Proceedings of the 2nd International Conference on Mathematical Physics and Computational Simulation
DOI: 10.54254/2753-8818/42/2024CH0209

A Review of Pluralistic Image Completion

Sirui Wang

School of Engineering, Xiamen University Malaysia, Sepang, Malaysia

CST2209166@xmu.edu.my

Abstract. Image inpainting has always been a major problem in the field of scientific research. How to fill in the damaged area of the image to make the image look realistic is the main goal of the image inpainting task. In recent years, with the rapid development of machine learning, the researchers started using machine learning to assist in image inpainting. The most representative one is image inpainting based on generative models. The classic idea of restoration is to generate a restoration result with the highest quality and the most realistic appearance. However, as long as the result is reasonable, image inpainting allows for multiple restoration results. For this reason, the field of pluralistic image completion was born. This paper introduces pluralistic image completion and reviews past research in this field. This paper divides pluralistic image completion methods into three categories: VAE-based, GAN-based, and Transformer-based, and gives examples of various representative methods and the latest research in this field. This paper also introduces the available datasets and evaluation metrics. A discussion is given based on the performance of these three categories of methods and the prospects for the development of the field of pluralistic image completion. This review can serve as a reference for researchers in the field of image inpainting. It provides a table of datasets that can be used to evaluate and compare the listed methods and looks forward to possible developments and applications in this field in the future.

Keywords: Deep Learning, Image Inpainting, Image Completion, Multivariate Image Inpainting, Pluralistic Image Completion.

1. Introduction

Image inpainting is a task aiming at filling the missing part of an incomplete image. It asks humans to provide semantically appropriate and visually realistic reconstructing results. It has been a complicated and critical problem and has drawn great attention for a long time. It can be applied to many areas such as object removal, photo editing and image restoration. Its key problem is how to generate missing contents to maintain the integrity and consistency of the restored images and avoid incomplete fusion between the filled contents and the known visible contents. There are several ways to solve this problem.

The traditional methods fix the image by filling the missing area by searching for the most similar image patches and fill them in. Though the traditional methods can work for some cases such as a hole in a large area of lawn, they still face a lot of difficulties when dealing with complex situations and provide semantically correct contents. With the development of technology, deep learning has started to be used in image inpainting area. The deep learning-based image inpainting approaches were classified into three subcategories, which are sequential-based, Convolutional Neural Networks-based(CNN-based), and Generative Adversarial Networks-based(GAN-based) methods[1]. They show good

DOI: 10.54254/2753-8818/42/2024CH0209

authenticity, clarity and semantic rationality in restoration. Context Encoders: Feature Learning by Inpainting is one of the earliest research projects on deep learning-based methods[2]. Pathak et al used a convolutional neural network trained to generate the contents of an arbitrary image region conditioned on its surroundings to realize image inpainting. This method improves image quality and semantic accuracy. After several years, the deep learning-based methods were quickly developed. Generative image inpainting with contextual attention is one of the milestone research projects in deep learning-based image inpainting in recent years[3]. It used the contextual attention technique to solve the problem creating distorted structures or blurry textures that caused by the ineffectiveness of convolutional neural networks in explicitly borrowing or copying information from distant spatial locations. Recently, Transformer has also been applied in the field of image inpainting and achieved good restoration results. Learning Contextual Transformer Network for Image Inpainting is an early attempt to use transformer in image inpainting[4]. It was based on convolutional networks with attention modules. It stacks several transformer blocks to replace convolution layers to better model the long-range dependencies so that it can guarantee the high fidelity and diversity of recovered images.

However, most of the inpainting methods focus on how to generate the best inpainting result. They can only generate one result and ignore other possibilities. Given an incomplete image without additional constraints, image inpainting itself allows multiple solutions as long as it looks reasonable. This is especially true for images with large missing areas. Therefore, some researchers have proposed the research task of pluralistic image completion.

Pluralistic image completion is a new branch of image inpainting tasks. It studies how to generate semantically reasonable and diverse restoration results while generating realistic and clear restoration results. Pluralistic image completion has high application value. It can add a lot of creativity while completing the basic task of image restoration. For example, in the application of photo editing, users can generate restored images of faces with and without glasses at the same time, and users can also generate restoration results of different expressions. As a result, the application prospects are very broad.

