Proceedings of CONF-SEML 2026 Symposium: Computational Analysis and Modeling in Complex Intelligent Systems
DOI: 10.54254/2755-2721/2026.GU34279

Progress and Analysis of Incomplete Image Restoration Based on CNN, GAN, Diffusion Models, and Hybrid Models

Yuheng Yang

Virtual Reality (VR) Modern Industrial College, Jiangxi University of Finance and Economics, Nanchang, China
TYB0511@outlook.com

Abstract. With the rapid development of machine learning and deep learning technologies, incomplete image restoration has become a research hotspot in the field of computer vision, with significant applications in cultural heritage preservation, medical imaging, surveillance completion, old photo restoration, and other areas. This paper focuses on incomplete image restoration technology, systematically reviewing four mainstream technical approaches based on Convolutional Neural Networks (CNN), Generative Adversarial Networks (GAN), diffusion models, and hybrid models. It analyzes the core principles, advantages and limitations, and experimental results of each technique, and, in combination with existing research, identifies current gaps in the field regarding restoration under extreme scenarios, cross-domain adaptation, and the balance between real-time performance and accuracy. The results indicate that each of the four techniques has its own focus, with hybrid architectures demonstrating greater potential for adapting to complex scenarios, while existing methods still face multiple bottlenecks that require breakthroughs. This study provides a solid reference and clear guidance for future research on incomplete image restoration, facilitating broader practical applications of this technology across various fields and promoting the deep integration of artificial intelligence technologies in specific areas of computer vision.

Keywords: Incomplete image restoration, CNN, GAN, Diffusion models, Hybrid architectures

1. Introduction

In the current era of rapid development in computer vision technology, images, as the core medium for information transmission, directly affect the accuracy and effectiveness of information interpretation. Image inpainting, as an important branch in the field of computer vision, aims to use technical means to fill in missing, damaged, or corrupted areas of images, restoring the complete information and visual coherence of the images. It holds irreplaceable application value in critical areas such as cultural relic preservation, medical image analysis, surveillance video completion, and old photo restoration. Traditional repair methods, such as interpolation and texture synthesis, overly rely on manual intervention, resulting in low repair accuracy and poor scene adaptability, making it difficult to handle complex damaged scenarios such as large-area damage, irregular masks, and ambiguous semantics. With the rapid integration of machine learning and deep learning

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.GU34279

technologies, image inpainting technology has made leapfrog progress, giving rise to mainstream technical approaches such as Convolutional Neural Networks (CNNs), diffusion models, Generative Adversarial Networks (GANs), and hybrid architectures, forming a pattern of parallel development of multiple technologies. The main aim of this paper is to systematically summarize the four major mainstream technical approaches for image inpainting, comprehensively analyze the core principles and advantages and disadvantages of each technology, and, based on existing research findings, clarify the current research gaps in the field, providing solid reference materials and clear guidance for future related research, promoting the development of this technology towards greater efficiency and accuracy.

2. Summary of mainstream image inpainting techniques

Current mainstream image inpainting techniques can be divided into three major categories. Each category is based on different core principles and has developed its own technical characteristics for different inpainting scenarios. Combining representative formulas, open-source data, and analysis, the summary is as follows:

2.1. CNN-based incomplete image restoration technology

CNN-based incomplete image restoration technology centers on convolutional neural networks. By extracting local features of images and using contextual information to fill in missing regions, it represents one of the earliest mature and widely applied technical approaches. Its core logic involves capturing local features such as textures and edges of images through multiple layers of convolution and pooling operations, and then restoring image resolution via deconvolution to complete the missing areas. The primary advantages of this technology are high computational efficiency and low hardware requirements, making it capable of quickly handling conventional incomplete scenarios (such as small-area damages and regular mask restorations).

