Proceedings of the 5th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/19/20231009

# Research on image inpainting methods based on machine learning

**Zhengnan Liu[1,†], Mingyang Qin[2,3,†]**

[1]School of Intelligence and Electronic Engineering, Dalian Neusoft University of Information, Dalian, 116000, China
[2]Houston International College, Dalian Maritime University, Dalian, 116000, China

[3]dd250768415@dlmu.edu.cn
[†]All authors contribute equally

**Abstract.** The technique of restoring sections of a picture that have been lost or damaged is known as "image inpainting." In light of recent developments in machine learning, academics have begun investigating the possibility of using deep learning methods to the process of picture inpainting. However, the current body of research does not include a comprehensive review of the many different inpainting methods that are based on machine learning, nor does it compare and contrast these methods. This article provides an overview of some of the most advanced and common machine learning based image restoration techniques that are currently available. These techniques include Multivariate inpainting technology and Unit inpainting technology, such as Context-Encoder Network, Generative Adversarial Network (GAN), and U-Net Network. We examine not just the benefits and drawbacks of each method, but also the ways in which it might be used in a variety of settings. At the conclusion of the piece, we predict that machine learning-based inpainting will continue to gain popularity and application in the years to come.

**Keywords:** image inpainting, machine learning, unit inpainting, multivariate inpainting.

## 1. Introduction

The discipline of computer vision has long prioritized research in inpainting technologies. With the development of deep learning, inpainting based on deep learning has steadily grown in popularity as a study area. Deep learning technology can learn the characteristics in a large number of image data, so that inpainting algorithm can perform better in complex scenes. For damaged images, deep learning-based methods can learn more accurate pixel information from existing datasets, thereby obtaining higher quality repair results. In addition, due to the continuous development of deep learning technology, new models and algorithms are also emerging, which makes the inpainting algorithm based on deep learning have higher efficiency and better repair effect. This review paper will review and analyze inpainting methods based on deep learning in recent years (Figure 1). First, we will introduce the background and development of inpainting technology, and briefly introduce the deep learning technology and its application in computer vision. Secondly, we will classify and introduce the common inpainting algorithms based on deep learning. These algorithms include multivariate inpainting methods in unit inpainting methods and inpainting algorithms based on generative

DOI: 10.54254/2755-2721/19/20231009

antagonism network, convolutional neural network, and multivariate inpainting methods. Finally, we will compare and analyze these algorithms, evaluate their advantages and disadvantages, as well as their applicability in different scenarios. This paper aims to provide reference and enlightenment for researchers and developers, and summarizes the methods of inpainting grounded in deep learning.

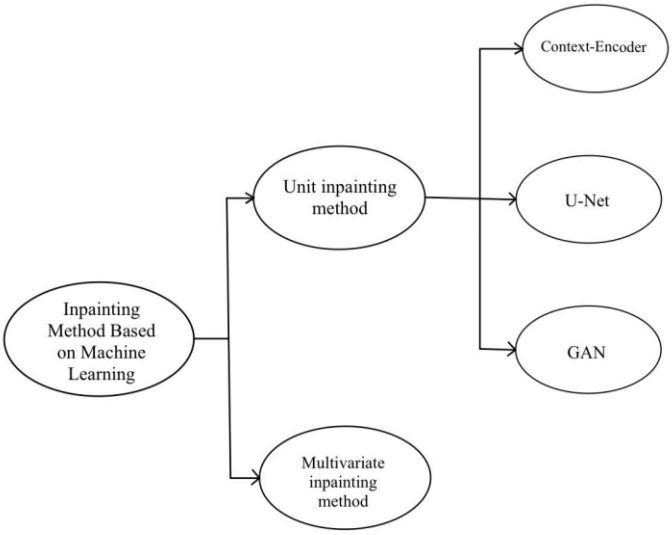


**Figure 1.** Overall classification of image inpainting methods based upon machine learning.

## 2. Image inpainting methods based on Context-encoder

The Context-encoder is a feature of the semantic network architecture that operates in an unsupervised manner [1]. It is derived from the Encoder-Decoder model architecture [2]. The structure of the model is illustrated in Figure 2. The context-encoder is comprised of two primary constituents, namely an encoder and a decoder. The procedure entails the compression of the input image into a low-dimensional depiction by the encoder, which is then supplied back to the decoder. Following this, the decoder generates a new image that is anticipated to exhibit similarities to the original image, despite the absence of certain elements.

