the advantages of the pyramid structure, VAE, and GAN, improving the learning ability of multi-scale features, significantly enhancing the clarity and detail of the inpainted images, although its training process is relatively complex. The CASI network [9] uses a fully convolutional design in the generator, combining Perceptual Loss and a joint loss function to capture high-level semantic information in images. Perceptual loss, through the integration of features from convolutional neural networks' intermediate layers, quantifies the visual variances between inpainted and original images to ensure that the restoration closely mirrors the original image. The complete convolutional architecture of the CASI network ensures the retention of spatial structure and enhances uniformity across various resolutions. Additionally, the network adopts a context aggregation mechanism, allowing it to effectively handle image distortion and blurriness in high-resolution image inpainting tasks, resulting in clearer and more natural inpainted images.

While the aforementioned image inpainting methods have made significant progress in terms of reconstruction quality and naturalness, Difficulties persist in creating superior, high-resolution

images without overextending computational resources. To address this issue, Zeng et al. [10] proposed an Aggregated Contextual Transformation GAN (AOT-GAN) for high-resolution image inpainting. AOT-GAN introduces an innovative generator architecture that enhances contextual reasoning by utilizing remote image context information through aggregated contextual transformation. This mechanism effectively integrates both global and local information, resulting in inpainted images with better overall structural consistency and richer details. Furthermore, AOT-GAN's discriminator is trained with a custom mask prediction task, focusing on fine-grained texture synthesis. This design overcomes common issues such as distortion and blurriness in high-resolution image generation, ensuring that inpainted images remain sharp and realistic even in high-resolution scenarios. Chen et al. [11] proposed the PMSCA-Res-U-Net, employing a partial multi-scale channel attention module. This strategy adjusts channel weights dynamically, tailored to the importance of specific image attributes, using multi-scale adaptive weighting. This mechanism not only improves the efficiency of feature extraction but also take full advantage of low-level feature information, significantly enhancing the quality of the inpainted images. While maintaining high-quality inpainting results, PMSCA-Res-U-Net effectively reduces computational resource consumption, achieving a good balance between inpainting quality and computational efficiency.

2.3. RNN-based Image Inpainting Methods

Recurrent Neural Networks (RNNs) belong to a group of neural network models equipped to process data in sequences, often used to handle information that exhibits continuity or depends on contextual relationships. In the field of image inpainting, the unique advantage of RNNs lies in their ability to effectively capture contextual relationships within image sequences, making them particularly well-suited for tasks involving sequential image restoration.

Xie et al. [12] introduced RNNs into image inpainting and proposed a model featuring Long Short-Term Memory (LSTM) networks. This variant of RNNs seeks to discern the relationships between the missing parts of an image and their immediate known surroundings, resulting in inpainting outputs with better continuity. By incorporating "memory" units and training on sequential data, LSTM networks can effectively learn the temporal and spatial characteristics of image content, ensuring consistency in the inpainting process. This method performs exceptionally well in handling the continuity features of images, especially in tasks such as video frame inpainting, where the sequential nature of the data is crucial.

Additionally, Liu et al. [13] proposed an image inpainting method based on Gated Recurrent Units (GRU). As another variant of RNNs, GRU has a simpler structure compared to LSTM while maintaining the long-dependency feature capability. Liu's research demonstrated that GRU performs well in image inpainting tasks involving large missing regions, producing more natural inpainting results. Furthermore, this method has lower computational complexity, making it suitable for real-time image inpainting applications. Zhang et al. [14] further expanded the application of RNNs in image inpainting by proposing a model based on bidirectional RNNs. Bidirectional RNNs consider both forward and backward information flows within an image, allowing for a more comprehensive capture of contextual relationships. By integrating information from both directions, this method delivers higher-quality and more consistent inpainting results, making the generated images appear more natural and realistic in terms of visual coherence.

Overall, the main advantage of RNN-based image inpainting methods is their ability to effectively process sequential image data, particularly excelling in video inpainting and tasks that require capturing long-term contextual dependencies. However, RNN models typically have a large number of parameters, requiring substantial training data for optimization. This results in relatively long training times and higher computational complexity, which can reduce efficiency when dealing with high-resolution images. As image inpainting technology continues to develop and RNN architectures are further optimized, RNN-based methods have the potential to play an increasingly significant role in tasks that require long-term dependencies and contextual understanding for image restoration.

2.4. CNN-based Image Inpainting Methods

Convolutional Neural Networks (CNNs), emblematic of advanced deep learning techniques, are broadly employed in numerous image processing endeavors. CNNs are structured primarily with convolutional layers, pooling layers, and fully connected layers. The convolutional layers use specific filters to process the input image and generate diverse feature maps while pooling layers perform downsampling and feature selection to retain essential information. Fully connected layers integrate the features extracted and pooled, ultimately outputting the restored image. This structure allows CNNs to preserve detailed information during feature extraction and image reconstruction while also effectively reducing computational complexity.

