IMAGE INPAINTING USING DEEP LEARNING TECHNIQUES: A REVIEW

Haritha L
Department of Electronics and Communication
College of Engineering Trivandrum
Trivandrum, India
tve21ecsp09@cet.ac.in

Prajith C A
Department of Electronics and Communication
College of Engineering Trivandrum
Trivandrum, India
caprajith@cet.ac.in

2023 International Conference on Control, Communication and Computing (ICCC) | 979-8-3503-3412-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICCC57789.2023.10165271

***Abstract*—Image inpainting is a significant research area in the field of computer vision, with a diverse range of applications in image processing. Traditional image inpainting techniques proved to ineffective in generating better inpainted results, as they tend to produce inconsistent and multiple image semantics. Due to the advancement of Deep learning (DL) in recent years, this can be employed to address the issues with conventional image inpainting techniques. Currently, researches are mainly focused on image inpainting techniques based on deep learning using CNN, GAN etc. In this article, several deep learning based image inpainting techniques are been reviewed.**

***Index Terms*—CNN, Deep learning, GAN, Image inpainting.**

I. INTRODUCTION

Image inpainting is a restoration technique which uses surrounding known information to infer the missing contents of an image. The known information may be structural, statistical, semantic etc [1] [2]. These different information are utilized by different techniques to be used for a wide range of applications such as removing obstructions in images, image super-resolution, face image editing, repairing damaged images during imaging etc.

Image inpainting research started in 1993, with a disocclusion research by Nitzberg's [3] for repairing images using a computer. Later in 2000, Bertalmio *et al.* [4] proposed an digital image inpainting algorithm that can automatically refill the missing region (hole) using the surrounding data and coined the term 'image inpainting'. One of the advantage of this technique is the automatic restoration of missing regions with different structures and backgrounds.

Image inpainting methods can be generally classified into two: Traditional and Deep learning based image inpainting methods. Some of the examples of the traditional image inpainting techniques are Diffusion approaches [5] [6], exempler methods [7] [8], and hybrid methods [9] [10]. Traditional methods uses the structural and texture information rather than the semantic contents leading to the imperfect restoration of the missing region of an image, even after balancing the structural and texture inpainting [11] gradually. Recent image inpainting research employs the statistical data and geometric structure.

The drawbacks of the traditional inpainting techniques are been resolved to an extent with the introduction of deep learning in image inpainting, which provided much superior and improved quality inpainted results in comparison. As the research progressed, the convolutional neural network representing the feed-forward neural network captured the abstract information in the images [12]. And later in 2014, another deep learning framework named Generative Adversarial Network (GAN) [13] was developed that exhibited spectacular inpainted outcomes and significant improvement in image inpainting quality. The inpainted image is analysed quantitatively using evaluation metrices like Peak Signal to Noise Ratio (PSNR), Frechet Inception Distance (FID), Structural Similarity Index measure (SSIM), Inception Score (IS), Multi-Scale Structural Similarity (MS-SSIM), Universal Quality Index (UQI), Mean Squared Error (MSE) and Learned Perceptual Image Patch Similarity (LPIPS).

Due to the rapid advancement of the deep learning based image inpainting techniques, a well-known model named, Context Encoder (CE) was established by Pathak *et al.* [14]. This was the first successful GAN implementation in image inpainting. CE is generally a convolutional neural network that is trained to generate the missing image region (hole) using it's surrounding information. During training process, it is tested using both standard pixel-wise reconstruction loss and reconstruction plus an adversarial loss. As it can handle multiple modes in the output better, whereas the latter gives significantly sharper results. CE learns the representation that captures both the appearance and semantics of visual features of the image.

The paper is structured into two sections: Section II covers the background of the deep learning-based image inpainting, followed by two subsections A and B, where the research on image inpainting using CNN and GAN architecture is reviewed respectively. In addition, the most effective model for image inpainting using deep learning methods among the reviewed techniques is presented towards the end of Section II. Section III summarises some of the challenges faced by existing methodologies, some recommendations and relevant conclusions.

