# A Survey of Face Image Inpainting Based on Deep Learning

Shiqi Su[1], Miao Yang[2], Libo He[3], Xiaofeng Shao[1], Yuxuan Zuo[1], and Zhenping Qiang[1(✉)]

[1] College of Big Data and Intelligent Engineering, Southwest Forestry University, Kunming 650224, China
qzp@swfu.edu.cn
[2] Yunnan Institute of Product Quality Supervision and Inspection, Kunming 650214, China
[3] Information Security College, Yunnan Police College, Kunming 650223, China

**Abstract.** In recent years, deep learning has become the mainstream method of image inpainting. It can not only repair the texture of the image, obtain high-level abstract features of the image, but also recover semantic images such as human faces. Among these methods, attention mechanisms, semantic methods, and progressive networks have become very promising image inpainting models. These models implement end-to-end image inpainting and generate visually reasonable and clear image structure and texture. This paper briefly describes the face inpainting technology and summarizes the existing face image inpainting methods. We try to collect most of the face inpainting methods based on deep learning, divide them into attentional, semantic-based, and progressive inpainting networks, and prorate the methods proposed by researchers in each category in recent years. Then we summarize the dataset proposed by the predecessors and the evaluation index of the algorithm performance. Finally, we summarize the current situation and future development trends of face inpainting.

**Keywords:** Face inpainting · Deep learning · Attention inpainting · Semantic inpainting

## 1 Introduction

Image is one of the common information carriers in all walks of life. Because a large amount of image information is destroyed, editing software that can edit images without leaving traces is not feasible. Therefore, an algorithm or system is needed that can edit images without leaving any trace. Image inpainting is a technical process to infer and restore the damaged or missing area content based on the known content of the image so that the image inpainting meets the needs of human visual perception as closely as possible. It can be extended to face inpainting. With the improvement of image processing tools and the flexibility


M. R. Khosravi et al. (Eds.): CloudComp 2021, LNICST 430, pp. 72–87, 2022.
https://doi.org/10.1007/978-3-030-99191-3_7

of digital image editing, automatic image inpainting has become an important application in computer vision and an important stimulating research topic in the field of image processing.

Nowadays, image inpainting has become an active research direction in the field of computer technology, and face inpainting, as one of its branches, is of great significance. Face inpainting is to restore damaged or occluded incomplete face images, but it is difficult to grasp the semantic structure. With the rapid development of image inpainting technology, the problems to be solved are becoming more and more complex. In real life, the lack of face image has caused major work-related problems in all walks of life, and face image inpainting technology can solve this problem intelligently and effectively. It can be applied to many practical applications related to face and has important research value. Due to different postures, expressions and occlusion, face inpainting is a difficult task. Therefore, a good inpainting algorithm should ensure the authenticity of the output, including the topology between eyes, nose, and mouth, as well as the consistency of posture, gender, race, and expression.

We summarize the face image inpainting method based on deep learning. The rest of the paper is organized as follows: Sect. 2 presents the review of the literature, including attention-based, semantic-based, and progressive inpainting methods. Section 3 presents datasets and scoring metrics commonly used in face image inpainting. Section 4 summarizes the current status and future development trends of face image inpainting.

## 2 Face Inpainting Method Based on Deep Learning

With the development of technology, the emergence of deep learning technologies, such as generative adversarial networks (GAN) [1]and convolutional neural networks (CNN) [2], has accelerated the development of the technology face inpainting. These deep learning-based image inpainting methods can already learn rich face semantic information from huge datasets, then fill in the missing information in the image end-to-end, and can achieve better effects. The key to facial inpainting is to maintain the correctness of the facial structure and the rationality of the detailed texture after the inpainting. The traditional image inpainting methods [3–5] did not have the ability to capture high-level semantics and consider the face as a whole, even more, are not suitable for the completion of large-area face images, so they cannot restore the facial image.

Although CNN can capture the abstract information of the image, and GAN can use supervised learning to strengthen the effect of generating the network, the results of these methods for face inpainting alone are not very satisfactory, most researchers now combine the two to inpainting images. In addition, there is also Shift-Net [6] proposed based on texture and CNN; since U-net [7] can use a few images for end-to-end training, some researchers have also proposed many inpainting methods based on this. As demand continues to increase, some researchers have proposed semantic inpainting methods. Pathak et al. [8] proposed the Encoder-Decoder network structure, although the context encoder can

capture the semantics of appearance and visual structure, the context encoder is used for semantic inpainting, the result is not ideal; the literature [9] had added the global context discriminator and the local context discriminator to the literature [8] to discriminate the consistency of the generation effect from both the global and local point of view, and can improve clarity and the contrast of the local area, and proposed a full convolution for image inpainting. And there are kinds of literature [10,11] that use this type of method. Although it can use full convolution to restore free template images of any resolution, the inpainting effect is not ideal when inpainting images with very complex semantic information.