The problem that needs to be solved in the pluralistic image completion task is how to generate various reasonable restoration results in the missing area. Since most of the previous deep learning-based image inpainting methods generate unique restoration results, some innovative research is needed to achieve multi-image restoration. In response to this problem, many researchers have realized high-quality pluralistic image completion methods based on different deep learning architectures. Representative ones are methods based on Variational Auto-Encoder(VAE) and methods based on GAN. Since the transformer has made a great splash in the field of image inpainting, in order to solve the potential problems of VAE and GAN in training and generation to improve the restoration quality, transformer has also been used in the research of pluralistic image completion. This article divides Pluralistic image completion into these three categories of methods and introduces representative studies of each category one by one. Then by comparing the experimental data of these studies, the development and prospects of pluralistic image completion are described.

2. Overview of Multivariate Image Technology

2.1. VAE

There are several methods using VAE as the base of them. One of the most prominent representatives is Zheng et al. [5], who first systematically proposed the problem of pluralistic image completion and provided a method. The name of this method is PICNet. They innovatively proposed a probabilistically principled framework and a Dual Pipeline Network Structure to solve the problem of how to generate diverse images while considering image quality. In order to generate a variety of image restoration results, Zheng et al. used Conditional Variational Autoencoder(CVAE) as the basic method, but if only CVAE is used, it is still difficult to generate images with rich diversity. Using a variant of CVAE (with a fixed potential prior), since there is only one training instance for each conditional label in the image completion scene, that is, the ground truth corresponding to the masked image, only one solution can be

DOI: 10.54254/2753-8818/42/2024CH0209

obtained. Therefore, as shown in Figure 1, Zheng et al. introduced the instance blind structure combined with CVAE to ensure the diversity of image restoration.

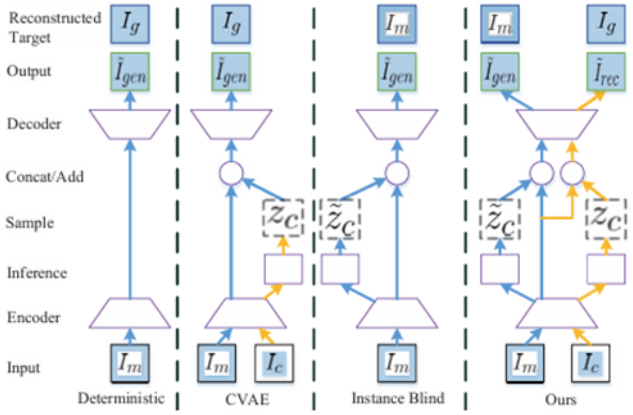

Figure 1. The architecture of CVAE and Instance Blind in PICNet [5]

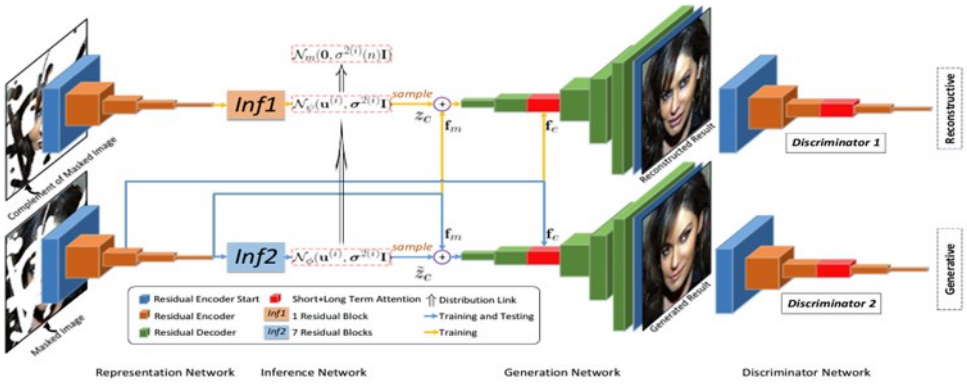

Figure 2. The network structure of PICNet [5]

As shown in Figure 2, this framework constitutes the Dual Pipeline Network Structure. The reconstruction network in the upper part combines the occluded image with the lost area for image reconstruction to train the prior network, so as to obtain a more accurate distribution and generate a variety of semantically consistent restoration results. The lower part is mainly used to train the generation network to generate more realistic and clear repaired images. GAN is applied to both parts of the training to achieve better generation results. At the same time, a short-term + long-term attention mechanism is created, so that the network can choose to focus on the fine-grained features in the encoder or the more semantically generated features in the decoder according to the situation. The PICNet method has excellently solved the problem of generating multiple inpainting results while ensuring high image quality, and has achieved a pioneering pluralistic image completion, achieving a qualitative breakthrough in this field. However, its structure is not stable enough, and the generated results are sometimes not realistic enough. This can be seen from the repair results in Figure 3.