II. IMAGE INPAINTING USING DEEP LEARNING TECHNIQUES

This section briefly presents the main deep learning networks that lay foundation for the future inpainting works. The two most significant neural networks used in the deep learning based image inpainting techniques are CNN and GAN.

A. Convolutional Neural Network (CNN)

CNN is a deep feedforward neural network consisting many convolutional layers and pooling layers in layered structure. The convolutional layer performs the input feature extraction process and outputs a feature map. Whereas the pool layers perform downsampling to reduce data dimension. This retains the main input features and prevent the overfitting problem efficiently. CNN is also employed with an activation function (like ReLU) for the nonlinear data transformation. The model extracts and recombine the features after numerous superpositions and learns the complicated feature learning [15]. Due to inefficiency of CNN in image inpainting, the variant form of it is preferred. U-Net and Fully Convolutional Network (FCN) are the two widely used modified CNN networks, which served as the base for the future inpainting algorithms. Both these networks use the Encoder-Decoder model which is discussed below prior to the two networks.

1) Encoder-Decoder Model: Encoder-Decoder is an important design model in deep learning. Encoder extracts the input features, preserve and learn the information contained in it. And the decoder uses this feature map to recover the corrupted image [16].

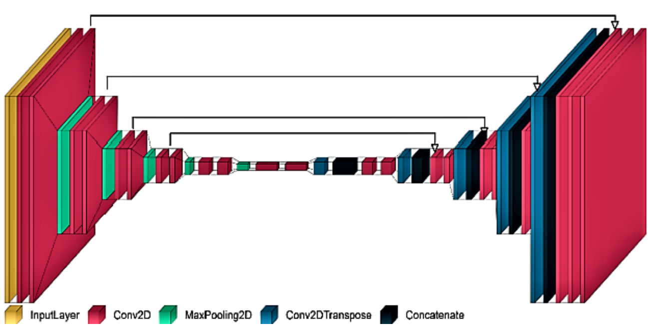

Fig. 1. Encoder decoder structure[35]

2) Fully Convolutional Network (FCN): FCN was initially proposed by Long *et al.* [17] for semantic segmentation and later it is been introduced into image inpainting techniques [18]. FCN replaces the full connection layers of conventional CNN with the deconvolution layers and upsamples the output back to the input size. The final convolutional layer of the FCN generates the most significant high-dimensional feature map of small size. This is fed to the deconvolutional layer to perform upsampling. In general, FCN uses the convolutional layer as an encoder [16] for feature extraction and noise removal whereas the deconvolution layer as decoder for image reconstruction.

FCN is also been utilized as a GAN generator due to it's spectacular inpainting results. Pathak *et al.* [14] introduced a context-based pixel prediction-based unsupervised visual feature learning system for inpainting images. This encoder-decoder system utilized the context encoders and the deconvolution network respectively. The encoder uses the AlexNet [15] architecture, whereas the channel-wise fully connected layers were employed for the encoder-decoder connection. The proposed methodology generated better inpainted images and also learned feature representations that are competitive with the pretrained existing models in semantic inpainting.

A multi-scale neural patch synthesis method was proposed by C Yang *et al.* [19] that preserves the contextual structure and also generates high-frequency details. The generation is accomplished by coordinating and updating patches with the most similar feature correlations of the deep classification network in the middle layer. This method generated a crisper and cohesive high-resolution inpainted images with better accuracy. In addition, this method also exhibited that FCN can effectively capture the global structure and the semantics of the image.

Y Song *et al.* [20] introduced a model that uses the segmentation information. Utilization of this information lead to clearer recovered boundary between diverse semantic regions and better texture within semantically consistent segments. This model uses two steps: Segmentation Prediction Network (SP-Net) followed by the Segmentation Guidance Network (SG-Net). The segmentation label is predicted at the first stage which is fed to the SG-Net to generate semantic guided inpainted results. This model inpainted images with an improved quality in comparison to the other methods. Hence, this interactive segmentation guidance led to possibility of the multi-modal image inpainting predictions.

