Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 32 No. 8s (2025)

Advances in Deep Learning Image Inpainting Architectures for the Effective Reconstruction of Damaged Regions- A Systematic Review

***Suma N[1], Dr. Kusuma Kumari B.M[2]**
[1]Research Scholar, Department of Studies and Research in Computer Applications, Jnanasiri Campus, Tumkur University, Bidrakatte,Tumkuru, Karnataka, India
[2]Assistant Professor, Department of Studies and Research in Computer Applications, Jnanasiri Campus, Tumkur University, Bidrakatte, Tumkuru, Karnataka, India
Corresponding Author : Suma N

Article History:

Received: *06-11-2024*

Revised: *13-12-2024*

Accepted: *02-01-2025*

Abstract:

Image inpainting is the method of reconstructing the missed or damaged region of the picture with certain rules that play a vital role in computer vision applications. Numerous image inpainting approaches have been recently proposed that effectively reconstruct the missed regions. Traditionally the missed regions are restored by schemes that are derived from the diffusion models, exemplar approaches, and sparsity approaches. Different hybrid schemes are derived from these traditional schemes that show reasonable performance in missed region reconstruction. Recently deep learning-based approaches which are derived from the architectures of Generative adversarial networks (GAN), Convolutional neural networks (CNN), Transformers, and U-Net are commonly used for effective image painting. More specifically numerous approaches have been derived using the advantage of GAN and CNN architectures. These schemes focus on the reconstruction of global structures and local texture components of the missed region from the known region. Therefore, the paper provides a review of different traditional, and recent deep-learning schemes.

Keywords: Image Inpainting, Exemplar approach, GAN approach, Convolutional neural network, GAN model.

1. Introduction

The art of restoring damaged photographs or paintings that have minor damages namely spots, dusts, cracks, and scratches is termed to be image inpainting [1]. The image inpainting procedure improves the visual appearance of the images. The digital image inpainting usually interpolates the neighborhood of the missed region to restore the missed area. i.e. the known picture information is utilized to estimate the damaged region. Image inpainting [2] has different applications namely old photo coloring, scratch removal, text removal, watermark removal, object/undesired region removal, etc. The different applications of image inpainting are illustrated in Fig. 1. Such applications can be broadly classified as restoration of damaged regions, Object removal, image completion, and artistic applications [3].

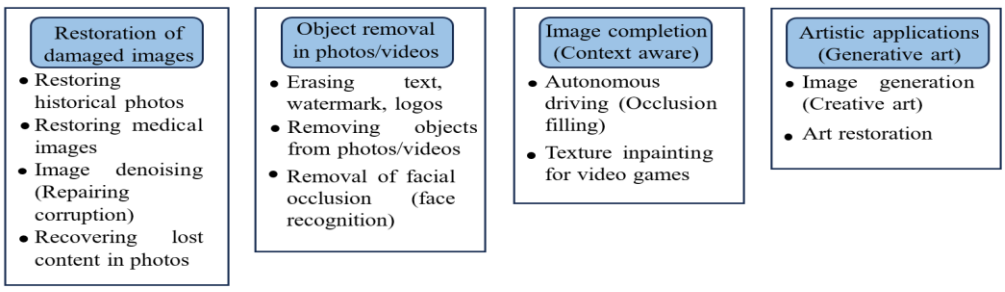

Fig. 1: Practical use of image inpainting algorithms

The image inpainting process should preserve the semantic structure while synthesizing the damaged or unknown region. The image inpainting also improves the resolution and helps to recover the images that are damaged due to storage, and transmission. The two broad classes of image inpainting schemes are the deep learning approach and the traditional approach as illustrated in Fig. 2. The traditional schemes include diffusion-based schemes [4], exemplar-based schemes [5], and sparsity-based schemes [6].

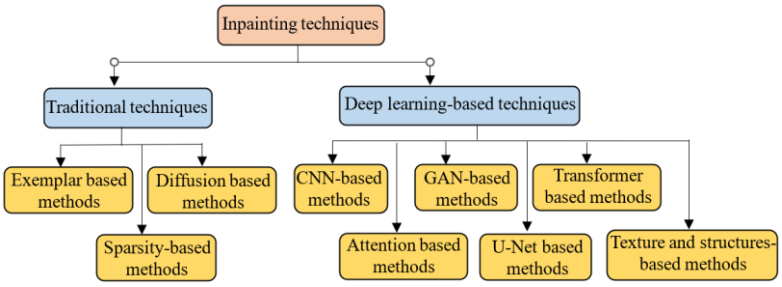

Fig. 2: Classification of various image inpainting schemes

Different deep learning architectures are derived from the CNN [7], GAN [8], Transformer [9], and U-Net structures [10]. There are two challenges in image inpainting application from a practical perspective. (i) Restoring the missed or damaged region from the known region to obtain the complete images. Such challenges are present in the reconstruction of old photographs. (ii) Removal of undesirable content from the picture, and replacing the region with plausible content. Such a challenge occurs mostly in photo editing for removing the objects and the watermark content. Also, the behavior of the picture inpainting scheme depends on the mask ratio.