Later on, under the influence of PICNet, many other pluralistic image completion methods were gradually born. Some of the solutions of early VAE-based pluralistic image completion methods may be of low quality due to structural distortion or texture blur, which may be due to the limitations of parameter distribution when facing natural images with complex textures. At the same time, since these VAE-based methods ignore the image structure information, multiple results with limited structural diversity will be generated, that is, posterior collapse. Therefore, Peng et al. proposed a pluralistic image

DOI: 10.54254/2753-8818/42/2024CH0209

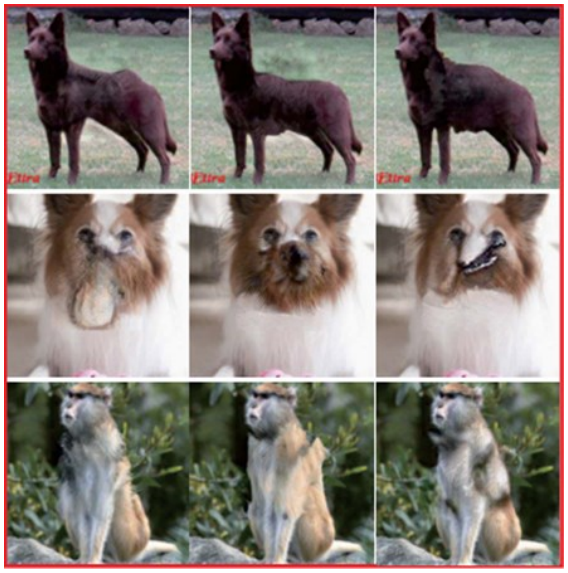

Figure 3. Examples of poor PICNet repair results [5]

completion method based on Hierarchical Vector Quantized Variational Autoencoder(VQ-VAE)[6]. In order to solve the limitations of pluralistic image completion methods, they proposed to use autoregressive modeling of discrete variables and split the structure and texture features for learning and training respectively. Finally, a network structure consisting of three parts is generated. As shown in Figure 4, the hierarchical encoder discretizes the original image structure and texture features, and then the decoder outputs the reconstruction result. Diverse Structure Generator generates multiple structural features when inputting an incomplete image, and Texture Generator uses the incomplete image and structural features to synthesize the image texture and generate the restoration result. Then, the previously pre-trained Hierarchical Encoder is used to assist in training to obtain better restoration quality.

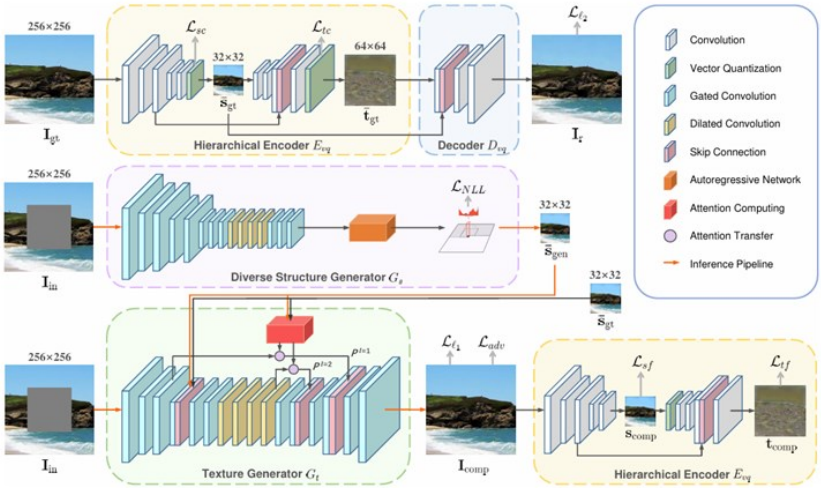

Figure 4. The Hierarchical VQ-VAE network structure [6]