X. Gao *et al.* [21] suggested an adversarial neural network based approach for face image inpainting. The global and local discriminator nets in the neural networks enhanced the face image inpainting quality. For improving the model's training stability, the weighted MSE loss and the GAN loss functions are used. The ultimate aim of the network training is to deceive the discriminator network's decision-making and allow the identification network to generate the actual image. This model is built and implemented on only 50% of the known image area availability.

3) U-Net: Ronneberger *et al.* [22] first introduced U-Net for the biomedical image segmentation and later on it is also utilized for semantic segmentation and image inpainting. It is similar to the FCN with an exception of being a symmetrical network. The first half of U-Net encodes the features using conventional convolution layer and the second part generates the results using upsampling [22]. Both the encoder and decoder consist of two convolution layers with max-pooling layer and upsampling layer in the encoder and decoder respectively. The feature map obtained from the encoder and decoder are concatenated during each upsampling step to retain more spatial information. And the next convolution layer uses this information to learn and produce a better detailed output.

Traditional CNN-based image inpainting methods using standard convolution for the damaged image areas with white or random noise produces unreasonable inpainted results. This occurs due to the convolution operation also been performed on invalid pixels. G. Liu *et al.* [23] introduced the use of partial convolution with an automated mask updation method for image inpainting. Here the convolution and re-normalization is performed on valid pixels only. This model produced better

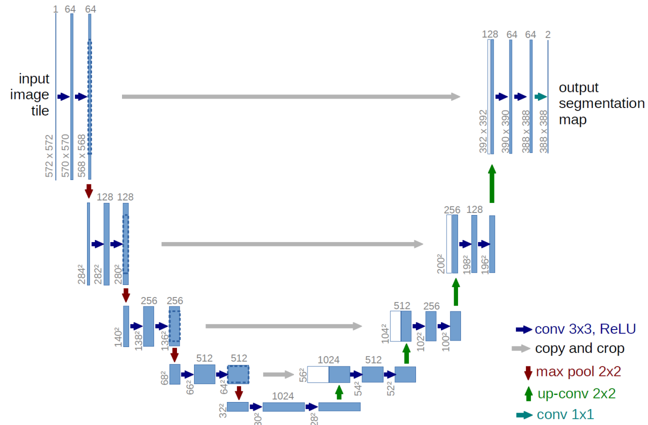

Fig. 2. U-Net architecture for medical image segmentation[22]

inpainting results for the images with irregular mask. This model can handle image holes of any size, shape and at any location. One of the drawback of this method is it's inability to restore a larger hole and the images with sparsely structured compositions like bars on the door.

Yu *et al.* [24] proposed a deep generative model that can generate new image structures and also use adjacent image attributes as references for prediction improvement during network training. This model consist of two stages: 1) a dilated convolutional layer trained with the reconstruction loss to delineate the missing image contents, 2) and then used a contextual attention module with convolution filter to identify and compare the most alike patch in the entire image with respect to the inpainting area using Softmax [24]. Here the inpainting results are influenced by the patch size used for background patch similarity computation. The proposed method produced greater quality inpainting results compared to other inpainting techniques.

A learning-based method was proposed by Z. Lin *et al.* [25] to provide visually coherent completion of a high resolution image with missing contents. The inpainting task is split into two distinct stages: inference and translation, with each stage been modelled with the deep neural network to overcome the challenge of directly learning the high dimensional image data distribution. The training becomes easy due to the functioning of two image-feature translation in a smaller space. This model produced a higher visual quality inpainted images. And it can also inpaint an image with realistic and precise details in a feed-forward fashion.

A progressive image inpainting technique using full-resolution residual network (FRRN) was suggested by Z Guo et al. [26]. The well designed architecture aids in easier feature integration and texture prediction. And N Blocks and dilation techniques are also used to further enhance the completion quality. The intermediate restoration efficiency is improved using a step loss function. This framework generated much better inpainted results both qualitatively and quantitatively.