A representative core formula of this technology is the partial convolution formula, which is crucial for solving irregular mask restorations, as follows:

$$\mathrm{P}(\mathrm{x}) = \frac{\sum_{\mathrm{k}}(\mathrm{w_k}\cdot\mathrm{x_k}\cdot\mathrm{m_k})}{\sum_{\mathrm{k}}\mathrm{m_k}+\epsilon}\cdot m + x\cdot(1\text{-}\mathrm{m}) \tag{1}$$

Here, x represents the input incomplete image feature map, m represents the mask matrix (with 0 for missing regions and 1 for intact regions), $\mathrm{w_k}$ represents the convolution kernel weight, ϵ is a small value to prevent a zero denominator, and $\mathrm{P}(\mathrm{x})$ is the output feature map of the partial convolution. The core advantage of this formula, as shown in Figure 1, is that the mask matrix m constrains the convolution process, using only pixels from the intact regions ($\mathrm{m_k}=1$) for convolution calculations. This avoids interference from invalid pixels on the restoration result while preserving the original information of intact regions, effectively improving the restoration accuracy for irregular holes [1].

DOI: 10.54254/2755-2721/2026.GU34279



Figure 1. Comparison of the performance of the PConv model on irregular mask inpainting tasks for facial images [1]

In the experiment based on Places365 and CelebA-HQ, the repair effect data of some convolution technology under irregular masks (mask rate 30%-70%): On the Places365 dataset, PSNR (peak signal-to-noise ratio, The higher the value, the higher the repair accuracy) reached 28.6 dB, and the SSIM (structural similarity, the closer to 1, the better the repair effect) reached 0.89, and on the CelebA-HQ dataset, the PSNR reached 32.1 dB and the SSIM reached 0.92 [1]. Compared with the traditional convolutional repair technology (Places365 dataset PSNR=25.3dB, SSIM=0.81), the repair accuracy of some convolutional techniques is significantly improved, especially in high mask rate (60%-70%) scenes, the repaired image semantics are more coherent and the texture is clearer.

2.2. Defective image restoration technology based on the diffusion model

The core principle is to gradually noize the image to random noise through the forward diffusion process, and then restore the complete image from the noise through the reverse diffusion process, and use the randomness and continuity of the diffusion process to achieve natural completion of the defective area. The core advantages of this type of technology are high repair quality, strong semantic consistency, and the ability to cope with extreme incompleteness scenarios (such as large-scale damage and semantically ambiguous area repair).

The core forward diffusion formula of the diffusion model is clearly proposed in the RePaint model, as follows:

$$q(x_t|x_{t-1})=N\left(x_t;\sqrt{1-\beta_t}x_{t-1}\beta_t I\right) \tag{2}$$

x_{t-1} is the image of step t-1, x_t is the image after the noise of step t, β_t is the noise coefficient of step t (the image gradually approaches random noise with the number of steps), N is the normal distribution, and I is the unit matrix. The forward diffusion process converts the complete image into random noise through gradual noise addition, providing the basis for reverse repair [2].

The reverse diffusion repair formula is:

$$p(x_{t-1}|x_t)=N\left(x_{t-1};\mu_\theta(x_t t)\sigma_t^2 I\right) \tag{3}$$

$\mu_\theta(x_t t)$ is the mean predicted by the model (output by the neural network θ), and σ_t^2 is the variance (fixed or predicted by the model), which gradually restores the complete image from the noise through the reverse denoising process, in which the information of the defective region is completed by the global semantic features learned by the model [2].

The experiment compares the performance of the two image restoration models, and the experiment shows that the DiffIR model is better than the RePaint model in terms of image

DOI: 10.54254/2755-2721/2026.GU34279

restoration quality and inference efficiency, as shown in Figure 2, which can not only generate restoration results that are closer to the real image and have more natural visual effects, but also have a significant lead in inference speed, taking into account the restoration effect and operation efficiency, and have more prominent practical application advantages.



Figure 2. Comparison of the effects of various image restoration models such as DiffIR/RePaint on face datasets [2]

2.3. GAN based mutilation image restoration method

Based on generative adversarial network (GAN) is a classic technical route in the field of deep learning restoration, juxtaposed with CNN and diffusion model-based methods, its core logic is to achieve high-fidelity and natural completion of defective areas through the adversarial training of generators and discriminators, and to achieve high-fidelity and natural completion of defective areas in scenes with rich textures and regular structures (such as faces). , natural landscape image restoration), and is still widely used in some practical scenarios [3, 4].