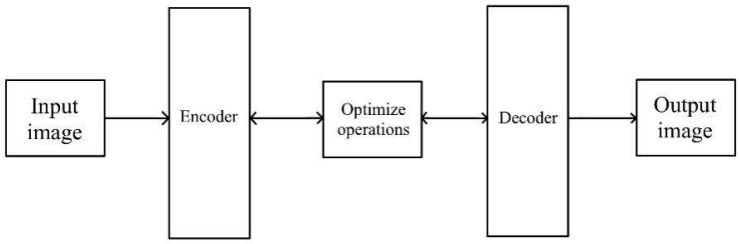


**Figure 2.** Structure of Encoder-Decoder model.

These methods can automatically generate content that matches the original image based on the known content in the image and the feature information around the area to be repaired and supplement the missing parts of the image, demonstrating superior performance in repair tasks. Furthermore, owing to its uncomplicated model architecture, the context encoder technique has emerged as a prominent restoration approach in contemporary times and has found utility in diverse domains, including but not limited to image manipulation, object elimination, and video restoration.

In the training of the model, the context encoder references both input reconstruction loss and adversarial loss to ensure the accuracy of the repaired image. Among them, reconstruction loss can

DOI: 10.54254/2755-2721/19/20231009

obtain the overall structure and contextual consistency of the missing part of the image, while adversarial loss can improve the flexibility of image generation. However, this method ignores the relationship between the repaired area and the overall area, which can lead to a deviation in the connection between the repaired area and the overall image.

Liao et al. have proposed a context decoder that relies on edge perception [3]. They have employed a fully convolutional network to restore the edge information of the missing image. The context encoder was utilised to repair the overall image, image defects, and the edge information that had been repaired. Yang et al. replaced all convolutional layers in the structure with a residual block structure to obtain prior information about the image, greatly improving the stability of the model during training [4]. Wang et al. decomposed the original network architecture into three parallel branch networks, each of which contains information sampled by different receptive field and feature resolution components of the image to extract and predict the overall and local information of the image at different levels [5]. Vo et al. divided model training into two stages [6]. In the first stage, they cited structural losses from joint reconstruction and feature reconstruction. In the second stage, they use adversarial losses to optimize the model structure. This structure can further increase the repair effect and accuracy of the model in different scenarios. In general, each methodology possesses unique merits, and the selection of the optimal approach may be contingent upon the particular demands of the image restoration undertaking. In cases where the edge information of an image is deemed crucial, Liao et al.'s approach is deemed optimal, whereas Wang et al.'s technique is deemed more appropriate for capturing multi-scale information.

## 3. Image inpainting methods based on U-Net

The U-Net network structure includes encoder subnets and decoder subnets, each with four sub modules. Two convolutional layers, one activation layer, and one maximum pooling layer make up the encoder submodule, which is used to downsample input feature maps (Figure 3). A transposed convolutional layer, a concatenation layer, and two convolutional layers make up the decoder submodule. These layers are used to upsample the input feature map and cut the number of convolutional kernels in half. The U-Net network structure uses the 3x3 convolution core, uses the linear rectification function (ReLU) as the activation function, and uses the sigmoid function as the activation function of the convolution layer. Finally, the U-Net network better retains the image feature information through the connection and jump connection of the encoder and decoder, thus achieving excellent inpainting effect.

The U-Net network contains two features: firstly, the U-Net network concatenates and fuses feature maps of the same scale of the encoder and decoder, while there is a significant semantic gap between feature maps of different scales; The second issue is that the U-Net network cannot fully utilize the surface features of the input image, which mainly contain global information of the image, including target location and semantic relationships [7].