Y. Zeng *et al.* [27] proposed a deep generative model using Pyramid-context ENcoder Network (PEN-Net). The PEN-Net restores the image by encoding contextual semantic contents from the high resolution input and then decodes the learnt semantic features back into images. For boosting the U-Net capacity, three main tailored components are also suggested namely: pyramid-context encoder, multi-scale decoder and an adversarial loss. The PEN-Net used the attention transfer network (ATN) in the pyramid-context encoder to learn and transfer the region affinity between the corrupted and uncorrupted area from the high-level semantic feature map to the low level feature map. Since the holes are been filled by shifting attention in a pyramidal pattern, both semantic and visual coherence for inpainted images can be guaranteed. It emphasised two primary variations between the ATN, cross-layer attention transfer and pyramid filling which adds additional specificity to low-level characteristics from high level semantic information. Moreover, this model has faster convergence and produced more realistic test results.

N Wang *et al.* [28] proposed a dynamic selection network (DSNet) which used the valid contents in the known image area for image inpainting. The DSNet consist of two dynamic selection modules: the Validness Migratable Convolution (VMC) and Regional Composite Normalisation (RCN) module, both using a dynamic selection mechanism for the effective utilization of valid pixels. The VMC dynamically chooses the spatial sampling locations during the convolution phase enabling a better feature extraction. The RCN module normalizes the feature regions selectively and combines several normalization techniques. This model produced a realistic and finely detailed images and outperformed the state-of-the-art methods both statistically and qualitatively. But it possess some drawbacks: 1) High processing time due to unoptimized convolution and normalisation procedure and 2) Difficulty in restoring original information in case of severely damaged regions.

M. Givkashi *et al.* [29] suggested an image inpainting technique using a network similar to U-Net which can extract various features from the image and produce superior inpainting results. The final inpainted result is enhanced by replacing the damaged image pixels with the restored pixels of the output images. This model produced a high quality inpainted results in comparison to the conventional inpainting methods.

B. Generative Adversarial Network (GAN)

GAN is a class of machine learning frameworks developed by Ian Goodfellow [13] and his coworkers in June 2014. It comprises of two neural networks: Generator and Discriminator, which compete against each other like a zero-sum game. Due to the strong data creation capabilities, GAN is well suited for image inpainting tasks. Image generation, video generation and voice generation are some of it's widely used applications. Context Encoder (CE) network [14] was the first GAN and CNN combined image inpainting network.

An innovative idea of training the generative model was put forth by M Mirza *et al.* [30]. This model is a conditional version of the generative adversarial net with an additional input such as class labels or datas from other modalities which

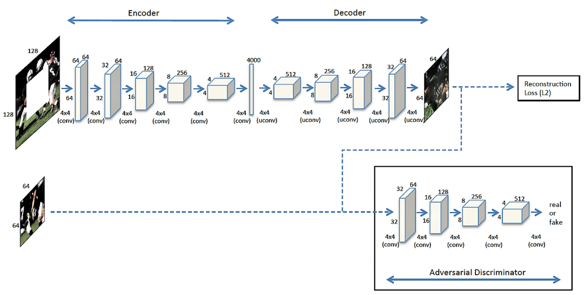

Fig. 3. Contextual Encoder Decoder Structure [14].

are conditioned on to both the generator and discriminator. In response to the class labels, this model generates the MNIST digits. Moreover, this model could also be used for image tagging approach.

A novel image inpainting approach using Exemplar GANs (ExGANs) was proposed by B. Dolhansky *et al.* [31], where the identity of the object to be removed or modified is preserved and considered at the inference time. This is a conditional GAN (cGAN) model that uses exemplar information in the form of either the reference image of the region to be inpainted or the perceptual code defining the object for generating high-quality, personalized inpainting results. This methodology is performed for closed to open eye inpainting task that generated photo-realistic, personalized inpainting results which are both perceptual and semantically plausible. This model produced a much superior perceptual inpainting result by utilizing the exemplar information.