2.2. GAN

In addition to VAE-based methods, some GAN-based pluralistic image inpainting methods have also achieved good results in this field. Some early methods such as BicycleGAN, MUNIT, DR, etc. encourage diverse mappings from potential codes to outputs, so that pictures of diverse styles can be

DOI: 10.54254/2753-8818/42/2024CH0209

generated. At the same time, some GAN-based methods in the field of image generation can also generate diverse pictures by sampling noise from probability distributions and mapping them to image space. However, these methods are usually difficult to achieve good restoration results. First, because each conditional label, that is, masked image, has only one corresponding training instance, it is impossible to obtain diverse generation results. Secondly, since the image restoration scenario requires that the image in the repaired area should be as consistent as possible with the original image in texture and structure, the use of the GAN method is more likely to cause mode collapse.

In order to solve the problems above and realize pluralistic image completion based on GAN, Zhao et al. proposed the UCTGAN method[7]. This method is mainly implemented by two branches and three network modules. The first is a manifold projection module, and then a generation module. These two modules constitute the main branch. By projecting the conditional completion image space and the instance image space into a common low-dimensional manifold space, a one-to-one mapping between images is achieved, thereby improving the variance of the repaired image, preventing the occurrence of mode collapse, and improving the diversity of the restoration. At the same time, a new attention mechanism, cross semantic attention layer, is used to enhance the authenticity and consistency of the repaired image. The second branch consists of a conditional encoding module. It plays the role of conditional constraint and avoids the occurrence of mode collapse. UCTGAN proposes a relatively good solution to the common mode collapse of GAN architecture, which ensures the diversity of restoration results and good image quality.

In addition to UCTGAN, there is another representative method, namely the PD-GAN method proposed by Liu et al. [8]. Before this, pluralistic image completion methods were almost all based on encoder-decoder networks, using the encoder to model the current masked image as a Gaussian distribution and then decode it into a complete image. This makes the distribution change greatly limited by the masked image itself, and the diversity is severely limited. At the same time, over-reliance on reconstruction loss to optimize the encoder and decoder can easily lead to a reduction in restoration diversity. Therefore, they decided to use GAN for generations. Different image contents are generated based on different random noise inputs without sending the input image to CNN. As shown in Figure 5, this method first makes a rough prediction of the restoration result based on the pre-trained partial convolutional network and uses it as prior information. Then a noise vector z is randomly sampled from the normal distribution. After that, this vector z is imported into the SPDNorm residual block created by the paper for debugging. This SPDNorm residual block consists of Hard SPDNorm and Soft SPDNorm. Hard SPDNorm increases the probability of obtaining diversity but reduces the quality of the result. In contrast, Soft SPDNorm can stabilize training and dynamically learn the state of prior information but lacks diversity. Therefore, they are combined to form the SPDNorm residual block. Multiple residual blocks are combined to form a single decoder similar to vanilla GAN, decoding the latent vector z into the image space. PD-GAN successfully implements a pluralistic image completion method based on GAN and improves the diversity of restoration results.

DOI: 10.54254/2753-8818/42/2024CH0209

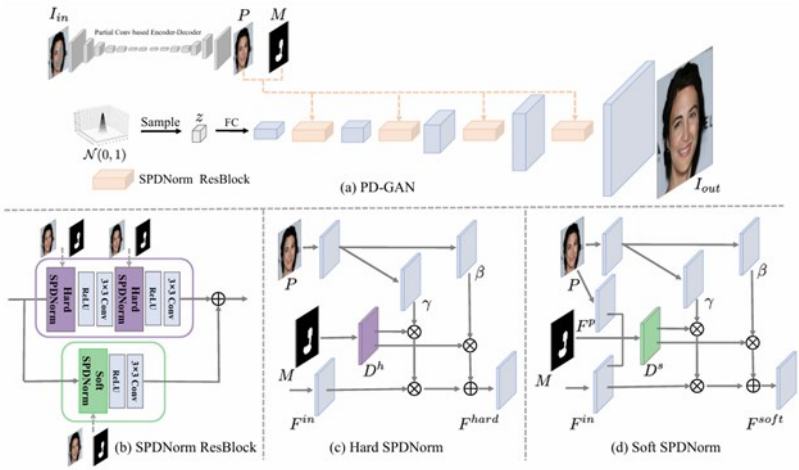

Figure 5. The network structure of PD-GAN and SPDNorm residual blocks [8]