The paper has the following forthcoming sections. Section 2 deliberates the outperforming traditional inpainting schemes and section 3 discusses the works that are related to deep learning. Section 4 discusses the dataset and important evaluation measures used for the analysis of image inpainting schemes. Section 5 shows the performance comparison of different approaches and lastly, the conclusion of the paper is illustrated in Section 6.

2. Traditional Schemes

The traditional image inpainting schemes include the approaches that are derived from the diffusion model, exemplar approach, and sparsity approach. This approach uses mathematical rules to derive the missed pixels.

2.1 Diffusion model-based approach

Diffusion models [11] are utilized in diffusion-based image inpainting. The two processes involved in diffusion-based inpainting are reverse and forward processes. The forward procedure adds the noise in multiple steps so that the image becomes completely noisy. The Stochastic-differential equation is used to represent the forward process. The noise added an image at iteration i can be represented as,

$$I_i = \sqrt{1 + \gamma_i}\rho_i + \sqrt{\gamma_i} I_o \quad (1)$$

Here, I_o resembles the input image, γ_i controls the noise over iteration i, ρ_i represents the noise that is added at iteration i. The reverse noise process removes the noise to reconstruct the original image. This was also performed iteratively. The two regions used in image inpainting include the unmasked region and the masked region. The visible part of the image is termed be unmasked region. The unmasked region helps to identify the mask area content. The mask area resembles the picture region, where the image pixels are corrupted, missed, or intentionally modified. It also may be due to

an occlusion (unwanted objects). The model learns to recover the visual content from the noisy image on the unmasked region. The same recovery procedure is followed in the masked region to complete the image inpainting process. The forward process depends on parameters such as the number of diffusion steps, noise type, and noise schedule. The reverse process depends on the model, noise, predictor, and conditioning mechanism. Diffusion schemes namely Denoising diffusion implicit models (DDIM) [12], and denoising diffusion probabilistic models (DDPM) [13] are commonly used diffusion models.

2.2 Exemplar-based schemes

The exemplar approach [14] aims to minimize the energy function

$$P = \alpha P_d(I) + \beta P_s(I) \quad (2)$$

$P_d(I)$ resembles how the inpainted image agrees with the known pixels and $P_s(I)$ resemble the continuity between the inpainted, and original region. α, and β resembles the weight to control the data term, and smoothness term respectively. The exemplar approach uses the principle of borrowing the information (known as exemplars) from the content available region to estimate the content of the masked area. This approach uses a patch-based reconstruction [15] instead of using pixel-wise interpolation. The main process involved in this exemplar approach is selecting the patch from the known region (unmasked region). The selected patch should closely match the neighborhood of the masked region. The matching is done based on criteria such as contextual consistency, texture similarity, or color similarity [16]. During the patch transfer, the patches are aligned at the boundaries. To have a smooth transition between the patched region, and the boundaries, the region is blended [17]. Usually, the masked regions are filled with patches step by step. The performance of the exemplar-based approach depends on parameters such as search window size, patch overlap, patch size, boundary propagation strategy, matching criteria, inpainting iterations, patch matching distance, and patch search strategy. The Greedy search algorithm [18] is commonly used to search the patches.

2.3 Sparsity-based schemes

This approach uses the sparsity concept to estimate the missed region. The principle of this approach is that a natural image can be compressed to a few coefficients. These coefficients act as a dictionary element or basis function. Fourier, wavelet, and DCT transform are utilized to estimate the coefficients. The performance of this approach depends on the transform type, stopping criterion, optimization type, and sparsity regularization parameter λ. The reconstructed image in sparse based approach can be represented as,

$$I_r = \underset{c}{argmin}\{\|Q(C - Cp)\|_2^2 + \lambda\|Zc\ \|_1\} \quad (3)$$

Q represents the projection operator. C_p and λ represents the partially observed image, and regularization factor respectively. $\|Zc\ \|_1$ resembles the $l_i - norm$ of the sparse coefficients, and C resembles the input image. The sparsity approach [19] uses controllable generation and content initialization to perform the image inpainting. A multi-pixel window mechanism was also used [20] to capture more fine details.