U. Demir *et al.* [32] presented a PGGAN-based image inpainting technique that combines a global GAN (G-GAN) architecture with discriminator network using the PatchGAN approach. For capturing both the extensive global image features and the local image texture continuity, first PGGAN shares the network layers between PatchGAN and G-GAN, then separates pathways to generate two adversarial losses. Both the quantitative and visual evaluations of this approach exhibited significant improvements in comparison to the state-of-the-art methods.

A method for pluralistic image completion was suggested by C. Zheng *et al.* [33]. A unique and probabilistically based GAN framework with two parallel approaches are suggested. 1) A reconstructive approach that uses only one given ground truth to determine the prior distribution of the missing portions and uses this to recreate the original image. 2) A generative path for which the distribution produced by the first path is linked to the conditional prior. Moreover, a novel short+long term attention layer is put forth to enhance the appearance consistency by using the distant relationship between decoder and encoder characteristics. This method generated superior quality completion results and diversified plausible output.

K. Nazeri *et al.* [34] developed a novel image inpainting method that can reproduce fine detailed and plausible structures in the missing regions. The proposed two-stage adversarial model, EdgeConnect consists of an edge generator followed by an image completion network. The image completion network completes the missing areas (holes) by filling it by utilizing the hallucinated hole edges (both regular and irregular) that were produced by the edge generator.

A Contextual Residual Aggregation (CRA) technique was suggested by Zili Yi *et al.* [35] which could generate high-frequency residuals for missing contents by weighted aggregating residuals from contextual patches, necessitating a low-resolution prediction from the network. Advantages of this model are: 1)Requires a less high-resolution training datasets, 2)Better inpainting quality, 3)Inpaint images of size 8K with fairly large hole sizes, 4) Faster performance (3x-6x) than other inpainting methods on high resolution images of same size between 1K and 2K, 5) Reduced memory and the computing resource cost. The increase in resolution and hole size neither lengthened the processing time nor degraded the inpainting quality.

C. Wang *et al.* [36] presented a dual-pyramidal inpainting architecture named DPNet to overcome the drawbacks of image inpainting due to improper representation and erroneous regularisation of existing features. This autoencoder network combines dynamic regularisation and adequate feature learning. The layer-wise pyramidal convolution in the encoder extracts multiscale features and the Pyramidal Attention Mechanism (PAM) in the decoder address the patch degrading issue in prior cross-scale non-local schemes and acquire finer patches from the learning layers directly. And a dynamic normalising technique utilizing the spatial mask information updated in the encoder is introduced to further ensure the feature consistency and integrity and eliminate the mask error accumulation in existing works. The dynamic normalising and the dual-pyramidal structure greatly enhanced the inpainting quality. The DPNet outperformed the existing techniques in terms of producing realistic and finely detailed inpainted images.

Image inpainting techniques are also been used for damaged medical image restoration. These images need to be repaired for a variety of reasons, such as information loss during the imaging process due to external noise, biological structure prediction to aid in therapy and sampling algorithm optimization.

Armanious Karim *et al.* [37] proposed ip-MedGAN, a GAN model for the medical image inpainting. This model consists of two patch-based discriminator networks with extra stylistic and perceptual losses for the realistically detailed and contextually consistent inpainting of missing information. On two distinct medical modalities, the proposed system performed better than the existing natural image inpainting approaches both quantitatively and qualitatively.