To make full use of the mask information, different researchers have proposed different convolution methods [12–16]. Jo et al. [12] proposed an encoder-decoder architecture similar to U-net. All convolution layers are gated convolution networks, which enable an image editing system. It is a system with a free-form mask, sketches, and color as inputs. Liu et al. [16] used partial convolution, where convolution was limited to valid pixels to reduce artifacts caused by differences in the distribution between masked and uncovered areas, but this method can create damaged structures when the missing areas became continuous and relatively large.

In addition, image inpainting in the practical application includes free-form or irregular holes. Compared with regular holes, these holes needed different optimization processes or attention mechanisms [17], so an attention-based face image inpainting method was introduced; there are also methods based on semantic; furthermore, some also proposed to divide the task of image inpainting into several subproblems for progressive inpainting. The multi-stage network model generally has higher efficiency. These methods are introduced in turn below.

### 2.1 Attention-Based Image Inpainting

In the field of computer vision, an attention mechanism is introduced to process visual information. It looks for the most important part of result generation and improves the performance of segmentation, re-recognition, and tracking algorithms. Attention mechanism is a technology that enables the model to focus on important information and make full use of it. Most of the research work on the combination of deep learning and visual attention mechanism focuses on using a mask to form the attention mechanism. The principle of the mask is to identify the key features in the picture data through another layer of new weight. Through learning and training, the deep neural network can learn the areas that need attention in the picture, which forms attention.

In recent years, attention based on the relationship between context and mask is often used in image inpainting tasks [18–32]. Yu et al. [20] innovatively added a context-aware module to their coarse-to-fine architecture, which focused on relevant feature patches at any location to improve the inpainting results. Literature [21] used partial convolution instead of vanilla convolution on the basis of [20]. Xie et al. [19] designed a two-way attention map estimation module for feature

renormalization and mask update in the feature generation process. He et al. [22] proposed an image inpainting model based on the inside-outside attention layer (IOA). IOA can generate images with free-form masks while maintaining high contextual semantic consistency and visual reality. Liu et al. [30] used a coherent layer of semantic attention in a refined network to ensure semantic correlation between exchange features. Wang et al. [33] developed a multi-scale attention module in the architecture to make flexible use of background content. The Pluralistic image completion (PIC) method [34] employed a self-attention layer that uses short-term and long-term context information to ensure a consistent appearance. Zeng et al. [35] applied an attention mechanism to build an attention delivery network that used advanced semantic information to inpainting low-level image features. These image inpainting algorithms demonstrate the effectiveness of the attention mechanism.

The attention of each round of the traditional attention mechanism is calculated independently of each other and will interfere with each other during fusion. The convolutional neural network can not explicitly borrow or copy information from distant spatial positions, which is the reason for image structure distortion and texture blur. Although the context attention layer can improve the performance compared with the traditional convolution, and the facial image inpainting model uses the attention mechanism to borrow features from the background, the inpainting results still lack fine texture details, and the pixels are inconsistent with the background.

### 2.2 Semantic-Based Image Inpainting

The attention-based image inpainting method obtains information from the background area far away from the mask to propagate to the mask area, but in the process of propagation, it will produce fuzzy results because of the misleading part of the information of the newly recovered mask. Facial images usually contain unique patterns, a few repetitive structures, and the semantic content is specific. When the face image is lost, it is more difficult to complete and restore, which makes face inpainting a challenging problem. Because the face image is highly structured and has several key semantic components, such as eyes and mouth, the semantic information of the face can be used to inpainting it better.

Semantic inpainting needs to fill a large number of missing areas according to known data. Such as [30,33,36–43], it infered the content of any large missing area in the image according to the semantic information of the image. Face inpainting is the most representative type of semantic inpainting. The general context will produce fewer ideal results. For example, the context encoder (CE) proposed by Pathak et al. [8] first used the deep neural network to generate missing regions, and the context encoder fills the loophole by extracting features from the original image. However, the disadvantage of this method is that the generated image contains too many visual artifacts. Semantic inpainting is not an attempt to reconstruct real images, but to fill this loophole with realistic content. Raymond et al. [37] proposed a new semantic image inpainting method. This method combined weighted semantic loss in the trained generative model

to determine the most similar coding information between the implicit space and the missing image and then predicts the missing content through the generative model. This method is superior to the common semantic-based algorithm CE and can generate reasonable and clear edge information, but there are examples of inpainting failures. Many methods use the prior knowledge or strategies in specific fields to solve the corresponding problems, such as the face super-resolution (SR) method, which uses the prior knowledge of the face to better SR inpainting the face. Shen et al. [44] used face semantic tags as global a priori and local constraints to eliminate ambiguity. They used the face analysis network to generate the semantic tags of the fuzzy input image, and then took the fuzzy image and semantic tags as the input to inpainting the image from coarse to fine network. Zhang et al. [39] proposed a new semantic image inpainting method-the squeeze excitation network deep revolution general adaptive network (SE-DCGAN). Zhang et al. [40] proposed a progressive generative network (PGN), which regarded semantic image inpainting as a step-by-step learning process, but its model did not have good inpainting results for free-form or complex images (e.g., face images).