3. Deep learning-based image inpainting

Several researchers developed image inpainting schemes that are derived from, CNN, GAN, Transformer, U-Net, Attention, and texture-structure models.

3.1 CNN-based schemes

The CNN-based approach [21] usually has an encoder, and decoder unit where the encoder structure collects the high-level descriptors from the visible region sequence of convolutional filters that increase the feature maps. A bottleneck layer is used after the encoder which shows the compressed representation of the feature map. Quaternion CNN [22] was proposed by Miao et al. that overcome the two important challenges in quaternion matrix estimation. Firstly the performance of image inpainting for the quaternion matrix approach highly depends on the type of regularizer. A regularizer won't show a better performance for all-natural images. Secondly, due to the non-commutativity property in multiplication for the quaternion matrix, optimizing the quaternion approach is difficult. The two challenges are overcome by the use of untrained quaternion CNN [23]. The loss function in the CNN-based approach can be expressed as,

$$\eta(I_r, I_i) = \lambda \eta_r(I_r) + \eta_d(I_r, I_i) \quad (4)$$

$\eta_r(I_r)$ resembles the regularization term, whose weight is controlled by the hyperparameter λ. $\eta_d(I_r, I_i)$ resembles the data loss between the inpainted picture I_r, and ground truth picture I_i.

An adaptive fusion process [24] was introduced by Zhu et al. that transfers visual style descriptors of known regions to unknown regions. The transformer-based approach provides a better global correlation, while the CNN-based approach can effectively reconstruct the local patterns. The global, and local advantages of the transformer, and CNN are combined to derive a local-global mixture [25]. Inception networks [26], [27], [28] architectures, ResNet [29], and VGG [30], structures are derived from the traditional CNN approach.

3.2 GAN-based schemes

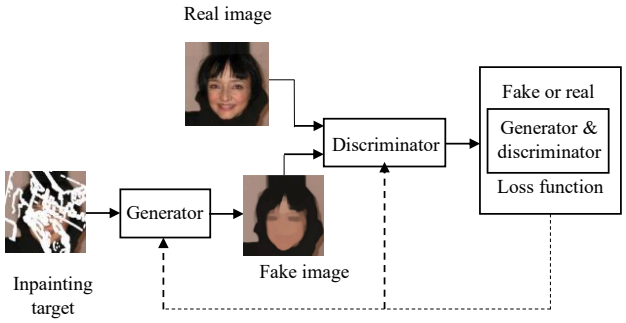

Fig. 3: GAN-based architecture in image inpainting

The two major components to perform image inpainting [31] in GAN architecture are the Generator module and Discriminator module as depicted in Fig. 3. The total objective function in GAN is

$$\eta_t = \lambda \eta_{pix} + \eta_G \quad (5)$$

Here, η_{pix} represents the pixel-wise loss, and η_G represents the generator loss.

The author Li et al. [32] analyzed the inability of the traditional GAN in reconstructing the texture, and structures of the missed region simultaneously. This approach introduces a texture, and structure-guided model, that uses the subnetworks such as refinement network, texture, and structure reconstruction process. The texture, and structure reconstruction process can restore more fine texture, and coherent structures than the traditional GAN. Finally, the refinement network combines the structure, and texture to obtain the inpainted picture. To reduce the computational burden of the GAN structure, the generator unit was included with a multi-task learning process [33]. This approach was used in inpainting the biomedical images in which the encoder-decoder performs three

processes namely detecting the edges, generation of organ boundary, and completion of resultant images. The structure of the GAN uses regularization functions and an optimum number of layers.

A dual-stage GAN [34] was derived where the first and second stage uses structure-aware and texture-aware learning mechanisms. The structure learning recovers the low-frequency component, while the texture-aware learning recovers the high-frequency component. The performance of GAN was enhanced by models namely GAN-Gradient Penalty [36], least squares GAN [35], and Wasserstein GAN [37] architectures.

3.3 Transformer based schmes

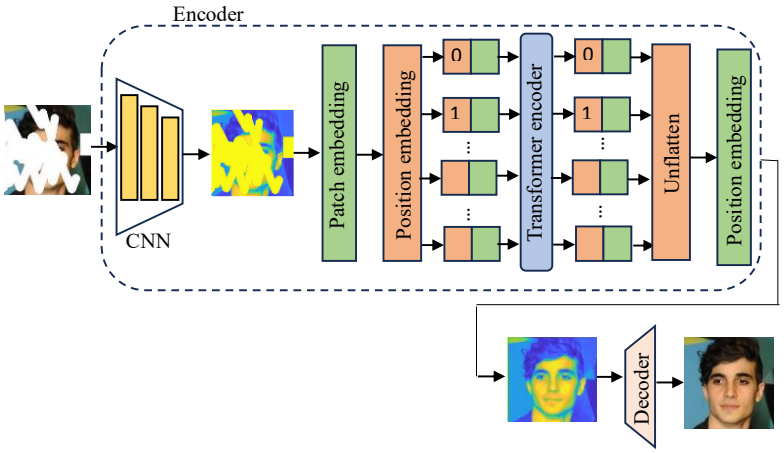

Fig. 4: Representation of transformer model in image inpainting

Fig. 4 shows the structure of the transformer model-based image inpainting. The total loss in transformer-based schemes is