2.3. Transformer

After CNN gradually dominated the field of pluralistic image completion, the defects of CNN model itself in completing pluralistic image completion tasks were gradually exposed: it could not understand the global structure well; Because the same convolution kernel operates on the features of all positions, repeated patterns or artifacts often appear in the restoration results, especially in the restoration of large-scale missing images. In the meanwhile, it is difficult to control the balance between image quality and diversity when using CNN model for pluralistic image completion. In order to achieve diversity in results, image quality will be significantly affected.

Transformer can solve this kind of problem very well. It can capture the global structure well and generate realistic and well-structured images. It can naturally support multi-dimensional output by directly optimizing the underlying data distribution. However, transformer itself is inefficient and has difficulty handling image restoration tasks with arbitrary shape missing.

Therefore, Wan et al. proposed the ICT method[9]. It combines the global structure understanding ability of transformer and the efficient local texture refinement ability of CNN to generate high-fidelity and diverse restoration results. The ICT method consists of two networks, the upper and lower networks. The upper network is a bidirectional transformer. This architecture can output the probability distribution of the missing area at the end. Then it samples multiple times from this distribution to reconstruct the low-resolution image and use it as the appearance prior. Subsequently, CNN is used to upsample the appearance prior in the lower network under the guidance of the masked image. This method combines the advantages of transformer and CNN well, and achieves the generation of higher fidelity, higher quality, and higher diversity restoration results than using CNN alone.

However, the restoration based on ICT, generating low-resolution images through the transformer architecture, and quantizing RGB pixels to improve computational efficiency are prone to information loss. To solve this problem, Liu et al. designed the PUT method[10]. This method combines the Transformer with the Patch-based Vector Quantized Variational Auto-Encoder(P-VQVAE) architecture. The basic network structure of PUT is similar to ICT, except that the encoder part has become a design based on P-VQVAE.P-VQVAE can avoid information loss during downsampling while ensuring computational efficiency, thereby retaining more information for reconstruction. This method optimizes transformer-based methods such as ICT, allowing such methods to generate more realistic and clear images.

Recently, in the field of pluralistic image completion, Zheng et al., the team that originally proposed PICNet, also developed a new method based on the strong potential of the transformer architecture in the field of image restoration[11]. Although the transformer-based method has achieved great success, it still has problems such as image quality degradation and long inference time. After the globally visible

DOI: 10.54254/2753-8818/42/2024CH0209

pixels are degraded by arbitrary masks, how to correctly bridge and utilize the globally visible pixels is also a problem. To solve these problems, they designed the PICFormer method. The main implementation steps are divided into three steps. First, in order to embed the image into a discrete vector, an Encoder-Decoder network and a learnable codebook, that is, a quantizer, are learned. Then a weighted bidirectional transformer is applied to infer the composition of the original embedding index. Finally, an optimization network is used to optimize the global appearance and improve the resolution. This method achieves high-quality and high-diversity generation results at a faster inference speed.

3. Experiments results and Results analysis

3.1. Experiment dataset

The commonly used training and experimental databases for pluralistic image completion are the same as those for regular image inpainting. They all use several of the most common large-scale public databases, including Paris, CelebA-HQ, Places2, ImageNet, and FFHQ. These are also the experimental databases used in the research cases listed in Table 1.

Table 1. Experiment dataset

Dataset	Introduction	Scale	Main type of pictures
Paris	a set of graphs collected from Google Street View	contain ing over 15000 street images of several cities	nature
CelebA-HQ	a high-quality, high-resolution facial image dataset with related attribute labels	consists of 30,000 images at 1024×1024 resolution	faces
Places2	a Large-Scale Database for Scene Understanding	consists over 100,000 high-quality natural images	nature
ImageNet	a Large-Scale image database organized according to the WordNet hierarchy	contains over 14 million images divided into 1000 types	nature
FFHQ	a high-quality, high-resolution facial image dataset	consists of 70,000 high-quality PNG images at 1024×1024 resolution	faces

3.2. Method of evaluation

The evaluation indicators used in multivariate image restoration are basically the same as those used in conventional image restoration, including the common Peak signal-to-noise ratio (PSNR), Structural similarity index measure (SSIM), Learned Perceptual Image Patch Similarity (LPIPS), and Fréchet inception distance (FID), but the standards of some indicators may be slightly different.