In the above two methods, if semantic and attention are combined, better processing results will be achieved. Liu et al. [30] used a coherent semantic attention layer to ensure the semantic correlation between exchange features, proposed a coarse-fine network, and added the fine inpainting network of coherent semantic attention (CSA) layer. By modeling the semantic correlation between hole features, it can not only maintain the structure of context but also predict the missing parts more effectively. Wang et al. [41] improved the original image inpainting algorithm model [30] based on U-net and Visual Geometry Group Network (VGG) and designed a semantic focus layer to learn the relationship between missing region features in the image inpainting task. Qiu et al. [38] proposed a two-step confrontation model semantic structure reconstructor and texture generator, which used the semantic structure graph based on unsupervised segmentation to train the semantic structure reconstructor and maintain the consistency between the missing part and the whole picture, spatial channel attention module (SCA) was introduced to obtain fine-grained texture.

### 2.3 Progressive-Based Image Inpainting

Although some of the above methods have good results, these methods are still challenging for the task of face image inpainting because the generated image should have a good visual effect and fast processing speed. Optimization-based methods such as [45] can produce inpainting results that were visually considered natural, but its calculation speed was very slow, and it was slower to process images with high resolution; the CE [8] is very fast, but sometimes it cannot restore a satisfactory structure. Therefore, researchers have developed a new method, the progressive-based method [11,13,15,18,20,22,30,40,41,46–54], so as to reduce the difficulty of training depth in the inpainting network.

For example, [13,20,22,30,48,52], these methods are two-stage network structures composed of a coarse-fine network. In the first stage, the image need to be

processed is roughly inpainting, and then the roughly processed image is used as the input of the fine stage, so as to further strengthen the structure and texture details of the face image. Xiong et al. [47] proposed a foreground aware inpainting method, which involved three stages: contour detection, contour completion, and image completion, so as to eliminate structure inference and content illusion. Song et al. [11] introduced additional manual labels in segmentation prediction and guidance Network (SPG-net), but it was often unavailable in practical application, so this method was difficult to be directly used for image restoration. Huang et al. [52] proposed a new semantic aware context aggregation module (SACA) to solve the problem of generating fuzzy content. By using the internal semantic similarity of the input feature graph, the remote context information is aggregated from the semantic point of view. SACA suppresses the influence of misleading hole features in context aggregation by learning the relationship between pixels and semantics and significantly reduces the computational burden. The multi-stage method can alleviate the difficulty of deep maintenance network training.

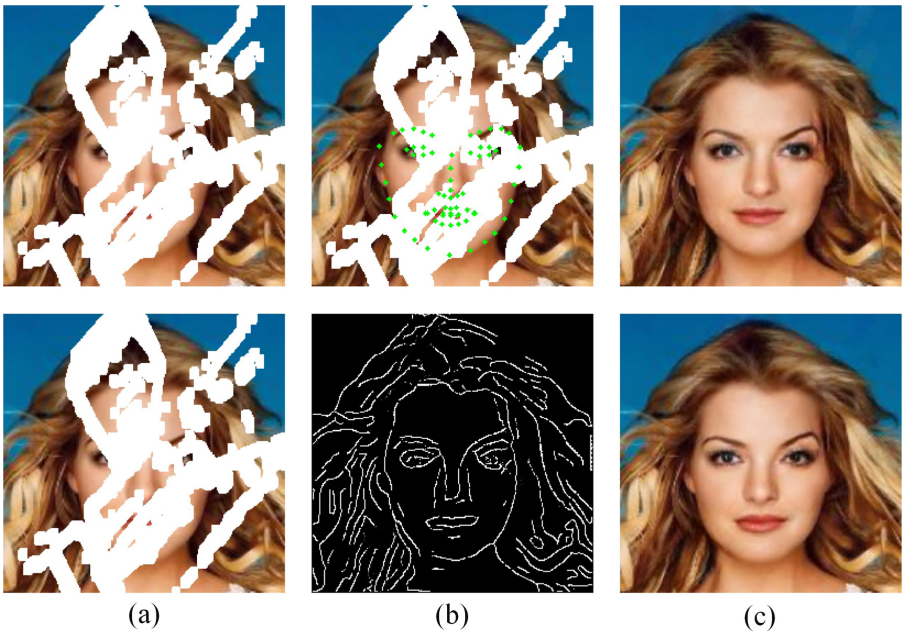


**Fig. 1.** Results of Lafin (upper) and edge connect (lower) algorithms. (a) input image, (b) presents the version of the landmark on the masked image (top), generated edges (bottom), (c) generated result.