$$\eta_t = \frac{\lambda_{per}}{N}\sum_{n=1}^{N}\|\delta_n(I_i) - \delta_n(I_r)\|_2^2 + \frac{1}{|\rho|}\sum_{j\in\rho}\|I_r(j) - I_i(j)\|_2^2 \qquad (6)$$

ρ resembles the indices of masked pixels, $\delta_n(.)$ resembles the features extracted at the n^{th} layer. N resembles the number of layers and λ_{per} resembles the constant. The transformer model [38] used by Huang et al. uses a U-Net structured transformer named self-attention transformer. This transformer reduces the complexity of the multi-headed attention structure. This approach replaces the traditional attention estimation. The irrelevant features are excluded by the sparse attention map that uses the ReLU function instead of the canonical function. Transformer-based schemes such as the ZITS network [39], MAE model [40], and transfill model [41] also provide reasonable performance.

The performance of the transformer reduces if the missed region is large. To address this challenge, a visual transformer structure [42] is introduced that uses a variable hyperparameter. This approach initially obtains multi-scale patches by partitioning the descriptor maps. The computation burden of the self-attention process is balanced by assigning the feature map to different heads. This approach also minimizes the number of feature map channels by a depth-wise convolution strategy. A dual discriminator approach [43] was used to improve the local and global discrimination in GAN. Also, the texture-aware module in GAN architecture [44] reduces the complexity of the traditional GAN structure.

3.4 Attention-based schemes

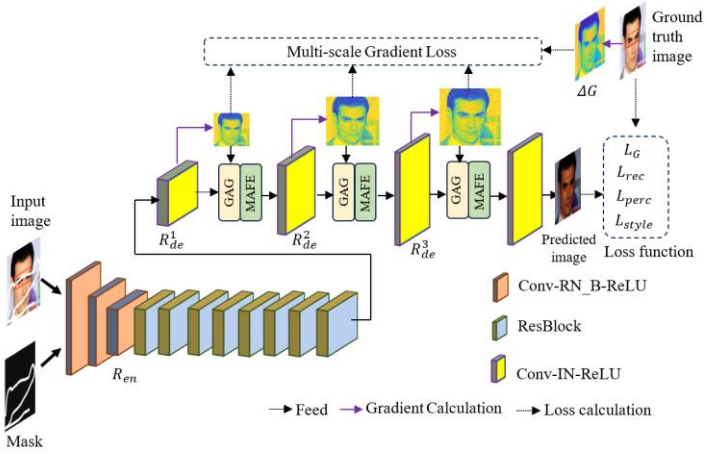

Fig. 5: Representation of attention-based image painting

The attention-based approach provides a high preference for certain regions while reconstructing the missed region. Fig. 5 illustrates the representation of attention-based image painting [24] which has the single encoder section with Convolutional filters and ResBlocks represented as (R_{en}), while the decoder has three sections of convolutional filters and attention modules represented as $(R^1_{de}, R^2_{de}, R^3_{de})$. Two attention modules namely multiscale attention-based feature extraction (MAFE) and gradient attention guidance (GAG) are utilized in the three sections of the decoder. Four different loss functions namely style loss (L_{style}), perception loss (L_{perc}), reconstruction loss (L_{rec}) and gradient loss (L_G) are utilized to update the multiscale gradient loss (ΔG) of the decoder. Different attention approaches namely coherent semantic attention [46], contextual attention [45], and multistage attention [47] provide fine texture information during the recovery of missed regions.

3.5 U-Net based schemes

The structure of the U-Net and Encoder-Decoder structures are represented in Fig. 6. Deep networks and attention processes are combined [48] to improve the texture distortion during the reconstruction process. The generator was constructed using the Res-U-Net structure. The residual structure was used in the encoder of the U-Net-based backbone. The skip connection uses the attention module which improves the effectiveness of the inpainting process.

