



Progressive Image Restoration with Multi-stage Optimization

Jiaming Yang , Weihua Zhang , and Yifei Pu

School of Computing, Sichuan University, Chengdu 610065, Sichuan, China
zhangweihua@scu.edu.cn

Abstract. With the recent development of deep learning technology, researchers have achieved significant results in small-scale image inpainting. However, when the missing area is large, undesirable artifacts and noise are introduced into the inpainting area. Hence, we present a multi-stage progressive image inpainting framework based on the well-known generative adversarial network(GAN) to solve this problem. In our MOPR-GAN method, generator uses a progressive inpainting module(PIM) and an image optimization module(IOM), while discriminator combines a patchGAN with an attention mechanism and a globalGAN. The PIM can gradually repair the image loss area and generate an attention map simultaneously. The IOM optimizes the details of the generated image based on the information provided by the attention map. The discriminator can capture the local continuity and universal global features of the image better. When comparing the test results with the latest research, the model showed a significant effect in both qualitative and quantitative analyses.

Keywords: Progressive image inpainting · Attention map · GAN

1 Introduction

Image inpainting aims to fill in the missing areas of damaged images to achieve the maximum possible authenticity. These algorithms are usually used for picture editing tasks, such as filtering unwanted objects [2, 16] or repairing old pictures [22]. When the missing area of the image is large, less information can be obtained; thus, increasing the available information of the image is critical to the restoration task. Moreover, the generated portion of the image may not have the same style as the real image; consequently, restorations tasks require both partial continuity and overall similarity.

Most traditional image inpainting methods diffuse the background data to the missing area using a differential operator [4]. Subsequently, with the explosive growth in data volume, patch-based inpainting methods were introduced, where

Supported by the Key Research and Development Program of the Sichuan Province (Grant No. 2022YFQ0047).

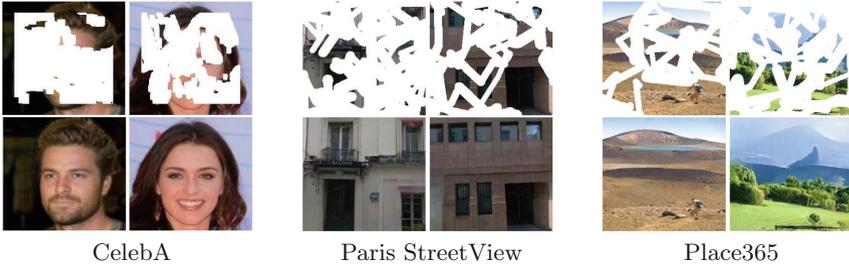


Fig. 1. Some inpainting results using the proposed framework on different datasets. Among them, (Up) Input image with missing areas. The missing areas are shown in white. (Down) Repaired picture using MOPR-GAN(ours).

the algorithm identifies patches in several source pictures to fill in the missing areas, to obtain the maximum similarity [10]. However, these methods perform poorly when restoring a complex detailed texture in a missing area.

Recently, with the rapid development of deep learning technologies, deep convolutional networks have begun to demonstrate extraordinary capabilities in the field of image inpainting. These methods replace the missing parts of the content through continuous learning of existing data, thereby generating a coherent structure in the missing area, which is difficult to achieve using traditional methods. However, images generated by these method are often blurred or have artifacts, which are dissatisfying in terms of visual performance.

To solve this challenging problem, in 2014, Goodfellow et al. proposed a generative adversarial network (GAN) [9], which is a network model composed of two deep convolutional networks called generator and discriminator. The generator uses data to generate images to deceive the discriminator, whereas the discriminator learns the difference between the real and generated images to efficiently identify the generated image, and returns an adversarial loss to improve the generator. Subsequently, several GAN-based models [7, 17] have been developed, many of which use encoder-decoder architectures as their generator.

Researchers began to make bolder attempts. They identified that global inpainting does not directly conform to the way in which the human brain thinks; therefore, they developed several model frameworks for multi-stage inpainting [14, 20]. For example, the EdgeConnect framework, which was proposed by Kamyar et al. [14], divides the task into two stages: edge prediction and edge-based repair. Although using the edge as the pre-information for repairing the image has a good effect, obtaining a good edge is a difficult task. Almost all these algorithms have similar shortcomings.