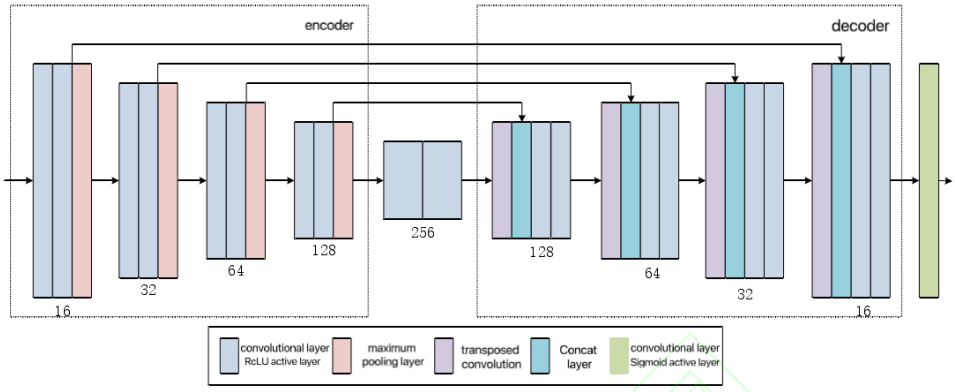


**Figure 3.** U-Net backbone network structure.

DOI: 10.54254/2755-2721/19/20231009

On this basis, Wang Weihua proposed inpainting algorithm based on depth feature rearrangement of double transition network [8]. On the basis of the U-Net network, two transfer connection layers are added to estimate the missing parts of the image by moving the encoder characteristics of the known area. A content loss function based on decoder features is introduced, and finally, the feature distance between the decoder and encoder is reduced to repair the image. Convolutional block attention module (CBAM) and residual network (ResNet) are introduced into Liu et al.'s proposed improved U-Net to handle large-scale data loss, notably recovering partial payload images [9].The improved U-Net adopts a simpler and shallower network structure specifically designed for repairing image rendering.

## 4. Image inpainting methods based on GAN

Goodfellow et al. devised the Generative Adversarial Network (GAN), a conventional unsupervised learning model [10]. The previously discussed context-encoder method is the first attempt to apply the concept of game confrontation to image restoration, and it has produced outstanding results, providing a solid foundation for future image repair research. Due to issues with image restoration, producing pictures that align with the desired image is fundamentally challenging. However, with the introduction of zero-sum game generation, the resulting defective images are no longer limited to the training image set. This approach offers greater versatility and diversity than earlier methods. That is why GAN class methods have been adopted in recent years to tackle picture restoration issues.

The structure of the conventional GAN approach is illustrated in Figure 4, with its core components consisting of the generator and discriminator. The generator produces images by leveraging random noise input, while the discriminator evaluates the fidelity of the generated content by comparing it to the original image. As the iterative process proceeds, both the discriminator and generator compete until they reach an equilibrium.

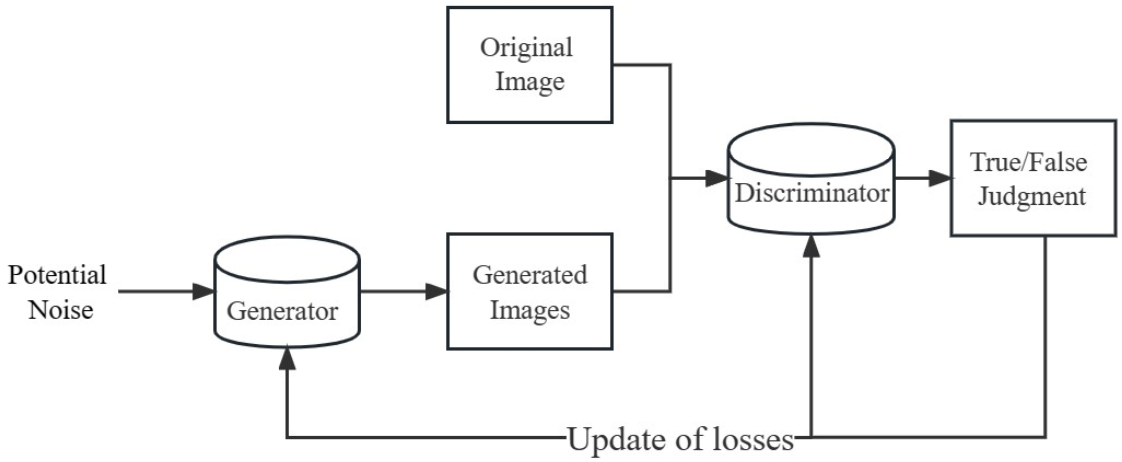


**Figure 4.** Structure of GAN model.

The structural diagram in Figure 4 illustrates that the GAN model's limitations during training are modest and that it is challenging to precisely manage mistakes at any given moment, which restricts the method's applicability and, in this approach, Mirza et al. suggested a generative adversarial network (CGAN) that may impose extra limitations [11]. The fundamental contribution of CGANs is the incorporation of additional information, specifically constraints, into the input data of the generator and discriminator in the original GAN.As a result, the CGAN approach enables GAN to build a specific picture with the provided information during the test stage and train the image with the matching extra information. Later, utilizing reference data concealed in known photos, Dolhansky et al. finished the eye restoration problem in natural images [12]. As illustrated in Figure 5, the network exhibits negative training by utilising a pair of images, namely Xi and ri, from a common training set.