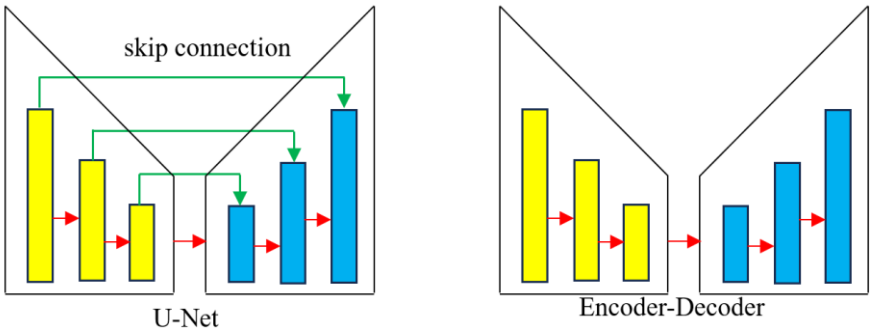

Fig. 6: Difference between U-Net architecture and Encoder-Decoder structures

The loss function in U-Net-based image inpainting can be represented as,

$$\eta(I_r, I_i) = \lambda_t \eta_t(I_r, I_i) + \lambda_p \eta_p(I_r, I_i) \quad (7)$$

Here, η_p resembles the perceptual loss represented as,

$$\eta_p(I_r, I_i) = \|\delta(I_r) - \delta(I_i)\|_2^2 \quad (8)$$

Here, $\delta(\cdot)$ represents the descriptor extraction function. I_r and I_i resembles the reconstructed image and input image respectively. In equation (7), the total variation loss function η_t can be denoted by,

$$\eta_t(I_r) = \sum_{u,v}\{(|I_r(u+1,v) - I_r(u,v)|) + (|I_r(u+1,v) - I_r(u,v)|)\} \quad (9)$$

λ_p, and λ_l are hyperparameters to control the perceptual, and total variation loss respectively. Different networks are derived from the U-Net structure that involves Pyramid-Network [49], partial convolution [50], deep fusion network [51], and Shift-Net [52]. The block attention process was included with the ResNet architecture to derive the improved inpainting architecture [53] that shows a lesser variation between the boundaries of the mask region and the un-mask region. The self-attention-based symmetric connected U-Net [54] estimates fine texture on the missed regions.

3.6 Texture and structure based schemes

In this scheme, the texture of the unmasked area is utilized to fill the masked area. The texture pattern of the masked area is synthesized based on the structure, and texture of neighboring unmasked pixels. Wang et al. [55] used a small receptive field-based shallow, and deep structure to minimize the formation of artifacts during reconstruction and to perform local refinement. The global refinement was performed by a large receptive field-based attention process.

Liu et al. [56] combined the pros of the transformer model, and CNN model to derive a bidirectional stream network. The CNN network is utilized to collect more local information, while the transformer is utilized to collect relevant features throughout the image. The CNN, and transformer model are developed on a hierarchical encoder and decoder architecture that minimizes the complexity. The degradation of GAN performance for large missed regions is addressed by the authors Chen et al. [57] by proposing a backbone network. The backbone network uses a non-pooling structure to reconstruct the texture descriptors. Thus the network has an image refinement network that follows the texture reconstruction and structural reconstruction network as illustrated in Fig. 7.

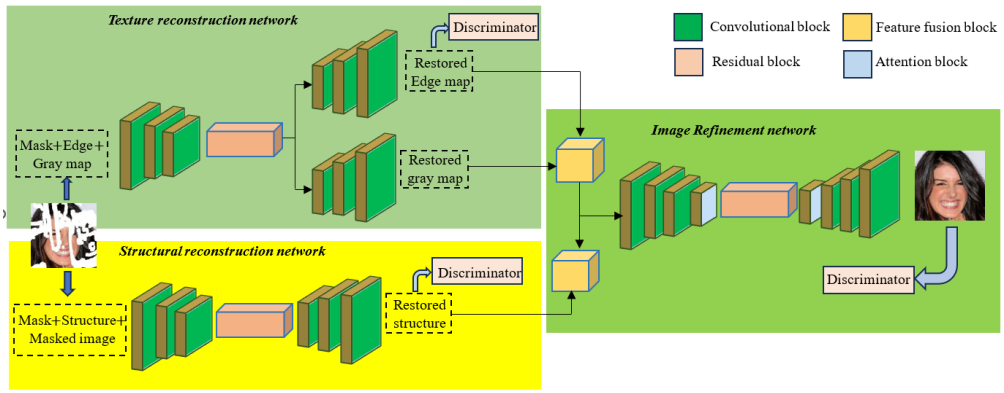

Fig. 7: Structure of the texture and structural-based approach

Multi-stage inpainting schemes are also commonly used in recent years. Usually, the number of refined networks is improved to enhance the inpainting performance. But increasing the refined networks also degrades the image due to the presence of artifacts, and blurring effects. To address this issue a perspective field structure [58] is introduced that uses a U-network with a multi-head attention structure. Five receptive field sub-networks are used in decreasing order of receptive field perspective. The influence of local pixels in the convolution process is minimized by the usage of the TransConv structure. An adaptive feedback process is utilized to derive an adaptive feedback