Nazeri et al. [15] proposed a two-stage confrontation model called edgeconnect, which first predicted the edge of the missing area, and then generated the edge-guided inpainting results. However, the edge is not an ideal semantic structure because it loses a lot of region information and color information. Yang et al. [55] introduced a generator called landmark guided face painter (Latin), which was composed of face landmark prediction subnet and image inpainting subnet to solve the problem of face inpainting. The face landmark prediction subnet module reflects the topology, pose, and expression of the target face to be restored. The image inpainting subnet uses the predicted landmark as a guide, and uses the spatial context to connect the temporal feature mapping to ensure the consistency of attributes. Table 1 is a brief comparison of edgeconnect, lafin method in terms of network structure, generator, and input images required for image processing. Figure 1 shows the results of edgeconnect and Lafin algorithms.

**Table 1.** Table comparison between Edgeconnect [15] and Lafin [55] methods.

| | Edgeconnect | Lafin |
|---|---|---|
| Networks | Edge generator, image completion network | Landmark prediction module, image inpainting module |
| Input | Mask, Edge map, Grayscale | Corrupted image, landmarks |
| Convolution | Dilated conv | Dilated conv, gated conv |
| Defect | The edge generator model cannot accurately depict edges in highly textured areas | The inpainting results after missing many central areas are not so ideal, so are landmarks |
| Advantage | The prior information with high correlation and low generation difficulty is selected as the prior information of the next stage. The edge information restored by the algorithm is accurate and will not appear as false content | Face key points are neat, sufficient, and robust, which can be used as the supervision of face inpainting. It is a simple and a reliable way of data expansion |

## 3 Datasets and Evaluation Indicators

### 3.1 Dataset

For face image inpainting, researchers have proposed many public datasets and large datasets to evaluate the applicability of their algorithms. Face image inpainting also belongs to a part of image inpainting. Adding images such as the natural and street view in the training process will improve the inpainting result. Table 2 introduces some datasets used by predecessors, and Fig. 2 shows sample images of common datasets.

**Table 2.** Datasets introduction.

| Dataset | Year | Number | Attribute | Affiliated |
|---|---|---|---|---|
| CAS-PEAL [56] | 2008 | 99,594 | Different postures in a specific environment | Institute of computing technology, Chinese Academy of Sciences |
| ImageNet [57] | 2009 | 14,197,122 | 21841 categories | Stanford University Vision Research Laboratory |
| Helen Face [58] | 2012 | 2330 | All images are marked with 68 feature points | University of Illinois, Urbana-Champaign and Adobe Systems Inc |
| CASIA-WebFace [59] | 2014 | 494,414 | 10575 people | National Laboratory of Pattern Recognition; institute of Automation, Chinese Academy of Sciences |
| Places2 [60] | 2017 | 10 million+ | Including more than 400 unique scene categories | Massachusetts Institute of Technology |
| CelebA [61] | 2018 | 202,599 | Each image is marked with features | Chinese University of Hong Kong |

### 3.2 Evaluating Indicator

Generally, in face inpainting, the evaluation index is usually used to objectively evaluate the advantages and disadvantages of image processing algorithms, and highlight the advantages of the algorithms propose by researchers in comparative experiments. Generally speaking, each subdivided field has corresponding indicators. In image inpainting, the evaluation indicators are basically universal and can be evaluated by comparing the image inpainting with the real image, that is, there are reference image evaluation indicators. Accordingly, a series of indicators are proposed: mean square error (MSE), structural similarity index (SSIM), peak signal to noise ratio (PSNR), etc. these indicators are common. PSNR and SSIM are mostly used in image inpainting, so the results are more convincing.

**Structural Similarity Index (SSIM).** SSIM [62] is a comprehensive reference image quality assessment index proposed by the University of Texas at Austin, it is used to estimate the similarity between image inpainting and original image. When the two pictures are exactly the same, the value of SSIM is 1. Structural similarity theory holds that natural images are highly structured, that is, there is a strong correlation between pixels, which contains important information of object structure in the visual scenes. The mean is used as the estimate of brightness, the standard deviation as the estimate of contrast, and the covariance

as the measure of structural similarity. Given two images x and y, the structural similarity of the two images is expressed as:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{1}$$