An Edge-Guided Generative Adversarial Network (EG-GAN) was put forth by Y Chai *et al.* [38] to recover brain MRI images using the missing through-plane slices as image masks. Here the image restoration task is split into two: edge connection and contrast completion with both of them utilizing GAN for the inpainting task. On comparing the EG-GAN to conventional imputation methods like the Densely Connected Super Resolution Network with GAN and the

TABLE I
SUMMARY OF DEEP LEARNING BASED IMAGE INPAINTING TECHNIQUES

Method	**Architecture**	**Dataset**	**Results**
[14] An unsupervised system for learning visual features that uses context-based pixel prediction	Encoder-Decoder	Paris StreetView, Imagenet	PSNR: 18.58
[19] Multi- Scale Neural Patch Synthesis Method	Contextual Encoder	ImageNet,Paris StreetView	PSNR: 18.00
[20] Segmentation Prediction and Guidance Method	FCN	Helen Faces, CityScapes	PSNR: 34.26, SSIM: 0.9591
[21] Based on Generative Adversarial framework	CNN	CelebA	SSIM: 0.89, PSNR: 31.86
[23] Partial Convolution with automated mask updation	U-Net	CelebA, Places2, ImageNet	For mask ratio (0.5,0.6) with borders; SSIM: 0.484, PSNR: 19.04
[24] Feed-forward generative network utilizing contextual attention mechanism	Fully convolutional neural network	CelebA, DTD, Places2, CelebA-HQ	PSNR: 18.91
[25] Learning based inpainting framework	U-Net in GAN	COCO, ImageNet	SSIM: 0.56
[26] A FRRN module with dilation layer to generate residuals and update mask for image completion	Full Resolution Residual Network	CelebA, Places2	For mask ratio of 40-50%; SSIM: 0.776, PSNR: 22.08
[27] Deep Generative inpainting model based on Pyramidal Context encoder Network	U-Net	Places2, DTD, Facade CelebA-HQ	MS-SSIM: 0.7809
[28] Dynamic Selection Network using VMC and RCN	U-Net	Places, CelebA Paris StreetView	For CelebA images with mask ratio (0.5,0.6); MS-SSIM: 0.8012
[29] Replaces damaged image pixels with the output reconstructed pixels using Autoencoder & pixel prediction	Encoder-Decoder model inspired with U-Net	ADE20K and KODAK dataset	For variable & thick mask (KODAK); PSNR: 28.59, SSIM: 0.9495
[30] Automated image taggging with multi-label prediction	cGAN	MIR Flickr 25,000 dataset	Effectively generated MNIST digits for each labels
[31] Use exempler information to retain the semantic content and personalize the output image	Ex-GAN	Face images	MS-SSIM: 0.0078
[32] Patch based Image Inpainting using GAN framework	PGGAN	Google StreetView, Paris StreetView, Places	For a 512x512 image; SSIM: 0.884, PSNR: 18.9
[33] Pluristic Image Completion with prior image distribution from one ground truth in two parallel paths	GAN	Paris StreetView, ImageNet, CelebA-HQ	PSNR: 20.10
[34] EdgeConnect generates a hallucinated hole edge and use the surrounding information to fill it	Two stage Adversarial framework	Places2, CelebA Paris StreetView	For fixed regular mask; SSIM: 0.823, PSNR: 21.75
[35] Contextual Residual Aggregation mechanism	GAN	CelebA-HQ, Places2, DIV2K	For a 4K image resolution; MS-SSIM: 0.8840
[36] Based on Dual Pyramidal framework	GAN	Places, CelebA, Paris StreetView	For CelebA-HQ images with random shaped mask; PSNR: 27.01, SSIM: 0.870
[37] Based on ip-MedGAN framework	CGAN	FLAIR(scanned image of brain region)	SSIM: 0.3818, PSNR: 18.32
[38] Based on Edge guided-GAN framework	Patch-GAN	Human Connectome Project Datasets	PSNR: 40.4849, SSIM: 0.9745

Context Encoder, it has a better SSIM, PSNR, signal texture and conspicuity.

Among all of the inpainting techniques, the models [38], [20] and [21] have proven to be more effective at producing the inpainted results with higher PSNR values. EG-GAN [38] restored the fine details of the brain anatomy in the missing region with visually plausible imputation results. [20] produced high quality inpainting results with improved recovered texture and boundaries. [21] produced an excellent face image reconstruction, particularly for the eye and hair regions.