network [59]. This network have the ability to derive the inpainting result, and uncertainty map simultaneously. As the iteration proceeds, the uncertainty is gradually reduced by the feedback network, which also improves the inpainting result.

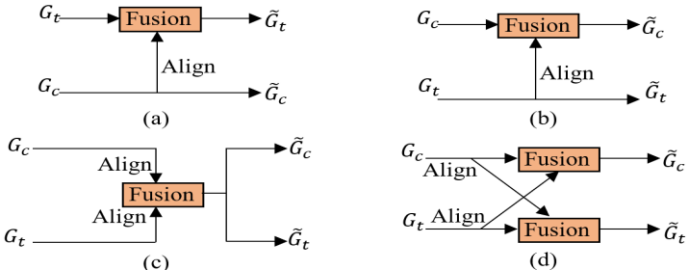

Fig. 8: Feature fusion approaches (a) Fusion at transformer stream (b) Fusion at CNN stream (c) Unified fusion (d) Fusion in both transformer and CNN stream

The features obtained by individual networks are combined by the fusion process. i.e. the texture features obtained by one network and the structural feature obtained by another network are combined by one of the fusion processes illustrated in Fig. 8. Here four feature fusion approaches are utilized to merge the transformer features (G_t) and the CNN feature (G_c).

4. Conclusion

This paper delivers a systematic review of different traditional and deep learning schemes used for picture inpainting applications. The paper discusses the uniqueness of the different research works that are derived from exemplar, diffusion, and sparsity-based schemes. The work also reviews the various deep learning structures that are used to perform image inpainting that effectively reconstructs the texture and structure of the missed area. More specifically the paper also focuses on the hybrid architectures that combine the advantage of CNN and GAN architecture in the reconstruction of local texture and global structure. The paper also reviews the architectures that are derived from the attention approaches, U-Net structures, and transformer-based approaches that show better performance for higher mask ratios.

References

[1]. Bugeau, A., Bertalmío, M., Caselles, V., & Sapiro, G. (2010). A comprehensive framework for image inpainting. *IEEE transactions on image processing*, *19*(10), 2634-2645.
[2] Guillemot, C., & Le Meur, O. (2013). Image inpainting: Overview and recent advances. *IEEE signal processing magazine*, *31*(1), 127-144.
[3]. Xu, Z., Zhang, X., Chen, W., Yao, M., Liu, J., Xu, T., & Wang, Z. (2023). A review of image inpainting methods based on deep learning. *Applied Sciences*, *13*(20), 11189.
[4]. Corneanu, Ciprian, Raghudeep Gadde, and Aleix M. Martinez. "Latentpaint: Image inpainting in latent space with diffusion models." *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 2024.
[5]. Arias, P., Facciolo, G., Caselles, V., & Sapiro, G. (2011). A variational framework for exemplar-based image inpainting. *International journal of computer vision*, *93*, 319-347.
[6]. Shen, B., Hu, W., Zhang, Y., & Zhang, Y. J. (2009, April). Image inpainting via sparse representation. In *2009 IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 697-700). IEEE.
[7]. Li, Z., & Wu, J. (2019). Learning deep CNN denoiser priors for depth image inpainting. *Applied Sciences*, *9*(6), 1103.
[8]. Zhang, X., Wang, X., Shi, C., Yan, Z., Li, X., Kong, B., ... & Mumtaz, I. (2022). De-gan: Domain embedded gan for high quality face image inpainting. *Pattern Recognition*, *124*, 108415.
[9]. Ko, K., & Kim, C. S. (2023). Continuously masked transformer for image inpainting. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 13169-13178).
[10]. Wei, R., & Wu, Y. (2022). Image Inpainting via Context Discriminator and U-Net. *Mathematical Problems in Engineering*, *2022*(1), 7328045.
[11]. Sridevi, G., & Srinivas Kumar, S. (2019). Image inpainting based on fractional-order nonlinear diffusion for image reconstruction. *Circuits, Systems, and Signal Processing*, *38*, 3802-3817.