In image inpainting tasks, CNNs are typically used as feature extractors and restoration tools, learning the distribution of features in the input image and filling in the damaged or missing parts with plausible content. The multi-layered structure of CNNs enables them to capture a range of features, resulting in more natural and realistic inpainting outcomes. SRCNN [15] is among the pioneering CNN-based strategies for super-resolution image inpainting. This strategy involves a three-layer convolutional network that elevates low-resolution images to higher resolutions. First, Convolutional layers derive features from the low-resolution image, then the middle layers perform non-linear mapping, and finally, the reconstruction layer generates a high-resolution output. The simplicity and effectiveness of the SRCNN method have made it a foundational approach in subsequent super-resolution research.

3. Experiment

3.1. Datasets

In the evaluation of image inpainting algorithms, the choice of datasets and performance metrics is crucial for determining the effectiveness and accuracy of the methods. Various datasets are used to test different inpainting techniques, each with unique characteristics and challenges. Widely adopted datasets for research in image inpainting comprise CelebA, Paris StreetView, Places2, and MS COCO. These datasets are comprised of a vast range of images with different resolutions, textures, and complexities, making them suitable for testing the generalization ability and performance of inpainting algorithms.

CelebA is an extensive dataset filled with celebrity face images, utilized for testing the efficacy of models dedicated to facial inpainting. The challenge with this dataset lies in accurately restoring facial features and maintaining consistency in texture and structure across the entire image. Paris StreetView is a dataset composed of street view images from Paris. This dataset includes diverse urban scenes with complex textures, making it ideal for evaluating the ability of algorithms to handle intricate structures and details in inpainting tasks. Places2 is a large dataset containing various scenes and environments. It is widely used in general image inpainting tasks to test the algorithm's ability to generalize across different contexts and maintain global structural consistency. MS COCO encompasses a broad assortment of object categories. In the context of image inpainting, this dataset presents a challenging variety of textures and objects that require accurate restoration.

3.2. Evaluation Metrics

To gauge the performance of diverse inpainting algorithms, a variety of metrics are used to assess the quality of the restored images. Each metric focuses on a different aspect of the inpainting process, offering a comprehensive evaluation of the algorithm's effectiveness. The commonly used metrics for the image inpainting includes:

Peak Signal-to-Noise Ratio (PSNR): This widely recognized metric evaluates how similar a restored image is to its original by calculating the pixel-by-pixel differences between them. A higher PSNR value indicates better image restoration quality, with less distortion or loss of information.

Structural Similarity Index (SSIM): Structural Similarity Index (SSIM): SSIM assesses the perceptual quality of a restored image through a comparison of its structural details with those of the original image. It considers factors such as luminance, contrast, and texture. A higher SSIM value suggests that the restored image is closer in quality to the original image from a human visual perception perspective.

Visual Information Fidelity (VIF): VIF measures the fidelity of the visual information in the restored image, considering how well the algorithm preserves the important visual details. VIF is more aligned with human perception than PSNR or SSIM and gives a better sense of how realistic the restored image looks.

Mean Absolute Error (MAE): MAE quantifies the average absolute deviation in pixel values between the restored and the original images. A lower MAE value indicates more accurate restoration, as it signifies that the difference between the two images is minimal.

Fréchet Inception Distance (FID): FID is often used in GAN-based methods to evaluate the quality of the generated images. A lower FID score means that the restored images are closer to the real images in terms of their feature distribution, which is important for producing realistic results.

By combining these datasets and performance metrics, researchers can gain a comprehensive understanding of how well different image inpainting algorithms perform. These metrics not only measure the overall quality of the inpainted images but also highlight the strengths and weaknesses of each method in handling various image restoration challenges.

3.3. Algorithm Performance Comparison

The rapid development of deep learning techniques has introduced various image inpainting methods, including Autoencoders, Generative Adversarial Networks (GANs), Recurrent Neural Networks (RNNs), and Convolutional Neural Networks (CNNs). Each of these methods has unique architectures and advantages, providing diverse solutions for image inpainting tasks. However, these methods are suitable for different scenarios, and their performance varies across different datasets. Therefore, a comprehensive performance comparison is crucial.

To assess the effectiveness of these methods, this segment methodically evaluates the performance of several representative image inpainting approaches. Specifically, we will analyze their performance on standard datasets such as CelebA, Paris StreetView, Places2, and MS COCO using commonly used evaluation metrics such as PSNR and SSIM. The goal is to reveal the strengths and weaknesses of each method, providing valuable insights for upcoming studies.