In this paper, we propose an image inpainting network called Multi-Stage Optimized Progressive Restoration with GAN (MOPR-GAN). The framework follows the basic principles of GAN, and is divided into generators and discriminators. The generator consists of two parts: (1) a Progressive Inpainting Module (PIM) and (2) an Image Optimization Module (IOM). The PIM is responsible for

progressively repairing the missing area and generating an attention map. The IOM refers to the attention map to optimize the details of the initial repaired image, so that it can generate high-quality images. The discriminator is a hybrid of a local discriminator (patchGAN) and a global discriminator (globalGAN). In contrast to the discriminator model proposed by Iizuka et al. [17], we added the attention map generated by the PIM and the IOM; consequently, the discriminator can identify the repaired area more efficiently.

We verified the performance of our model using three standard datasets CelebA [21], Paris Street View [5], and Place365 [3] (certain results are shown in Fig. 1), and compared our method with some of the most advanced frameworks. The primary contributions of this paper are as follows:

- We propose an Image Optimization Module (IOM), that uses a form of competition within the region to extract the best matching distribution of the real image.
- An attention mechanism, called Global Adaptive Attention (GAA), is developed, which acts on the entire network. Under the joint constraints of structure and texture, it gradually updates the attention score following the progress of the network to obtain finer details, thereby enhancing the potential and efficiency of the network.
- We propose a Multi-Stage Optimized Progressive Restoration GAN (MOPR-GAN) framework, which can generate images with better results in terms of details and overall performance; the framework has achieved good performance in both qualitative and quantitative analysis.

Moreover, we performed certain ablation experiments to verify the accuracy of the contributions.

2 Method

In this section, we first introduce the structure of each part of the proposed network framework. And then, we introduce the GAA scheme based on a multi-stage network. Finally, we explain the loss function. The pipeline of our network model is shown in Fig. 2.

2.1 Generator

The generator comprises two modules: 1) PIM, which is used for preliminary image restoration, and 2) IOM, which is used for updating image details. Meanwhile, the generator of the entire network is divided into two forms in the two phases: (1) contains only PIM modules; (2) contains PIM and IOM modules. We will introduce specific training strategy in Sect. 3.1. Here, we explain the two modules in detail.

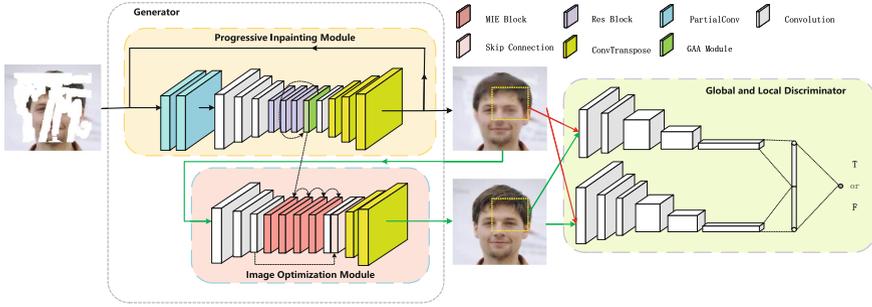


Fig. 2. The overall architecture of the proposed network model is divided into two phases, which we use the red line to represent the first stage and the green line to represent the second one in the diagram (the specific process is described in Sect. 3.1). And then, the network inputs are images and masks, Progressive Inpainting Module (PIM) is used to generate pre-inpainting images recursively, the Image Optimization Module (IOM) optimizes details, and the global and local discriminator (Up is local, Down is global) enhances the network repair potential. (Color figure online)

PIM. The design idea for PIM is derived from RFR [13], which using recursive reasoning to achieve a gradual inpainting process from the edge to the center of the missing area. The difference is that we modified the reasoning network structure and used our own attention mechanism. The details of this part are as follows:

Referring to the network architecture proposed by Johnson et al. [11], we built an encoding-decoding network for the reasoning network. The encoder is composed of three convolutions. The decoder is composed of three deconvolutions with strides of 1/2, 1/2, and 1. Three residual blocks, a GAA and a convolution were added to the middle. The residual blocks are used to avoid the disappearance of the gradient, and the GAA is used to calculate the current attention score and update the feature map. The GAA calculation process is described in Sect. 2.3. Furthermore, the role of the convolutional layer is concatenating the reconstructed feature map with the input one in order to get the final feature.

In general, the PIM iterates multiple times until the feature map is completely filled. Thus, a good pre-inpainting image can be generated.