The core of the GAN repair method is the adversarial training mechanism, which learns the feature distribution of the complete image through the generator, generates contextually appropriate repair content, and distinguishes the real complete image from the generated repair image through the Discriminator. The core loss function of this method is composed of adversarial loss, perceptual loss, and L1 loss, which work synergistically to take into account the authenticity, semantic consistency, and pixel accuracy of the repair [5].

Core adversarial loss formula:

$$L_{ADV}(G,D)=E_{x,y}[\log D(x,y)]+E_{x,z}[\log(1-D(x,G(x,z)))] \tag{4}$$

Among them, x represents the masked mutilated image, y is the original complete image (Ground Truth), z is the random noise, $G(x,z)$ is the repaired image output by the generator, $D(x,y)$ represents the probability of the discriminator judging the real image, and $D(x,G(x,z))$ is the probability of the discriminator judging the generated image [3]. The core function is to make the repair content generated by the generator close to the real image texture, avoid blurring, semantic confusion and other issues, and make the repair results more visually coherent.

In order to compensate for the lack of semantic consistency of adversarial loss, the mainstream GAN repair model introduces perceived loss:

$$L_{PERC}(G)=E_{x,y}[\|\phi(y)-\phi(G(x,z))\|_1] \tag{5}$$

Where $\phi(\cdot)$ is a pre-trained VGG16 feature extractor, which improves the semantic consistency of the repaired content by comparing the high-level feature differences between the real image and the repaired image, and solves the texture distortion problem that GAN models are prone to [4].

The final total loss function is:

DOI: 10.54254/2755-2721/2026.GU34279

$$L_{TOTAL}=\lambda_{ADV}L_{ADV}+\lambda_{PERC}L_{PERC}+\lambda_{L1}\|y-G(x,z)\|_1 \quad (6)$$

λ_{ADV}、λ_{PERC} and λ_{L1} balance the weights of each loss to achieve synergistic optimization of authenticity, semantic consistency, and pixel accuracy [5].

2.4. Defective image restoration technology based on hybrid architecture

The core of the incomplete image restoration technology based on hybrid architecture is to integrate the advantages of CNN, Transformer, diffusion model and other technologies to make up for the limitations of a single technology. The core idea of this type of technology is to use CNN to capture local features, Transformer to model global semantic relationships, and diffusion models to improve the naturalness of repair, which is suitable for complex defective scenarios (such as large-scale damage in high-resolution images and superposition of multiple types of defects) [6].

3. The current research gap in the field of mutilated image restoration

Combined with the research status of the above mainstream technologies, open source experimental data and practical application requirements, the current research gaps in the field of mutilated image restoration mainly focus on the following three aspects:

3.1. The ability to repair extreme scenes is insufficient

From the open-source experimental data, it can be seen that the repair accuracy of the existing mainstream models is significantly reduced in the scenario of ultra-large area damage (mask rate exceeding 80%): the PSNR of the LaMa model drops to 24.7 dB and the SSIM to 0.78 on the FFHQ dataset (mask rate of 80%) [7]. The PSNR and SSIM of the RePaint model dropped to 25.3 dB and SSIM to 0.76 [2] on the CelebA-HQ dataset (mask rate of 80%), both of which had obvious semantic confusion and texture blur. For low-resolution defective images (such as below 256×256), existing models are prone to aliasing and artifacts after repair, and relevant open source data show that the PSNR of low-resolution image restoration is generally 3-5dB lower than that of high-resolution images.

3.2. Poor cross-field adaptability

Existing repair models are mostly designed for single-domain datasets and lack general cross-domain repair capabilities. For example, the RePaint model, suitable for natural image restoration, has a PSNR of only 23.1 dB and an SSIM of 0.72 on the open-source medical imaging dataset BraTS (CT images of brain tumors), which is much lower than its restoration effect on natural images [2]. In the LaMa model suitable for cultural relics restoration, on the open source monitoring dataset UCF101 (dynamic monitoring image), the restored image has obvious dynamic blur, and the SSIM is only 0.75 [7]. There is an obvious gap in the research on the generalization ability of the cross-domain model. The root cause of this problem lies in the lack of effective cross-domain knowledge transfer capabilities of models, which is one of the core research directions for improving the generalization ability of artificial intelligence [8, 9].