PSNR is a ratio between the maximum possible value (power) of a signal and the power of distorting noise that affects the quality of its representation. the higher the PSNR, the better degraded image has been reconstructed to match the original image and the better the reconstructive algorithm.

SSIM is an index that evaluates the structural similarity of two images by integrating image contrast, structural difference, and brightness

LPIPS evaluate the distance between image patches. Higher means further/more different. Lower means more similar. In Pluralistic image completion task, this paper uses LPIPS to measure the diversity of the methods. Higher LPIPS means higher diversity.

FID is a metric for quantifying the realism and diversity of images generated by GANs. The lower FID is, the better the inpainting results are.

DOI: 10.54254/2753-8818/42/2024CH0209

3.3. Result analysis

Table 2. The results in CelebA-HQ dataset

Method	Mask Ratio	Architecture	Dataset	Results			
				CelebA-HQ			
				SSIM	PSNR	LPIPS	FID
UCTGAN (2020)	/	GANs	Paris, CelebA-HQ, Places2, and ImageNet	0.8862	26.38	0.03	
Hierarchical VQ-VAE (2021)	/	VQ-VAE	CelebA-HQ, Places2, and ImageNet	0.8675	24.56		9.784

Table 3. The results in Imagenet dataset

Method	Mask Ratio	Architecture	Dataset	Results			
				ImageNet			
				SSIM	PSNR	LPIPS	FID
PICNet (2019)		CVAE	Paris, CelebA-HQ, Places2, and ImageNet	/	20.1	0.029	/
ICT (2021)	random	Transformer	FFHQ, Places2 and ImageNet	0.835	23.775	/	35.842
	20%-40%			0.888	24.757	/	28.818
	40%-60%			0.721	20.135	/	59.486
PUT(2022)	10%-60%	Transformer	FFHQ, Places2 and ImageNet	0.818	23.264	/	27.648
	20%-40%			0.875	24.238	/	21.272
	40%-60%			0.704	19.742	/	45.153

Table 4. The results in FFHQ and Place2 dataset

Method	Mask Ratio	Architecture	Dataset	Results							
				FFHQ				Places2			
				SSIM	PSNR	LPIPS	FID	SSIM	PSNR	LPIPS	FID
PD-GAN (2021)	10%-20%	GANs	Paris StreetView, CelebA-HQ, and Places2	/	/	/	/	0.935	29.2	/	19.98
	20%-30%			/	/	/	/	0.88	26.75	/	34.84
	30%-40%			/	/	/	/	0.839	25.48	0.1238	44.24
	40%-50%			/	/	/	/	0.782	23.15	/	52.68
ICT (2021)	random	Transformer	FFHQ, Places2 and ImageNet	0.91	26.68	/	14.529	0.832	25.788	/	25.42
	20%-40%			0.948	27.92	/	10.995	0.88	26.503	/	21.598
	40%-60%			0.845	22.61	/	20.024	0.724	22.215	/	33.853
PUT(2022)	10%-60%	Transformer	FFHQ, Places2 and ImageNet	0.906	25.94	/	14.554	0.806	24.492	/	22.121
	20%-40%			0.936	26.88	/	12.784	0.861	25.452	/	19.617
	40%-60%			0.845	22.38	/	21.382	0.703	21.528	/	31.485
PICFormer (2024)	20%-30%	Transformer	FFHQ, ImageNet, and Places2	/	/	/	/	0.8631	24.26	0.0854	15.72
	30%-40%			/	/	/	/	0.8631	22.08	0.1249	15.46
	40%-50%			/	/	/	/	0.7329	20.63	0.1662	15.83