As shown in Table 1, different methods show significant differences in performance. Autoencoder-based methods (such as Context Encoders), being developed earlier, perform well in handling simple structures and maintaining global consistency. However, when faced with high-resolution images and complex textures, the performance of autoencoder methods tends to be relatively weaker. This indicates that autoencoders have limitations in addressing complex details and high-resolution scenarios.

In contrast, Methods based on Generative Adversarial Networks (GANs), such as DeepFill v2, EdgeConnect, and AOT-GAN, employ adversarial training involving generators and discriminators. This approach helps them preserve realism and ensure visual consistency during image restoration. DeepFill v2, with the introduction of gated convolutions, significantly improves the handling of image details, while EdgeConnect's edge-first strategy effectively enhances the restoration of edges and regions with complex textures. These GAN-based methods perform excellently in metrics such as PSNR and SSIM, demonstrating their strong ability to generate visually realistic and detail-rich inpainted images.

Based on the data in the table and the discussion above, it is evident that GANs and their variants currently lead in image inpainting tasks, particularly when dealing with high-resolution and complex scenes. Future research can build upon this foundation by exploring the combination of GANs with other deep learning methods (such as RNNs and autoencoders) to further improve restoration performance. Additionally, as image resolution and complexity increase, optimizing computational

resource consumption and enhancing real-time inpainting capabilities will also become key research priorities in the future.

Table 1. Performance comparison of representative methods

Method	Datasets	PSNR↑	SSIM↑	MAE↓	FID↓
Context Encoders	Paris StreetView	17.59	/	17.59	/
IIZUKA et al.	Paris StreetView	24.37	0.81	/	/
Liu et al.	CelebA	30.10	0.93		
EdgeConnect	CelebA	30.02	0.93	1.38	1.08
EdgeConnect	Paris StreetView	28.26	0.89	1.91	2.49
DeepFill v1	Places2	25.92	0.89	/	/
DeepFill v2	Places2	29.52	0.92	/	/
Pyramid-VAE-GAN	MS-COCO	27.50	0.88	/	/
CASI	Paris StreetView	25.30	0.85	/	/
AOT-GAN	Places2	26.03	0.890	2.11	1.57
PMSCA-Res-U-Net	Places2	28.95	0.911	/	/
PMSCA-Res-U-Net	CelebA	32.42	0.952	/	/
LSTM	MS-COCO	26.00	0.86	/	/
Liu et al.	MS-COCO	25.40	0.84	/	/
Zhang et al.	CelebA	27.21	0.89	/	/

4. Discussion

Through an in-depth investigation of previous methods, there still exist several key challenges, despite the significant progress made in deep learning-based image inpainting techniques:

(1) Lack of naturalness and detail fidelity in inpainting results. Current image inpainting methods often struggle with complex scenes, leading to over-smoothing or unnatural textures. This issue arises primarily because models find it difficult to simultaneously grasp both the comprehensive architecture and specific details within an image, resulting in inpainted images that lack visual realism. Future research should focus on developing more refined model architectures that can simultaneously capture both global structure and local details. Further optimization of Generative Adversarial Networks (GANs) and attention mechanisms will be essential to improve the realism and detail fidelity of inpainted images.

(2) Problems with computational efficiency and processing speed. Existing image inpainting algorithms require substantial computational resources, especially when handling high-resolution images and large-scale datasets. This leads to slower inpainting processes, which limits the scalability of these techniques in real-world applications. The issue is particularly pronounced in real-time applications where speed is critical. Researchers can explore model compression techniques, acceleration algorithms, and multi-modal integration methods to improve robustness and efficiency.

(3) Dependency on large-scale labeled data. Deep learning models generally require vast amounts of labeled data for training. However, Procuring labeled data incurs significant costs and consumes substantial time, making it challenging to perform effective image inpainting in unsupervised or weakly supervised environments. Future research could focus on developing more

advanced unsupervised or weakly supervised learning methods. Additionally, researchers could explore ways to generate small labeled datasets that maximize the use of limited data resources.

(4) Security and trustworthiness concerns. As the application of image inpainting techniques expands into sensitive fields such as healthcare imaging and legal evidence analysis, the security and reliability of inpainting algorithms become critical. However, current models still have limitations in terms of interpretability and robustness, and they could potentially be misused for malicious image manipulation. To this end, it will be important to strengthen research on model interpretability and robustness to ensure the reliability of inpainting results in various application scenarios. Furthermore, in response to potential ethical issues related to image manipulation, future research should explore stricter algorithm design and application standards to guarantee the ethical application of these technologies.