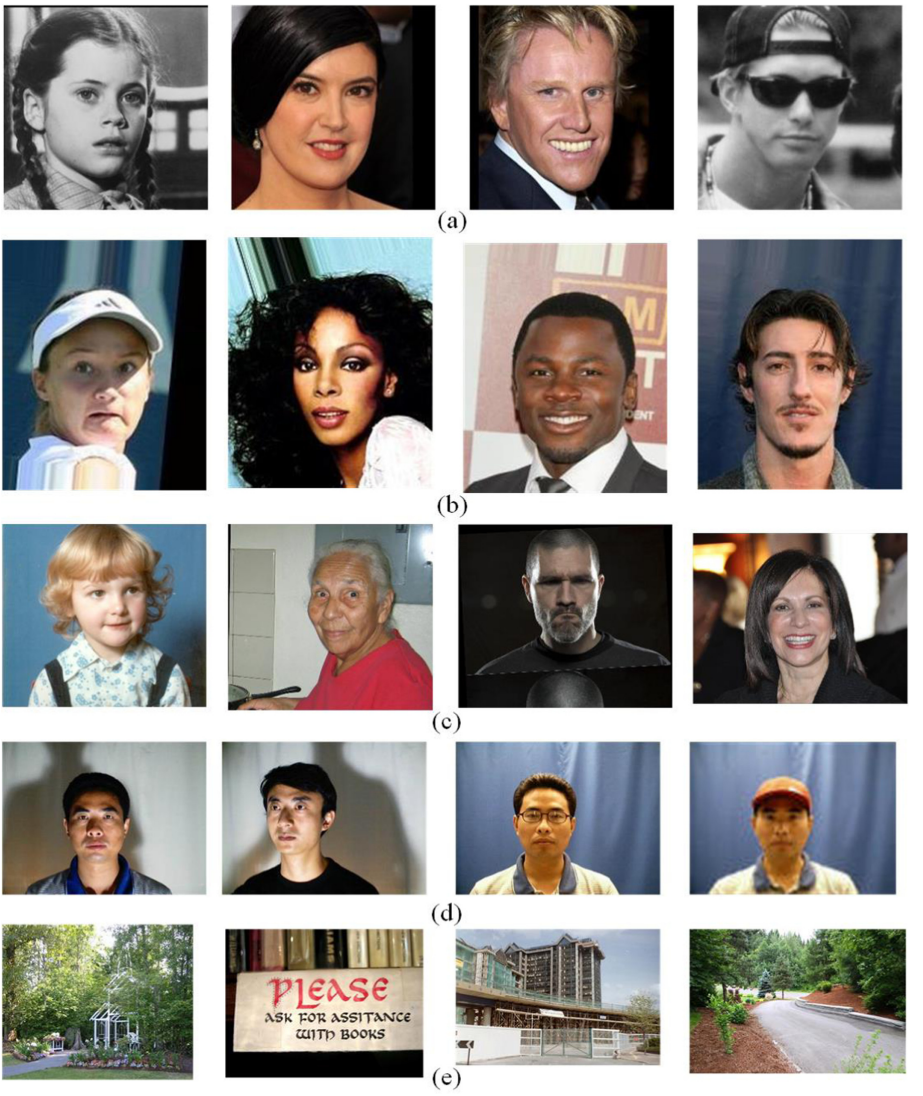


**Fig. 2.** Common dataset samples. From top to bottom are CASCIA dataset (a), CelebA dataset (b), Helen Face dataset (c), CAS-PEAL dataset (d), Place2 dataset (e).

here $\mu_x$, and $\mu_y$ are the mean values of x and y respectively, $\sigma_x^2$ and $\sigma_y^2$ are the variances of x and y respectively, $\sigma_{xy}$ is the covariance of x and y; $c_1 = (k_1 L)^2, c_2 = (k_2 L)^2$ are two constants, in case the fraction is zero, L is the range of pixel values, $k_1 = 0.01, k_2 = 0.03$ are the default values. The value range of SSIM is 0–1.

**Peak Signal to Noise Ratio (PSNR)).** PSNR [62] is the most widely used objective standard for evaluating images. It is generally used to evaluate the quality of the repaired face image compared with the real face image. The higher the PSNR, the smaller the distortion after compression. PSNR is widely used, but its value cannot well reflect the subjective feeling of human eyes. General value range: 20–40. The larger the value, the better the video quality. Usually, after image compression, the output image will be different from the original image to some extent. PSNR performs statistical analysis based on the gray value of image pixels. Due to the differences in human visual characteristics, the evaluation results are usually inconsistent with people's main feelings, but it is still a valuable evaluation index. In order to measure the quality of the processed face image, we usually refer to the PSNR value to measure whether a processing program is satisfactory. PSNR can be simply defined by mean square error MSE. Given an original image I with a size of $m \times n$ and a processed face image K, the mean square error is defined as:

$$MSE = \frac{1}{mn} \Sigma_{j=0}^{m-1} \Sigma_{j=0}^{n-1} [I(i,j) - K(i,j)]^2 \tag{2}$$

among them: $I(i,j), K(i,j)$ respectively represent the pixel value at the corresponding coordinate, and m and n are the height and width of the image respectively. The formula of PSNR is as follows:

$$PSNR = 10lg(\frac{MAX_I^2}{MSE}) \tag{3}$$

among them: $MAX_I^2$ is the maximum value of the picture color.

The evaluation results of some inpainting methods on common datasets are shown in Table 3. It can be seen from Table 3 that the values of PSNR and SSIM based on the progressive face inpainting method are relatively stable, and there is no low score, and SSIM almost exceeds 90%.