[12]. J. Song, C. Meng, and S. Ermon, "Denoising diffusion implicit models," 2020, arXiv:2010.02502.
[13]. J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in Proc. Adv. Neural Inf. Process. Syst., vol. 33, 2020, pp. 6840–6851.
[14]. Zhang, N., Ji, H., Liu, L., & Wang, G. (2019). Exemplar-based image inpainting using angle-aware patch matching. *EURASIP Journal on Image and Video Processing*, *2019*, 1-13.
[15]. Xu, Z., & Sun, J. (2010). Image inpainting by patch propagation using patch sparsity. *IEEE transactions on image processing*, *19*(5), 1153-1165.
[16]. Yi, Z., Tang, Q., Azizi, S., Jang, D., & Xu, Z. (2020). Contextual residual aggregation for ultra high-resolution image inpainting. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 7508-7517).
[17]. Prados, R., Garcia, R., & Neumann, L. (2014). *Image blending techniques and their application in underwater mosaicing* (Vol. 13). Cham, Switzerland: Springer.
[18]. Ghrabat, M. J. J., Ma, G., Abduljabbar, Z. A., Al Sibahee, M. A., & Jassim, S. J. (2019). Greedy learning of deep Boltzmann machine (GDBM)'s variance and search algorithm for efficient image retrieval. *IEEE Access*, *7*, 169142-169159.
[19]. Li, Fan, et al. "MagicEraser: Erasing Any Objects via Semantics-Aware Control." *European Conference on Computer Vision*. Springer, Cham, 2025.
[20]. Han, Ruyi, et al. "Transformed sparsity-boosted low-rank model for image inpainting with non-convex γ-norm regularization and non-local prior." *Optics & Laser Technology* 181 (2025): 111865.
[21]. Li, Z., & Wu, J. (2019). Learning deep CNN denoiser priors for depth image inpainting. *Applied Sciences*, *9*(6), 1103.
[22]. Miao, J., Kou, K. I., Yang, Y., Yang, L., & Han, J. (2024). Quaternion matrix completion using untrained quaternion convolutional neural network for color image inpainting. *Signal Processing*, *221*, 109504.
[23]. Jia, Z., Ng, M. K., & Song, G. J. (2019). Robust quaternion matrix completion with applications to image inpainting. *Numerical Linear Algebra with Applications*, *26*(4), e2245.
[24]. Zhu, Y., Wang, C., Geng, S., Yu, Y., & Hao, X. (2022). Multi-scale gradient attention guidance and adaptive style fusion for image inpainting. *Journal of Visual Communication and Image Representation*, *89*, 103681.
[25]. Woo, S., Ko, K., & Kim, C. S. (2024). Local and global mixture network for image inpainting. *Journal of Visual Communication and Image Representation*, *104*, 104312.
[26]. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2015, pp. 1–9.
[27]. C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 2818–2826.
[28]. C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-ResNet and the impact of residual connections on learning," in Proc. AAAI Conf. Artif. Intell., 2017, vol. 31, no. 1, pp. 1–7.b
[29]. O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet large scale visual recognition challenge," Int. J. Comput. Vis., vol. 115, no. 3, pp. 211–252, Dec. 2015.
[30]. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, arXiv:1409.1556.
[31]. Wang, W., Niu, L., Zhang, J., Yang, X., & Zhang, L. (2022). Dual-path image inpainting with auxiliary gan inversion. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 11421-11430).
[32]. Li, Z., Zhang, Y., Du, Y., Wang, X., Wen, C., Zhang, Y., ... & Jia, F. (2024). STNet: Structure and texture-guided network for image inpainting. *Pattern Recognition*, *156*, 110786.
[33]. Rakibe, P. L., & Patil, P. D. (2024). Improved medical image inpainting using automatic multi-task learning driven deep learning approach. *e-Prime-Advances in Electrical Engineering, Electronics and Energy*, *9*, 100678.
[34]. Yeh, C. H., Yang, H. F., Chen, M. J., & Kang, L. W. (2024). Image inpainting based on GAN-driven structure-and texture-aware learning with application to object removal. *Applied Soft Computing*, *161*, 111748.
[35]. X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S. P. Smolley, "Least squares generative adversarial networks," in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Oct. 2017, pp. 2813–2821.
[36]. I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of Wasserstein GANs," in Proc. Adv. Neural Inf. Process. Syst., vol. 30, 2017, pp. 1–11.
[37]. M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in Proc. Int. Conf. Mach. Learn., 2017, pp. 214–223.
[38]. Huang, W., Deng, Y., Hui, S., Wu, Y., Zhou, S., & Wang, J. (2024). Sparse self-attention transformer for image inpainting. *Pattern Recognition*, *145*, 109897.