DOI: 10.54254/2753-8818/42/2024CH0209

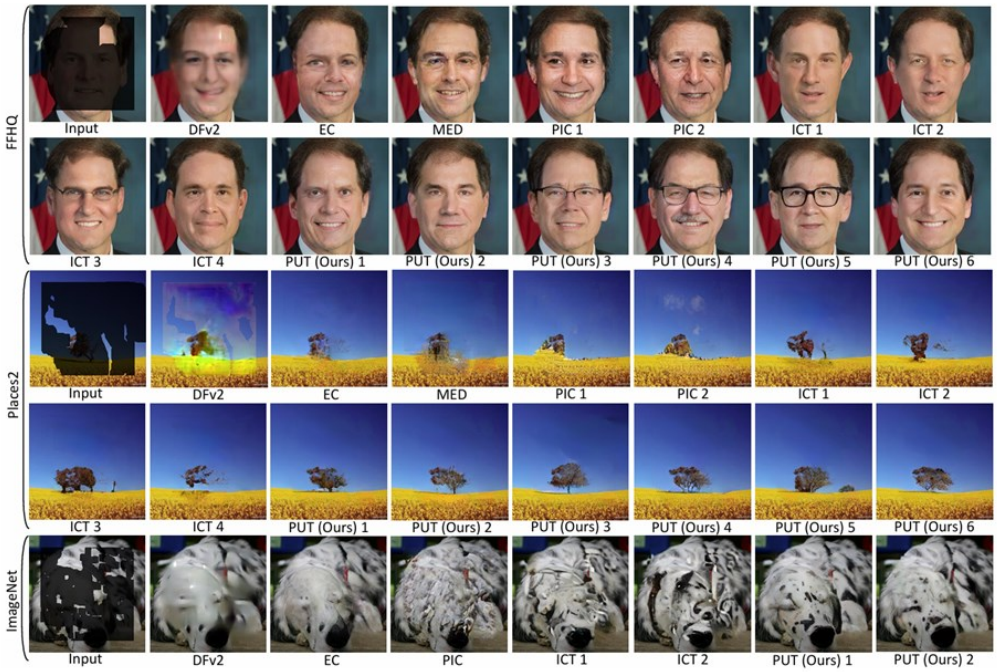

Figure 6. Demonstration of some multi-image restoration methods referenced from [10]

Based on the development of image inpainting, pluralistic image completion has shown excellent effects and potential in the diversity of restoration and has relatively good restoration performance. Figure 6 shows some restoration examples of pluralistic image completion methods. It can be easily seen that when repairing a large range of missing parts, multivariate image restoration shows high diversity and good image quality and authenticity. The data in the table compares the outstanding multi-image restoration methods in recent years. In Table 2, UCTGAN and VQ-VAE both show high image restoration quality and good diversity. In the comprehensive performance of the experimental results in Table 3 and Table 4, ICT, PUT and PICFormer are basically equivalent in terms of PSNR and SSIM, all at the highest level. The lowest one is PICNet. The gap in image quality and diversity between them is obvious, proving that the algorithm improvements in recent years are very effective. Over time, new research has an increasing trend in both PSNR and SSIM, while LPIPS has a decreasing trend. This fully demonstrates that the pluralistic image completion algorithm is gradually improving over time. The PSNR and SSIM of the three most recent methods: ICT, PUT, PICFormer are basically the same, but compared with VAE, GAN based algorithms have been greatly improved, which shows that the introduction of transformers is very meaningful for improving the image restoration quality and semantic rationality of pluralistic image completion methods. In the meanwhile, it can be seen from LPIPS that the diversity of images is gradually increasing based on ensuring semantic rationality and image authenticity. With the increase of mask rate, the diversity of images is also significantly improved.

However, it is worth noting that all pluralistic image completion methods are seriously affected by the mask ratio. When the mask ratio is below 40%, the restoration effect is generally good. As the amount of information lost in the image increases, the image restoration quality of these methods has a significant decline. Relatively speaking, most methods can achieve satisfactory results in the restoration of faces and general natural environments, but in the restoration of scenes with more complex textures such as ImageNet, the restoration effect still can be improved a lot.