DOI: 10.54254/2755-2721/19/20231009

The aforementioned visuals depict the reference image denoted as G, the defect image referred to as Zi, and the discriminator labelled as D.

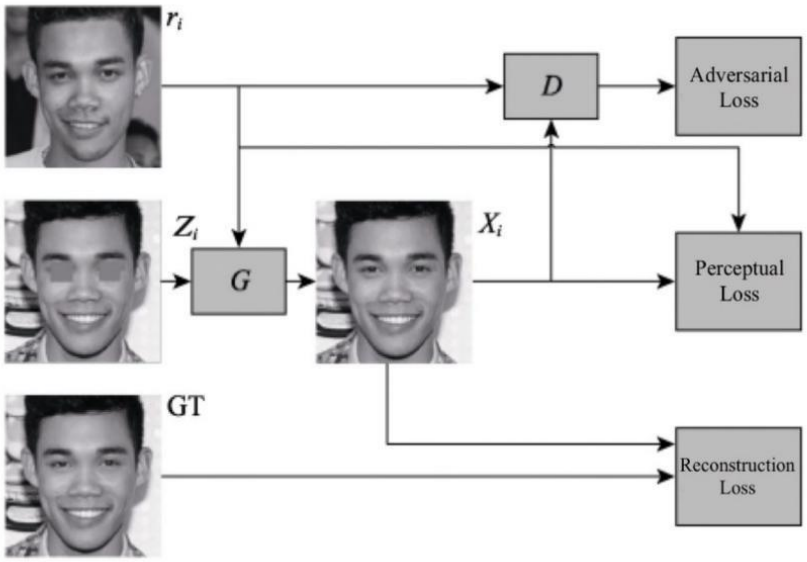


**Figure 5.** An eye repair network structure mentioned in reference [12].

Radford et al. created a deep convolution GAN (DCGAN) by tightly merging CNN and GAN in order to further improve the ability of model feature extraction [13]. The stability of network training and the quality of produced pictures were significantly improved by this design. The DCGAN model has introduced certain alterations to the conventional CNN framework. Substitute all pooling layers in the discriminator with strided convolutions and employ fractional-strided convolutions in the generator. Incorporate batch normalisation in both the generator and discriminator. To achieve deeper architectures, it is recommended to eliminate fully connected hidden layers. The Rectified Linear Unit (ReLU) activation function should be utilised in the generator for all layers, with the exception of the output layer, which should employ the hyperbolic tangent (Tanh) activation function. The utilisation of Leaky Rectified Linear Unit (LeakyReLU) activation function in the discriminator for all layers is recommended. The aforementioned instances serve as illustrations of the adaptability and benefits of generative adversarial networks (GANs) within the realm of image generation and restoration. Generative Adversarial Networks (GANs) have exhibited a wide range of capabilities in the fields of image generation and restoration. GANs have demonstrated versatility in image generation and restoration. CGAN adds extra information, Dolhansky et al. utilized reference data, and DCGAN improves stability and quality through deep learning techniques. These advancements showcase GANs' potential to produce high-quality and specific images in various applications. These advancements demonstrate the potential for GANs to produce high-quality and specific images in a variety of applications.

## 5. Image inpainting methods based on multivariate

Despite the fact that inpainting has made considerable strides, the majority of current research focuses on unit inpainting techniques. The unit inpainting method can only generate one image, while the multivariate inpainting method can generate many different images, thus improving the diversity and reliability of the repair results. Next, this paper will summarize the research status and progress of multivariate inpainting methods.

The feature of multivariate inpainting method is that it can generate a variety of different images, which are reasonable repair results. At present, researchers mainly use VAE, CVAE, GAN and other

DOI: 10.54254/2755-2721/19/20231009

model architectures to achieve multiple repair images, and propose some multivariate inpainting methods. Han et al., for instance, suggested a two-stage repair technique that generates the form and appearance of pictures using shape generation networks and appearance generation networks, respectively, and creates image diversity using an interactive encoder network [14]. Meanwhile, Dupont et al. created a pixel limited CNN model that can perform probabilistic semantic repair on picture distribution under specific visible pixel circumstances [15]. They also incorporated the PixelCNN model into repair tasks. These methods fully consider the characteristics of Semantic information and diversity, and can generate more reasonable and diverse repair results.