**Table 3.** Table lists the quantitative evaluation results of the above-mentioned various image inpainting algorithms based on deep learning, the values of which are from various original documents. "+" means that the larger the value of this index, the better.

| Category | Method | Dataset | Mask | PSNR$^+$ | SSIM$^+$ |
|---|---|---|---|---|---|
| Attention-based | [18] | Place2 | Center mask | 26.17 | 0.91 |
| | [19] | Place2 | Mask ratio = $(0.2, 0.3]$ | 25.59 | 0.785 |
| | [20] | Place2 | Rule mask | 18.91 | – |
| | [21] | CelebA | – | **40.86** | – |
| | [22] | CelebA | Center mask | 26.84 | 0.921 |
| | [23] | CelebA | Center mask | 32.23 | **0.933** |
| | [27] | CelebA | Mask ratio = $(0.3, 0.4]$ | 26.67 | 0.874 |
| | [30] | CelebA | Center mask | 26.54 | 0.931 |
| | [34] | ImageNet | Center mask | 20.10 | – |
| | [25] | Place2 | Random mask | – | 0.781 |
| Semantic-based | [36] | CelebA | Center mask | 19.18 | 0.920 |
| | [37] | CelebA | Random mask | 22.8 | – |
| | [38] | CelebA | Center mask | **32.23** | **0.933** |
| | [40] | CelebA | 50% mask | 19.10 | 0.802 |
| | [44] | Helen Face | Center mask | 21.45 | 0.851 |
| | [52] | CelebA | Mask ratio = $(0.3, 0.4]$ | 27.21 | 0.895 |
| Progressive-based | [15] | CelebA | Mask ratio = $(0.4, 0.5]$ | 25.28 | 0.846 |
| | [22] | CelebA | Center mask | 26.84 | 0.921 |
| | [41] | ImageNet | – | 25.74 | 0.934 |
| | [46] | CelebA | – | 26.60 | 0.920 |
| | [47] | Place2 | – | **29.86** | 0.938 |
| | [51] | CelebA | Mask ratio = $(0.2, 0.3]$ | 29.01 | **0.955** |
| | [53] | Place2 | Mask ratio = $(0.2, 0.3]$ | 25.66 | 0.914 |

## 4 Conclusion

With the continuous development of deep learning technology and urgent application needs, the task of face image inpainting has attracted the attention of researchers from all walks of life and become an important and challenging research topic in the field of computer vision. In this paper, different types of methods are introduced, including attention-based methods, semantic-based methods, and progressive inpainting methods; secondly, the commonly used datasets and performance evaluation indexes of face image inpainting in the existing literature are summarized. Through quantitative evaluation and comparison of inpainting effects, it is proved that the current face disocclusion technology based on deep learning has a good experimental effect.

Based on the classification and summary of the existing image inpainting methods, aiming at the problems still existing in the current research task, this paper makes the following prospects for its future research direction and development trend:

1) The essence of image inpainting is a computer vision task to guess and complete the missing area by mining known information. It is an inevitable demand to improve the inpainting quality to effectively extract the known information and establish the information associated with the unknown content. Improving the learning ability of image feature expression of inpainting model is still one of the problems worthy of in-depth exploration.
2) At present, most of the datasets used in the literature are European and American face datasets, so the test results of Asian faces are not ideal. Therefore, it is necessary to establish a dataset belonging to Asian face features to make the algorithm more consistent with Asian face attributes.
3) Generative adversarial network plays a key role in image generation and is also adopted by most image inpainting methods. However, at present, generative adversarial network still has its own defects, such as mode collapse and unstable training. How to solve these problems will also become a challenge in image inpainting research.

**Acknowledgements.** This work was funded by the National Natural Science Foundation of China (12163004), the basic applied research program of Yunnan Province (202001AT070135, 202101AS070007, 202002AD080002, 2018FB105).

## References

1. Goodfellow, I., et al.: Generative adversarial nets. In: Advances in Neural Information Processing Systems, vol. 27 (2014)
2. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. Adv. Neural. Inf. Process. Syst. **25**, 1097–1105 (2012)
3. Bertalmio, M., Sapiro, G., Caselles, V., Ballester, C.: Image inpainting. In: Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques, pp. 417–424 (2000)
4. Barnes, C., Shechtman, E., Finkelstein, A., Goldman, D.B.: PatchMatch: a randomized correspondence algorithm for structural image editing. ACM Trans. Graph. **28**(3), 24 (2009)
5. Huang, J.B., Kang, S.B., Ahuja, N., Kopf, J.: Image completion using planar structure guidance. ACM Trans. Graph. **33**(4), 1–10 (2014)
6. Yan, Z., Li, X., Li, M., Zuo, W., Shan, S.: Shift-Net: image inpainting via deep feature rearrangement. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) Computer Vision – ECCV 2018. LNCS, vol. 11218, pp. 3–19. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01264-9_1
7. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28

8. Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T., Efros, A.A.: Context encoders: feature learning by inpainting. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2536–2544 (2016)
9. Iizuka, S., Simo-Serra, E., Ishikawa, H.: Globally and locally consistent image completion. ACM Trans. Graph. **36**(4), 1–14 (2017)
10. Song, Y., et al.: Contextual-based image inpainting: infer, match, and translate. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) ECCV 2018. LNCS, vol. 11206, pp. 3–18. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01216-8_1
11. Song, Y., Yang, C., Shen, Y., Wang, P., Huang, Q., Kuo, C.C.J.: SPG-Net: segmentation prediction and guidance network for image inpainting. arXiv preprint arXiv:03356 (2018)
12. Jo, Y., Park, J.: SC-FEGAN: face editing generative adversarial network with user's sketch and color. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 1745–1753 (2019)
13. Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X., Huang, T.S.: Free-form image inpainting with gated convolution. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 4471–4480 (2019)
14. Xiao, Q., Li, G., Chen, Q.: Deep inception generative network for cognitive image inpainting. arXiv preprint arXiv:01458 (2018)
15. Nazeri, K., Ng, E., Joseph, T., Qureshi, F.Z., Ebrahimi, M.: EdgeConnect: generative image inpainting with adversarial edge learning. arXiv preprint arXiv:00212 (2019)
16. Liu, G., Reda, F.A., Shih, K.J., Wang, T.-C., Tao, A., Catanzaro, B.: Image inpainting for irregular holes using partial convolutions. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) ECCV 2018. LNCS, vol. 11215, pp. 89–105. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01252-6_6
17. Vaswani, A., et al.: Attention is all you need. In: Advances in Neural Information Processing Systems, pp. 5998–6008 (2017)
18. Xiao, Z., Li, D.: Generative image inpainting by hybrid contextual attention network. In: Lokoč, J., Patras, I. (eds.) MMM 2021. LNCS, vol. 12572, pp. 162–173. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-67832-6_14
19. Xie, C., et al.: Image inpainting with learnable bidirectional attention maps. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 8858–8867 (2019)
20. Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X., Huang, T.S.: Generative image inpainting with contextual attention. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5505–5514 (2018)
21. Mohite, T.A., Phadke, G.S.: Image inpainting with contextual attention and partial convolution. In: 2020 International Conference on Artificial Intelligence and Signal Processing (AISP), pp. 1–6. IEEE (2020)
22. He, X., Cui, X., Li, Q.J.I.A.: Image inpainting based on inside-outside attention and wavelet decomposition. IEEE Access **8**, 62343–62355 (2020)
23. Qiu, J., Gao, Y.: Position and channel attention for image inpainting by semantic structure. In: 2020 IEEE 32nd International Conference on Tools with Artificial Intelligence (ICTAI), pp. 1290–1295. IEEE (2020)
24. Wu, H., Zhou, J.: IID-Net: image inpainting detection network via neural architecture search and attention. IEEE Trans. Circ. Technol. Syst. Video (2021)
25. Wang, C., Wang, J., Zhu, Q., Yin, B.: Generative image inpainting based on wavelet transform attention model. In: 2020 IEEE International Symposium on Circuits and Systems (ISCAS), pp. 1–5. IEEE (2020)