III. CONCLUSION

Image inpainting is a very challenging and an active computer vision task in the image processing field. Several methods were employed to produce a visually pleasing and semantically meaningful inpainted results. This article provide a review of deep learning based methods in image inpainting using CNN and GAN. Even though these techniques produced much better inpainted results, it also faced some drawbacks which needs to be eliminated. The CNN based inpainting methods making use of FCN and U-Net is effective for the image reconstruction. But they posess challenges like optimization, high time consumption for inpainting small hole regions, inefficiency to inpaint large holes etc. So convolutional variants like partial and gated convolution and attention mechanisms were employed to enhance the quality of image inpainting. Due to the strong data creation capabilities and adversarial learning mechanisms GAN based inpainting model could generate high quality inpainted results. Moreover, it is used for medical image restoration due to the high reconstruction accuracy. The variant GAN based image inpainting techniques produced better inpainted results but only in specific situations. Another disadvantage is that these techniques cannot be reproduced further. Face image inpainting using GAN had a severe drawback of not preserving the human identity as such. Hence the future progress should be done in the directions to overcome these limitations and produce a better deep learning based inpainting methodology that can inpaint damaged images with multiple holes of any size at multiple locations with the semantic and texture content of the missing portion as such.

REFERENCES

[1] J. Sun, L. Yuan, J. Jia, and H.-Y. Shum, "Image completion with structure propagation," in *ACM SIGGRAPH 2005 Papers*, 2005, pp. 861–868.

[2] H. Liu, B. Jiang, Y. Xiao, and C. Yang, "Coherent semantic attention for image inpainting," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 4170–4179.

[3] M. Nitzberg, D. Mumford, and T. Shiota, *Filtering, segmentation and depth*. Springer, 1993, vol. 662.

[4] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, "Image inpainting," in *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, 2000, pp. 417–424.

[5] K. He and J. Sun, "Statistics of patch offsets for image completion," in *European conference on computer vision*. Springer, 2012, pp. 16–29.

[6] D. Tschumperlé and R. Deriche, "Vector-valued image regularization with pdes: A common framework for different applications," *IEEE transactions on pattern analysis and machine intelligence*, vol. 27, no. 4, pp. 506–517, 2005.

[7] M. Elad, J.-L. Starck, P. Querre, and D. L. Donoho, "Simultaneous cartoon and texture image inpainting using morphological component analysis (mca)," *Applied and computational harmonic analysis*, vol. 19, no. 3, pp. 340–358, 2005.

[8] J. Mairal, M. Elad, and G. Sapiro, "Sparse representation for color image restoration," *IEEE Transactions on image processing*, vol. 17, no. 1, pp. 53–69, 2007.

[9] C. Guillemot and O. Le Meur, "Image inpainting: Overview and recent advances," *IEEE signal processing magazine*, vol. 31, no. 1, pp. 127–144, 2013.

[10] A. Bugeau, M. Bertalmío, V. Caselles, and G. Sapiro, "A comprehensive framework for image inpainting," *IEEE transactions on image processing*, vol. 19, no. 10, pp. 2634–2645, 2010.

[11] C. Fan, K. Ren, L. Meng *et al.*, "Advances in digital image inpainting algorithms based on deep learning," *J. Sign. Process*, vol. 36, no. 01, pp. 102–109, 2020.

[12] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep learning for computer vision: A brief review," *Computational intelligence and neuroscience*, vol. 2018, 2018.

[13] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, 2020.

[14] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros, "Context encoders: Feature learning by inpainting," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2536–2544.

[15] Y. LeCun, K. Kavukcuoglu, and C. Farabet, "Convolutional networks and applications in vision," in *Proceedings of 2010 IEEE international symposium on circuits and systems*. IEEE, 2010, pp. 253–256.