IOM. Because of the limitation of the repair method based on the partial progressive algorithm, it is inevitable that the output of the preliminary repair image will have a certain degree of local artifacts or chromatic aberration especially the center of the missing area, even though we adaptively mix the previous attention score when computing the new one. Therefore, we launched the IOM to resolve this problem effectively, and generate images with good performance in terms of both details and overall.

Inspired by the multi-scale convolutional fusion block proposed by Yu et al. [19], the IOM of this paper is divided into three parts.

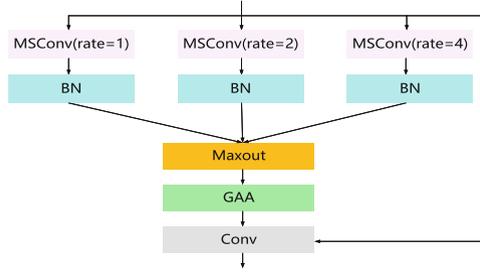


Fig. 3. The main part of the IOM. It can improve the image details in the feature space.

The first stage is the initial extraction of the image which comes from PIM. Through three convolutional layers, it can learn the basic feature information. We retained this current feature map for later use.

The second stage comprises four MIE blocks. The MIE block is shown in Fig 3, whose main body is a multi-scale competitive convolution with a scale of 3. Multi-scale convolution allows the convolution calculations to have a wider field of view, and can capture longer dependencies so that we can improve the network capability. Meanwhile, using Maxout prevents overfitting and provides a lightweight constraint, as well as selects the best feature value of the current position in multiple feature regions. Subsequently, we added a GAA and concatenate the resulting feature map with the input one through a convolution. The role of GAA is the same as in PIM, responsible for calculating the attention score and updating the feature map. The GAA calculation process is described in Sect. 2.3.

Finally, we skip concatenating the feature maps after multiple multi-scale convolutions with the retained feature map in the first stage, which can strengthen the consistency of the image structure and prevent the gradient from disappearing after optimization.

2.2 Discriminator

Iizuka et al. [17] proposed a multi-scale discriminator architecture including a local discriminator and global discriminator, which can enhance the detailed performance of the repaired area and improve the global consistency of the image. Therefore, the discriminator of our network also uses this scheme. The local discriminator uses four convolutional layers and one fully connected layer, whereas the global discriminator uses five convolutional layers and one fully connected layer. Spectral normalization and LeakyReLU [1] with a slope of 0.2 were added to each layer of the two discriminators. In addition, we also added the attention mechanism to the discriminator, which is mainly manifested in the loss function as shown in Sect. 2.4.

2.3 GAA

The attention mechanism can fill the missing area by determining the similarity in the background texture. However, the traditional attention mechanism only calculates the attention score of the feature map, and lacks direct supervision of the attention; therefore, the information learned is unreliable. Conversely, the self-attention method proposed by Peng et al. [15] specified this, but ignored the rich texture information. To solve this problem, we propose the GAA module, which first calculates the structural and texture attention scores, makes them constrain each other, and then performs adaptive accumulation.

Our attention mechanism is divided into two parts: attention calculation and attention transfer. First, in the attention calculation step, we calculate the truncated distance similarity [18] of structure attention with 3×3 patches in the input structure feature (\bar{d}^s) and the cosine similarity of texture attention (\bar{d}^i).

$$\bar{d}_{(x,y,x',y')}^s = \tanh\left(-\left(\frac{d_{(x,y,x',y')}^s - v}{\sigma}\right)\right), \quad (1)$$

where $d_{(x,y,x',y')}^s$ is the Euclidean distance between the patches at (x, y) and (x', y') ; v and σ are the mean value and standard deviation of $d_{(x,y,x',y')}^s$, respectively.

$$\bar{d}_{(n,m,n',m')}^i = \frac{\sum_{i,j \in (-k, \dots, k)} d_{(n+i, m+j, n', m')}^i}{k^2} \quad (2)$$

where $d_{(n,m,n',m')}^i$ is the cosine similarity of each part of the feature pixels between (n, m) and (n', m') .