Furthermore, certain research endeavours employ concealed images as antecedent knowledge to facilitate the enhancement of network diversification. The method of using mask images as prior information to guide network diversification repair was proposed by Zhao et al. This method consists of three modules: conditional encoding, manifold projection, and generation [16]. In the conditional encoding module, the network learns conditional distribution information by inputting mask images. The manifold projection module projects mask distribution information and instance image space onto a common low dimensional manifold space, and learns one-to-one mapping between the two spaces. Finally, in the generation module, the network generates diverse images. These methods can effectively ensure the quality and diversity of repair results.

In the task of multivariate inpainting, more and more studies have used Transformer to improve the repair ability of the model [17]. A bidirectional autoregressive Transformer model has been created by researchers like Yu et al. for learning the autoregressive distribution of pictures in order to efficiently repair their various structures and remote locations [18]. Wan et al. used a bidirectional Transformer model in a similar manner to accomplish diversified appearance reconstruction of low-resolution pictures, utilizing appearance priors and upsampling CNN networks to lead high-fidelity texture detail restoration of images [19]. The advantage of these methods is that they can ensure the diversity and rationality of repair results. However, these methods also have some shortcomings, such as the lack of prior feature information, making it difficult to generate more reasonable semantic structures.

The benefit of the multivariate inpainting approach over the unit inpainting method is that it may provide more accurate and varied restoration outcomes by taking into account the features of semantic information and variety. The multivariate inpainting technique, however, also has significant drawbacks. First, to increase the model's capacity for repairs and variety, these techniques call for additional processing power and training data. Second, because these techniques prioritize pixel-level restoration over semantic-based repair, it is challenging to produce more logical semantic structures with them. Additionally, these approaches may be impacted by the external environment and data distribution because to the diversity and complexity of the look and form of numerous complex sceneries and objects, which might lead to erroneous repair findings. Therefore, future research needs to further explore how to better integrate Semantic information and diversity in the task of multivariate inpainting, so as to improve the rationality and accuracy of the repair results.

## 6. Comparison of different image inpainting methods

U-Net network has good accuracy and efficiency. In the field of inpainting, U-Net network is usually used to reconstruct missing image information. Through training, the U-Net network can learn advanced features of images and perform pixel level image repair. However, this method often suffers from issues such as unsmooth image reconstruction and unnatural textures. On the basis of the U-Net network, compared with the two methods of adding a transmission connection layer to estimate the missing part of the image and introducing ResNet and CBAM to deal with large-scale data loss, the method of adding a transmission connection layer is relatively simple and can quickly complete inpainting, but it is easy to produce errors when dealing with large-scale data loss. The method of introducing higher-order modules such as ResNet and CBAM has better repair performance, but it has higher computational complexity and relatively longer training and testing time.

GAN network can generate images with high fidelity through orthogonalization generation and discriminator training. In the field of inpainting, the GAN network can generate the missing image

DOI: 10.54254/2755-2721/19/20231009

content by learning the feature distribution of the real image. Compared with the traditional U-Net network, the GAN network can better process the Semantic information in image tasks and provide higher image fidelity.On the basis of GAN network, the application of CGAN and DCGAN in the field of inpainting is compared. CGAN can ensure image consistency by adding constraints, but the training complexity is high. DCGAN extracts advanced features of images through convolutional neural networks, resulting in high-quality repaired images that need a lot of data to be trained on.

The Context Encoder network can repair missing image parts by learning the contextual information of image content. This method can reconstruct the overall structure of the image while also maintaining better image details. However, there are still issues with image block stretching in the Context Encoder network, which need to be further addressed through some methods. Compared to the three methods mentioned earlier based on Context encoder, edge aware context decoder and full convolutional network can improve the accuracy and accuracy of image restoration, but the computational speed is slower. The residual block structure replacement convolution layer can quickly realize inpainting, but it has high requirements for the design of the network. Decomposing the network architecture into three parallel branch networks has good repair performance and computational speed, but the complexity of network adjustment is relatively high.

Multivariate inpainting method and unit inpainting method have their own advantages and disadvantages. Multivariate inpainting can use multiple samples to improve the accuracy and quality of inpainting, but it requires a lot of data and computing resources. The unit inpainting method is fast in calculation and can handle small-scale data loss, but there may be some problems for large-scale data loss.