[39]. Q. Dong, C. Cao, and Y. Fu, "Incremental transformer structure enhanced image inpainting with masking positional encoding," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2022, pp. 11348–11358.
[40]. K. He, X. Chen, S. Xie, Y. Li, P. Dollár, and R. Girshick, "Masked autoencoders are scalable vision learners," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2022, pp. 15979–15988.
[41]. Y. Zhou, C. Barnes, E. Shechtman, and S. Amirghodsi, "TransFill: Reference-guided image inpainting by merging multiple color and spatial transformations," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2021, pp. 2266–2267.
[42]. Campana, J. L. F., Decker, L. G. L., e Souza, M. R., de Almeida Maia, H., & Pedrini, H. (2023). Variable-hyperparameter visual transformer for efficient image inpainting. *Computers & Graphics*, *113*, 57-68.
[43]. Chen, Y., Zhang, H., Liu, L., Chen, X., Zhang, Q., Yang, K., ... & Xie, J. (2021). Research on image inpainting algorithm of improved GAN based on two-discriminations networks. *Applied Intelligence*, *51*, 3460-3474.
[44]. Hedjazi, M. A., & Genc, Y. (2021). Efficient texture-aware multi-GAN for image inpainting. *Knowledge-Based Systems*, *217*, 106789.
[45]. Liu, J., Gong, M., Tang, Z., Qin, A. K., Li, H., & Jiang, F. (2022). Deep image inpainting with enhanced normalization and contextual attention. *IEEE Transactions on Circuits and Systems for Video Technology*, *32*(10), 6599-6614.
[46]. Liu, H., Jiang, B., Xiao, Y., & Yang, C. (2019). Coherent semantic attention for image inpainting. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 4170-4179).
[47]. Li, C., Xu, D., & Zhang, H. (2024). Multi-stage image inpainting using improved partial convolutions. *IET Image Processing*.
[48]. Chen, Y., Xia, R., Yang, K., & Zou, K. (2024). DNNAM: Image inpainting algorithm via deep neural networks and attention mechanism. *Applied Soft Computing*, *154*, 111392.
[49]. Y. Zeng, J. Fu, H. Chao, and B. Guo, "Learning pyramid-context encoder network for high-quality image inpainting," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2019, pp. 1486–1494
[50]. G. Liu, F. A. Reda, K. J. Shih, T.-C. Wang, A. Tao, and B. Catanzaro, "Image inpainting for irregular holes using partial convolutions," in Proc. Eur. Conf. Comput. Vis. (ECCV), 2018, pp. 85–100.
[51]. X. Hong, P. Xiong, R. Ji, and H. Fan, "Deep fusion network for image completion," in Proc. 27th ACM Int. Conf. Multimedia, Oct. 2019, pp. 2033–2042.
[52]. Z. Yan, X. Li, M. Li, W. Zuo, and S. Shan, "Shift-Net: Image inpainting via deep feature rearrangement," in Proc. Eur. Conf. Comput. Vis. (ECCV), 2018, pp. 1–17.
[53]. Liu, L., & Liu, Y. (2022). Load image inpainting: An improved U-Net based load missing data recovery method. *Applied Energy*, *327*, 119988.
[54]. Hou, Y., Ma, X., Zhang, J., & Guo, C. (2024). Symmetric Connected U-Net with Multi-Head Self Attention (MHSA) and WGAN for Image Inpainting. *Symmetry*, *16*(11), 1423.
[55]. Quan, W., Zhang, R., Zhang, Y., Li, Z., Wang, J., & Yan, D. M. (2022). Image inpainting with local and global refinement. *IEEE Transactions on Image Processing*, *31*, 2405-2420.
[56]. Liu, J., Gong, M., Gao, Y., Lu, Y., & Li, H. (2024). Bidirectional interaction of CNN and Transformer for image inpainting. *Knowledge-Based Systems*, 112046.
[57]. Chen, M., Zang, S., Ai, Z., Chi, J., Yang, G., Chen, C., & Yu, T. (2023). RFA-Net: Residual feature attention network for fine-grained image inpainting. *Engineering Applications of Artificial Intelligence*, *119*, 105814.
[58]. Meng, J., Liu, W., Shi, C., Li, Z., & Liu, C. (2024). Degression receptive field network for image inpainting. *Engineering Applications of Artificial Intelligence*, *138*, 109397.
[59]. Ma, X., Zhou, X., Huang, H., Jia, G., Wang, Y., Chen, X., & Chen, C. (2024). Uncertainty-aware image inpainting with adaptive feedback network. *Expert Systems with Applications*, *235*, 121148.