Then, we used the softmax function to generate structure and texture attention score maps respectively, which are referred to as $score^s$ and $score^i$.

$$score_{(x,y,x',y')}^s = softmax(\lambda \bar{d}_{(x,y,x',y')}^s) \quad (3)$$

$$score_{(n,m,n',m')}^i = softmax(\lambda \bar{d}_{(n,m,n',m')}^i) \quad (4)$$

where λ is set to 50. After that, we used $score^s$ as a constraint to adjust the value of $score^i$. The calculation process is as follows:

$$score'_{(n,m,n',m')} = softmax(2score_{(n,m,n',m')}^i score_{(x,y,x',y')}^s) \quad (5)$$

where $pixel(n, m) \in patch(x, y)$, $pixel(n', m') \in patch(x', y')$. At this point, we can calculate the final attention map. If $score_{(n,m,n',m')}^{i-1}$ represents the attention score computed at the previous iteration, λ is a learnable parameter, so the final attention map $score_{(n,m,n',m')}$ is:

$$score_{(n,m,n',m')} = \begin{cases} \lambda score'_{(n,m,n',m')} + (1 - \lambda) score_{(n,m,n',m')}^{i-1} & \text{if } \exists score_{(n,m,n',m')}^{i-1} \\ score'_{(n,m,n',m')} & \text{otherwise} \end{cases} \quad (6)$$

In particular, $score_{(n,m,n',m')}^{i-1}$ in the first MIE submodule equals to the attention scores generated by the last iteration of PIM.

The next step is attention transfer. We used the score map to reconstruct the feature map. If f represents the input feature and f' represents the reconstructed feature, the formula is

$$f'_{(n,m)} = \sum_{n' \in 1, \dots, W} \sum_{m' \in 1, \dots, H} \text{score}_{(n,m,n',m')} f_{(n',m')} \quad (7)$$

Finally, we reserve the attention score calculated for the next iteration.

2.4 Loss Function

Because our network is a GAN-based model, the loss function is divided into two parts: the loss of the discriminator and the generator. The first we calculated is the loss of the discriminator. Because the attention mechanism is added to it to increase the accuracy when identifying the authenticity of the image, this part is composed of adversarial and attention losses. If D is the discriminator, G is the generator, A is the attention map, I_R represents the real image, and I_G represents the generated image, then its loss function (L_D) is calculated as follows:

$$L_{\text{attention}} = L_{MSE}(D(I_R), 0) + L_{MSE}(D(I_G), A); \quad (8)$$

$$L_D = -\log(D(I_R)) - \log(1 - D(I_G)) + \lambda_{\text{att}} L_{\text{attention}}, \quad (9)$$

where $\lambda_{\text{att}} = 0.2$ in our experiments. Meanwhile, the attention map A will be explained in Sect. 3.1.

Refer to the idea in [14], our generator is trained using a joint loss comprising l_1 , adversarial, perceptual, and style losses. The l_1 loss is normalized by the mask size, while the adversarial loss is provided by the mapping output of the generated image in the discriminator, which is a part of L_D , as

$$L_{\text{adv}} = \log(1 - D(I_G)). \quad (10)$$

The perceptual loss L_{per} and style loss L_{style} are two loss functions that were proposed by Johnson et al. [11]. Finally, the overall loss function of our generator is

$$L_G = \lambda_{l_1} L_{l_1} + \lambda_{\text{adv}} L_{\text{adv}} + \lambda_{\text{per}} L_{\text{per}} + \lambda_{\text{style}} L_{\text{style}}. \quad (11)$$

We finally set $\lambda_{l_1} = 1$, $\lambda_{\text{adv}} = 0.01$, $\lambda_{\text{per}} = 0.2$, and $\lambda_{\text{style}} = 200$. The generator's loss function combination in our model is similar to [14] and has been shown to be effective.

3 Experiments

In this section, we elaborately explain certain related settings and strategies to facilitate the reproduction of the network.

3.1 Training Setting and Strategy

We trained our model with a batch size of six. As this is a GAN-based architecture, there is a generator and discriminator that must be updated; therefore, we used the Adam optimizer. The entire network design strategy is divided into two steps: (1) We only use PIM as the generator until PIM converges. In this case, the attention map in the discriminator loss function is the one generated by the last iteration of the PIM. (2) We join IOM and make PIM and IOM work together as generator. At this time, the attention map becomes the one generated by IOM's last MIE. During each step, we both used learning rates of $2e^{-4}$ and $2e^{-5}$ to train the generator and discriminator, respectively. Then, we used $5e^{-5}$ to fine-tune the generator, while the discriminator used $5e^{-6}$. During the fine-tuning, we did not want to relearn all other network parameters; therefore, we froze all the batch normalization layers of our generator. It is worth noting that we don't unfreeze PIM's batch normalization layers during the second step. All experiments were performed using Python 3.7 on an Ubuntu 20.04 system, with an 11 G NVIDIA GeForce RTX 2080 GPU and Intel Xeon E5-1650 v4 3.60 GHz CPU.