## 7. Conclusion

In this essay, we summarize the methods of inpainting based on deep learning. We introduce the unit inpainting methods, including U-Net, GAN, Context Encoder, multivariate inpainting methods and the improved methods based on them, and compare their advantages and disadvantages from the theoretical and practical perspectives.

U-Net based inpainting method is excellent in pixel level image inpainting, but in some cases it is easy to have problems such as unsmooth image reconstruction and unnatural texture. The GAN network can better process Semantic information in image tasks and provide higher image fidelity, but its learning ability for constraints is relatively weak. The Context Encoder network can reconstruct the overall structure of an image while maintaining better image details, but it still needs to address issues such as image block stretching. In addition to the above methods, we also introduced CGAN, DCGAN, and other deep learning methods. These methods have special applications in different fields and scenarios, such as edge aware context decoders and full convolutional networks for repairing edge information of lost images. In addition, we also introduced some methods that combine deep learning models with more traditional algorithms.

Given the widespread usage of deep learning algorithms in the field of inpainting, we think that as technology advances, the effectiveness and performance of these techniques will also advance. In practical applications, we need to choose the appropriate method for inpainting according to the specific situation.

## References

[1] Pathak D, Kr Å Henb, etc. Context Encoder: Feature Learning Through Repair, 2016 IEEE Conf. Comp. Vis. Patt. Recog., 2536-2544.

[2] Rumelhart D E, Hinton G E, Willims R J. Learning internal representations through error propagation. Univ. California, San Diego, 1985.

[3] Liao L, Hu R, Xiao J, etc. Edge Aware Context Encoder for inpainting. 2018 Conf. Aco., Sp. Sig. Proc., 3156-3160.

[4] Yang J, Qi Z, Shi Y. Learn to incorporate structural knowledge into inpainting. 2020, AAAI Artif. Intel. 12605-12612.

[5] Wang Y, Tao X, Qi X, et al. inpainting by generating multi column convolutional neural networks, 2018 Annual Conf. Neur. Infor. Proc. Sys., 329-338.
[6] Vo H V, Duong N K, P É Rez P. Structural Repair, 2018 ACM Inter. Mul. Conf., 1948 1956.
[7] Ji L A region segmentation method based on full resolution attention U-Net neural network, 2023 Rad. Engin. 1-9
[8] Wang W. Repair algorithm based on U-Net network and its application in Yi language restoration Yunnan Univ., 2021.
[9] Liu Liqi, Liu Yanli. Load inpainting: an improved U-Net based load missing data recovery method. 2022 Applied Energy, 327.
[10] Goodfellow I, Mirza M, etc. Generating adversarial networks, 2014 Annual Conf. Neur. Infor. Proc. Sys., 2672-2680.
[11] Mirza M, Osidero S. Conditionally generated adversarial networks. ArXiv preprint arXiv: 1141.17842014.
[12] Dolhansky B, Canton Ferrer C. Using Examples to Generate Eyes for Adversarial Network Painting 2018IEEE Conf. Comp. Vis. Patt. Recog.: 7902-7911.
[13] Radford A, Metz L, Chintala S. Use deep convolution to generate adversarial networks for unsupervised representation learning. ArXiv preprint arXiv: 151106434015.
[14] Han X, Wu Z, Huang W, et al. FiNet: Compatible and diversified fashion inpainting, 2019 Inter. Comp. Vis. Conf., 4480-4490.
[15] Dupont E, Suresha S. Probabilistic Semantic Repair Using Pixel Constrained Cellular Neural Networks, 2019 Inter. Conf. Artif. Intel. Stat., 2261-2270.
[16] Zhao L, Mo Q, Lin S, et al. UCTGAN: Diversified inpainting based on unsupervised cross space translation, 2020, IEEE Conf. Comp. Vis. Patt. Recog. 5740-5749.
[17] Vaswani A, Shazeer N, Parmar N, etc. Attention is what you need. 2017 Annual Conf. Neur. Infor. Proc. Sys., 5998-6008.
[18] Yu Yi, Zhan Fang, Wu Rong, et al. Diversity inpainting based on bidirectional and autoregressive transformations 2021, ACM Inter. Multi. Conf., 69-78.
[19] Wan Z, Zhang J, Chen D, et al. Using Transformers to Achieve High Precision Multivariate Image Mosaic, 2021 Inter. Conf. Comp. Vis.,4672-4681.