4. Discussion

In summary, pluralistic image completion has been developed in depth and has made significant progress and good results. However, the current methods in this field generally have room of

DOI: 10.54254/2753-8818/42/2024CH0209

improvement when facing complex environment images, images with complex textures, and images with large defects. The algorithm can be improved to achieve further improvement. Researchers can continue to study the further application of transformer in the field of image restoration and improve it to achieve higher image quality and diversity of restoration results. At the same time, as the application of diffusion model in the field of image inpainting has gradually begun to be promoted in recent years, it can show good restoration quality and realism on images with large holes. The application of diffusion model in the field of pluralistic image completion deserves further exploration. Pluralistic image completion also has great application potential and can be applied to various scenes with the need to fill image holes. At present, there have been some related application, such as generating diverse facial expressions (EC-GAN) when restoring faces and can also be used to generate diverse facial features and environmental features for people to choose a restoration effect they want most [12]. In addition, researchers can also study how to apply it in further customized scenes, such as how to use text customized features in the occluded area of the picture to generate a variety of realistic restoration results that conform to the text and the image semantics. For example, you can cover the eyes and generate eyes of a specific pupil color or glasses of a specific color through text requirements. In this application scenario, users will have extremely high customization capabilities and can also generate more accurate restoration results that better meet their needs. It can be seen that pluralistic image completion is a field of image restoration with great development potential. It is worth further research in the future.

5. Conclusion

This paper first introduces the research field of image inpainting. Through past studies, the research purpose, content and significance of image inpainting are introduced. After that, by focusing on the limitations of classic image inpainting methods based on generative models, namely the problem of low image diversity, the field of multivariate image restoration is proposed to improve the diversity of restoration results. This paper divides multivariate image restoration methods into three categories based on different deep learning architectures, namely, VAE-based, GAN-based, and Transformer-based methods. By listing the most representative studies in each method, the past research methods, research progress and research results of pluralistic image completion are explained. Among them, the research based on VAE is PICNet and Hierarchical VQ-VAE; the research based on GAN is UCTGAN and PD-GAN; the research based on Transformer is ICT, PUT and PICFormer. By summarizing and comparing the experimental results of these studies, this paper summarizes the advantages of pluralistic image inpainting in generating multiple reasonable results, as well as aspects that need to be strengthened, such as restoration quality. Finally, based on the experimental results, some possible future development directions are proposed. For example, the application of diffusion model can be studied to achieve pluralistic image completion. At the same time, possible application areas of pluralistic image completion are proposed, such as combining it with text semantic recognition for application scenarios.

References

[1] Elharrouss, O., Almaadeed, N., Al-Maadeed, S., & Akbari, Y. (2020). Image inpainting: A review. Neural Processing Letters, 51, 2007-2028.

[2] Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T., & Efros, A. A. (2016). Context encoders: Feature learning by inpainting. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2536-2544).

[3] Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X., & Huang, T. S. (2018). Generative image inpainting with contextual attention. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 5505-5514).

[4] Deng, Y., Hui, S., Zhou, S., Meng, D., & Wang, J. (2021, October). Learning contextual transformer network for image inpainting. In Proceedings of the 29th ACM international conference on multimedia (pp. 2529-2538).

[5] Zheng, C., Cham, T. J., & Cai, J. (2019). Pluralistic image completion. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 1438-1447).

DOI: 10.54254/2753-8818/42/2024CH0209

[6] Peng, J., Liu, D., Xu, S., & Li, H. (2021). Generating diverse structure for image inpainting with hierarchical VQ-VAE. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 10775-10784).

[7] Zhao, L., Mo, Q., Lin, S., Wang, Z., Zuo, Z., Chen, H., ... & Lu, D. (2020). Uctgan: Diverse image inpainting based on unsupervised cross-space translation. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 5741-5750).

[8] Liu, H., Wan, Z., Huang, W., Song, Y., Han, X., & Liao, J. (2021). Pd-gan: Probabilistic diverse gan for image inpainting. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 9371-9381).

[9] Wan, Z., Zhang, J., Chen, D., & Liao, J. (2021). High-fidelity pluralistic image completion with transformers. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 4692-4701).

[10] Liu, Q., Tan, Z., Chen, D., Chu, Q., Dai, X., Chen, Y., ... & Yu, N. (2022). Reduce information loss in transformers for pluralistic image inpainting. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 11347-11357).

[11] Zheng, C., Song, G., Cham, T. J., Cai, J., Luo, L., & Phung, D. (2024). Bridging Global Context Interactions for High-Fidelity Pluralistic Image Completion. IEEE Transactions on Pattern Analysis and Machine Intelligence.

[12] Chen, Y., Yang, W., Fang, X., & Han, H. (2023). EC-GAN: Emotion-Controllable GAN for Face Image Completion. Applied Sciences, 13(13), 7638.