[16] X. Mao, C. Shen, and Y.-B. Yang, "Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections," *Advances in neural information processing systems*, vol. 29, 2016.

[17] J. Long, E. Shelhamer, and T. Darrell, "Fully con-volutional networks for semantic segmentation," *IEEE Transactions on Pattern Analysis Machine Intelligence*, vol. 39, no. 4, pp. 640–651, 2014.

[18] S. Chaudhury and H. Roy, "Can fully convolutional networks perform well for general image restoration problems?" in *2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA)*. IEEE, 2017, pp. 254–257.

[19] C. Yang, X. Lu, Z. Lin, E. Shechtman, O. Wang, and H. Li, "High-resolution image inpainting using multi-scale neural patch synthesis," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 6721–6729.

[20] Y. Song, C. Yang, Y. Shen, P. Wang, Q. Huang, and C.-C. J. Kuo, "Spg-net: Segmentation prediction and guidance network for image inpainting," *arXiv preprint arXiv:1805.03356*, 2018.

[21] X. Gao, M. Nguyen, and W. Q. Yan, "Face image inpainting based on generative adversarial network," in *2021 36th International Conference on Image and Vision Computing New Zealand (IVCNZ)*. IEEE, 2021, pp. 1–6.

[22] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[23] 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 *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 85–100.

[24] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang, "Generative image inpainting with contextual attention," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 5505–5514.

[25] Y. Song, C. Yang, Z. Lin, X. Liu, Q. Huang, H. Li, and C.-C. J. Kuo, "Contextual-based image inpainting: Infer, match, and translate," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 3–19.

[26] Z. Guo, Z. Chen, T. Yu, J. Chen, and S. Liu, "Progressive image inpainting with full-resolution residual network," in *Proceedings of the 27th acm international conference on multimedia*, 2019, pp. 2496–2504.

[27] Y. Zeng, J. Fu, H. Chao, and B. Guo, "Learning pyramid-context encoder network for high-quality image inpainting," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 1486–1494.

[28] N. Wang, Y. Zhang, and L. Zhang, "Dynamic selection network for image inpainting," *IEEE Transactions on Image Processing*, vol. 30, pp. 1784–1798, 2021.

[29] M. H. Givkashi, M. Hadipour, A. PariZanganeh, Z. Nabizadeh, N. Karimi, and S. Samavi, "Image inpainting using autoencoder and guided selection of predicted pixels," in *2022 30th International Conference on Electrical Engineering (ICEE)*. IEEE, 2022, pp. 700–704.

[30] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv preprint arXiv:1411.1784*, 2014.

[31] B. Dolhansky and C. C. Ferrer, "Eye in-painting with exemplar generative adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7902–7911.

[32] U. Demir and G. Unal, "Patch-based image inpainting with generative adversarial networks," *arXiv preprint arXiv:1803.07422*, 2018.

[33] C. Zheng, T.-J. Cham, and J. Cai, "Pluralistic image completion," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 1438–1447.

[34] K. Nazeri, E. Ng, T. Joseph, F. Z. Qureshi, and M. Ebrahimi, "Edge-connect: Generative image inpainting with adversarial edge learning," *arXiv preprint arXiv:1901.00212*, 2019.

[35] Z. Yi, Q. Tang, S. Azizi, D. Jang, and Z. Xu, "Contextual residual aggregation for ultra high-resolution image inpainting," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 7508–7517.

[36] C. Wang, M. Shao, D. Meng, and W. Zuo, "Dual-pyramidal image inpainting with dynamic normalization," *IEEE Transactions on Circuits and Systems for Video Technology*, 2022.

[37] K. Armanious, Y. Mecky, S. Gatidis, and B. Yang, "Adversarial inpainting of medical image modalities," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 3267–3271.

[38] Y. Chai, B. Xu, K. Zhang, N. Lepore, and J. C. Wood, "Mri restoration using edge-guided adversarial learning," *IEEE Access*, vol. 8, pp. 83 858–83 870, 2020.