5. Conclusion

This paper provides a comprehensive review of deep learning-based image inpainting techniques, covering a variety of methods. By analyzing the technical characteristics, application scenarios, and performance of these methods, we can observe the immense potential and advantages of deep learning in image inpainting tasks. GANs have steadily risen to prominence as the principal technology within the image inpainting sector, thanks to their outstanding performance in handling complex textures and maintaining global structural consistency. Meanwhile, Autoencoders, RNNs, and CNNs each have their own unique strengths in specific tasks, providing diverse solutions for image inpainting.

Although current image inpainting techniques have made significant progress in reconstructing image quality and ensuring detail fidelity, several challenges remain to be addressed. Balancing computational efficiency with further improvement in realism and detail quality of inpainted images remains a key focus for future research. Combining the strengths of multiple deep learning techniques, such as integrating GANs with Autoencoders, RNNs, and other methods, could potentially enhance the effectiveness of image inpainting. Additionally, advancements in unsupervised or weakly supervised learning approaches may reduce the dependency on large-scale labeled datasets, improving model generalization and practical application value.

In the future, as deep learning technology continues to advance and computational resources improve, image inpainting techniques are expected to achieve greater automation, stronger generalization capabilities, and faster processing speeds. These advancements will allow the widespread application of image inpainting in areas such as cultural heritage preservation, medical image restoration, and digital media creation, thereby driving these industries to higher levels of development.

References

- [1] D. Pathak, P. Krähenbühl, J. Donahue, T. Darrell and A. A. Efros, "Context Encoders: Feature Learning by Inpainting," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 2536-2544, doi: 10.1109/CVPR.2016.278.
- [2] Iizuka S, Simo-Serra E, Ishikawa H. Globally and locally consistent image completion. *ACM Transactions on Graphics*, 2017, 36 (4): 107. DOI:10.1145/3072959.3073659.
- [3] Liu G, Reda F A , Shih K J ,et al. Image Inpainting for Irregular Holes Using Partial Convolutions [J]. Springer, Cham, 2018. DOI:10.1007/978-3-030-01252-6_6.
- [4] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu and T. Huang, "Free-Form Image Inpainting With Gated Convolution," 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea (South), 2019, pp. 4470-4479, doi: 10.1109/ICCV.2019.00457.
- [5] Nazeri K, Ng E, Joseph T, et al. EdgeConnect: Generative image inpainting with adversarial edge learning. arXiv:1901.00212, 2019.

- [6] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu and T. S. Huang, "Generative Image Inpainting with Contextual Attention," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018, pp. 5505-5514, doi: 10.1109/CVPR.2018.00577.
- [7] ZHANG X, WANG X, SHI C, et al. De-gan: domain embedded gan for high quality face image inpainting [J]. *Pattern Recognition*, 2022, 124: 108415.
- [8] TIAN H, ZHANG L, LI S, et al. Pyramid-vae-gan: transferring hierarchical latent variables for image inpainting [J]. *Computational Visual Media*, 2023, 9 (4): 827- 841.
- [9] LI H, LI G, LIN L, et al. Context-aware semantic inpainting [J]. *IEEE transactions on cybernetics*, 2018, 49 (12): 4398-4411.
- [10] ZENG Y, FU J, CHAO H, et al. Aggregated contextual transformations for high-resolution image inpainting [J]. *IEEE Transactions on Visualization and Computer Graphics*, 2022.
- [11] CHEN Y, XIA R, YANG K, et al. Dnnam: image inpainting algorithm via deep neural networks and attention mechanism [J]. *Applied Soft Computing*, 2024, 154: 111392.
- [12] Xie et al. (2019) introduced RNNs into image inpainting and proposed a model that integrates Long Short-Term Memory (LSTM) networks, aiming to capture the relationships between the missing regions of an image and their surrounding known areas, resulting in better inpainting outputs with continuity .
- [13] Liu et al. (2020) proposed an image inpainting method based on Gated Recurrent Units (GRU). This method is noted for its simplicity and effectiveness in handling long-term dependencies, demonstrating good performance in image inpainting tasks involving large missing regions .
- [14] Zhang et al. (2021) expanded the application of RNNs in image inpainting by proposing a model based on bidirectional RNNs. This approach allows for a more comprehensive capture of contextual relationships by integrating information from both forward and backward information flows, enhancing the quality and consistency of the inpainting results .
- [15] Dong C, Loy C C, He K, et al. Image super-resolution using deepconvolutional networks [J]. *IEEE transactions on pattern analysis and machineintelligence*, 2015, 38 (2): 295-307.