26. Li, J., Wang, N., Zhang, L., Du, B., Tao, D.: Recurrent feature reasoning for image inpainting. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7760–7768 (2020)
27. Wang, N., Ma, S., Li, J., Zhang, Y., Zhang, L.J.P.R.: Multistage attention network for image inpainting. Pattern Recognit. **106**, 107448 (2020)
28. Huang, L., Wang, W., Chen, J., Wei, X.Y.: Attention on attention for image captioning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 4634–4643 (2019)
29. Song, L., et al.: Unsupervised domain adaptive re-identification: theory and practice. Pattern Recognit. **102**, 107173 (2020)
30. Liu, H., Jiang, B., Xiao, Y., Yang, C.: Coherent semantic attention for image inpainting. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 4170–4179 (2019)
31. Chen, B., Li, P., Sun, C., Wang, D., Yang, G., Lu, H.: Multi attention module for visual tracking. Pattern Recogn. **87**, 80–93 (2019)
32. Uddin, S., Jung, Y.J.: Global and local attention-based free-form image inpainting. Sensors **20**(11), 3204 (2020)
33. Jiao, L., Wu, H., Wang, H., Bie, R.: Multi-scale semantic image inpainting with residual learning and GAN. Neurocomputing **331**, 199–212 (2019)
34. Zheng, C., Cham, T.J., Cai, J.: Pluralistic image completion. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1438–1447 (2019)
35. Zeng, Y., Fu, J., Chao, H., Guo, B.: Learning pyramid-context encoder network for high-quality image inpainting. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1486–1494 (2019)
36. Vitoria, P., Sintes, J., Ballester, C.: Semantic image inpainting through improved Wasserstein generative adversarial networks. arXiv preprint arXiv:01071 (2018)
37. Yeh, R.A., Chen, C., Yian Lim, T., Schwing, A.G., Hasegawa-Johnson, M., Do, M.N.: Semantic image inpainting with deep generative models. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5485–5493
38. Qiu, J., Gao, Y., Shen, M.: Semantic-SCA: semantic structure image inpainting with the spatial-channel attention. IEEE Access **9**, 12997–13008 (2021)
39. Zhang, F., Wang, X., Sun, T., Xu, X.: SE-DCGAN: a new method of semantic image restoration. Cogn. Comput. **13**, 1–11 (2021)
40. Zhang, H., Hu, Z., Luo, C., Zuo, W., Wang, M.: Semantic image inpainting with progressive generative networks. In: Proceedings of the 26th ACM International Conference on Multimedia, pp. 1939–1947 (2018)
41. Wang, W., Gu, E., Fang, W.: An improvement of coherent semantic attention for image inpainting. In: Sun, X., Wang, J., Bertino, E. (eds.) ICAIS 2020. CCIS, vol. 1252, pp. 267–275. Springer, Singapore (2020). https://doi.org/10.1007/978-981-15-8083-3_24
42. Yang, W., Li, X., Zhang, L.: Toward semantic image inpainting: where global context meets local geometry. J. Electron. Imaging **30**(2), 023028 (2021)
43. Ciobanu, S., Ciortuz, L.: Semantic image inpainting via maximum likelihood. In: 2020 22nd International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC), pp. 153–160. IEEE (2020)
44. Shen, Z., Lai, W.S., Xu, T., Kautz, J., Yang, M.H.: Deep semantic face deblurring. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 8260–8269 (2018)

45. Yang, C., Lu, X., Lin, Z., Shechtman, E., Wang, O., Li, H.: High-resolution image inpainting using multi-scale neural patch synthesis. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6721–6729 (2017)
46. Ma, B., An, X., Sun, N.: Face image inpainting algorithm via progressive generation network. In: 2020 IEEE 5th International Conference on Signal and Image Processing (ICSIP), pp. 175–179. IEEE (2020)
47. Xiong, W., et al.: Foreground-aware image inpainting. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5840–5848 (2019)
48. Zeng, Yu., Lin, Z., Yang, J., Zhang, J., Shechtman, E., Lu, H.: High-resolution image inpainting with iterative confidence feedback and guided upsampling. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12364, pp. 1–17. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58529-7_1
49. Yen, S.H., Yeh, H.Y., Chang, H.W.: Progressive completion of a panoramic image. Multimedia Tools Appl. **76**(9), 11603–11620 (2017)
50. Karras, T., Aila, T., Laine, S., Lehtinen, J.: Progressive growing of GANs for improved quality, stability, and variation. arXiv preprint arXiv:10196 (2017)
51. Guo, Z., Chen, Z., Yu, T., Chen, J., Liu, S.: Progressive image inpainting with full-resolution residual network. In: Proceedings of the 27th ACM International Conference on Multimedia, pp. 2496–2504 (2019)
52. Huang, Z., Qin, C., Liu, R., Weng, Z., Zhu, Y.: Semantic-aware context aggregation for image inpainting. In: ICASSP 2021–2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2465–2469. IEEE (2021)
53. Li, J., He, F., Zhang, L., Du, B., Tao, D.: Progressive reconstruction of visual structure for image inpainting. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 5962–5971 (2019)
54. Zamir, S.W., et al.: Multi-stage progressive image restoration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 14821–14831 (2021)
55. Yang, Y., Guo, X., Ma, J., Ma, L., Ling, H.: LAFIN: generative landmark guided face inpainting. arXiv preprint arXiv:11394 (2019)
56. Gao, W., et al.: The CAS-PEAL large-scale Chinese face database and baseline evaluations. IEEE Trans. Syst. Man Syst. Cybernet. Part A Hum. **38**(1), 149–161 (2007)
57. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: ImageNet: a large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255. IEEE (2009)
58. Le, V., Brandt, J., Lin, Z., Bourdev, L., Huang, T.S.: Interactive facial feature localization. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012. LNCS, vol. 7574, pp. 679–692. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-33712-3_49
59. Yi, D., Lei, Z., Liao, S., Li, S.Z.: Learning face representation from scratch. arXiv preprint arXiv (2014)
60. Zhou, B., Lapedriza, A., Khosla, A., Oliva, A., Torralba, A.: A 10 million image database for scene recognition. IEEE Trans. Pattern Anal. Intell. Mach. **40**(6), 1452–1464 (2017)

61. Liu, Z., Luo, P., Wang, X., Tang, X.: Large-scale CelebFaces attributes (CelebA) dataset. Retrieved August **15**, 11 (2018)
62. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004)