DOI: 10.54254/2755-2721/2026.GU34279

3.3. The challenge of balancing real-time performance and accuracy has not yet been resolved

From open-source data, it can be observed that diffusion model-based restoration techniques offer high accuracy but are slow in computation; the RePaint model's inference time exceeds 600 ms, making it difficult to meet real-time restoration requirements (such as real-time monitoring completion or mobile device restoration) [2]. On the other hand, CNN-based restoration techniques are fast in computation (inference time is usually below 100 ms) but have limited accuracy, making it hard to handle complex defect scenarios [1, 10]. Hybrid architecture techniques attempt to take advantage of both, but they still face issues of model complexity and high computational cost. How to significantly improve computational speed while ensuring restoration accuracy, achieving real-time restoration, remains an important research gap in the current field [11].

4. Conclusion

This paper, based on the analysis of CNNs, diffusion models, GANs, and hybrid models, clarifies the core advantages and application scenarios of each technology, highlights the current research gaps in areas such as restoration in extreme scenarios, cross-domain adaptation, real-time performance balancing, and unsupervised techniques, and, based on technological development trends and practical application needs, provides a reasoned outlook on future research directions. Currently, image inpainting technology has made significant progress but still faces many technical bottlenecks, which are both unique issues in this field and common challenges in the broader field of artificial intelligence. Future research will focus on addressing existing gaps, promoting technology deployment, achieving coordinated improvements in restoration accuracy, efficiency, and generalization capability, supporting the widespread application of this technology in fields such as cultural heritage protection, medical imaging, and monitoring completion, and providing references for the deep integration of artificial intelligence technologies in specialized domains.

References

[1] Liu, G., Reda, F. A., Shih, K. J., et al. (2018). Image inpainting for irregular holes using partial convolutions. In Proceedings of the European Conference on Computer Vision (ECCV) (pp. 85–100).
[2] Lugmayr, A., Danelljan, M., Timofte, R., & Van Gool, L. (2022). RePaint: Inpainting using denoising diffusion probabilistic models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 11461–11471).
[3] Pathak, D., Krähnnbühl, P., Donahue, J., Darrell, T., & Efros, A. A. (2016). Context encoders: Feature learning by inpainting. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 2536–2544).
[4] Iizuka, S., Simoes, R., & Ishikawa, H. (2017). Globally and locally consistent image completion. ACM Transactions on Graphics, 36(4), 108.
[5] Yu, J., Lin, Z., Yang, J., et al. (2018). Generative image inpainting with contextual attention. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 5505–5514).
[6] Liu, H., Wang, Y., Qian, B., Wang, M., & Rui, Y. (2024). Structure matters: Tackling the semantic discrepancy in diffusion models for image inpainting. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 8038–8047).
[7] Suvorov, R., Logacheva, E., Mashikhin, A., et al. (2022). Resolution-robust large mask inpainting with Fourier convolutions. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 2149–2159).
[8] Li, X., Li, J., Zhu, L., & Shen, H. T. (2024). Efficient transfer learning of large models with limited resources: A survey. Chinese Journal of Computers, 47(11), 2491–2521.
[9] Serrano, S. A., Martinez-Carranza, J., & Sucar, L. E. (2024). Knowledge transfer for cross-domain reinforcement learning: A systematic review. IEEE Access, 12, 114552–114572.

DOI: 10.54254/2755-2721/2026.GU34279

[10] Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X., & Huang, T. S. (2021). Free-form image inpainting with gated convolution. IEEE Transactions on Pattern Analysis and Machine Intelligence, 43(11), 3851–3865.
[11] Chen, W., Wang, J., Zhang, Y., et al. (2025). On-device large language models: A survey of model compression and system optimization. Research Square. https: //doi.org/10.21203/rs.3.rs-7975734/v1