3.2 Datasets

We used datasets CelebA [21], Paris StreetView [5] and Place365 [3] to verify our model. Meanwhile, the irregular masks are automatically generated by scripts.

3.3 Comparison Models

We compared our experimental results with those of certain state-of-the-art methods, both qualitatively and quantitatively. These methods include CA [12], GLCIC [17], PIC [6], EC(EdgeConnect) [14], FE [8], and RFR [13].

4 Results

We conducted experiments on the three datasets, and compared the results with the methods mentioned in the previous section both qualitatively and quantitatively. Moreover, we conducted ablation tests to verify the necessity of the proposed module.

4.1 Qualitative Comparison

Figure 4, 5, and 6 show the visualization of our approach when compared to the four state-of-the-art methods on the three datasets. In comparison, our model showed excellent visual effects, and when the missing area became larger, the effect was more apparent. This shows the superiority of our Network.

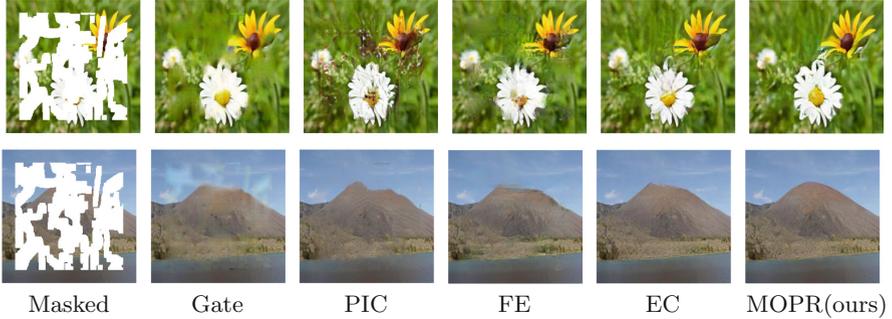


Fig. 4. Results on Place365 [3]



Fig. 5. Results on CelebA [21]



Fig. 6. Results on Paris StreetView [5]

Table 1. Quantitative results over three standard datasets with six models: Contextual Attention(CA) [12], Globally and Locally Consistent Image Completion(GLCIC) [17], Pluralistic Image Completion(PIC) [6], EdgeConnect(EC) [14], Recurrent Feature Reasoning(RFR) [13], MOPR-GAN(Ours). The best result of each group is bolded. *Higher is better. †Lower is better.

Dataset		Place365			CelebA			Paris Street View		
Mask ratio		10%-20%	30%-40%	50%-60%	10%-20%	30%-40%	50%-60%	10%-20%	30%-40%	50%-60%
SSIM*	CA	0.893	0.739	0.502	0.888	0.750	0.614	0.905	0.766	0.625
	GLCIC	0.862	0.686	0.535	0.865	0.689	0.560	0.878	0.724	0.588
	PIC	0.932	0.786	0.494	0.965	0.881	0.672	0.930	0.785	0.519
	EC	0.933	0.802	0.553	0.975	0.915	0.759	0.950	0.849	0.646
	RFR	0.939	0.819	0.596	0.981	0.934	0.819	0.954	0.862	0.681
	MOPR(Ours)	0.941	0.825	0.744	0.988	0.942	0.862	0.966	0.901	0.732
PSNR*	CA	24.36	19.13	16.56	25.32	19.94	17.18	26.09	20.74	18.17
	GLCIC	23.49	18.50	16.06	24.09	18.50	16.24	25.72	21.02	18.71
	PIC	27.14	21.72	17.17	30.67	24.74	19.29	29.35	23.97	19.52
	EC	27.17	22.18	18.35	32.48	26.62	21.49	31.19	26.04	21.89
	RFR	27.75	22.63	18.92	33.56	27.76	22.88	31.71	26.44	22.40
	MOPR(Ours)	29.15	24.60	20.77	34.81	27.75	24.79	31.54	26.67	23.44
Mean $l1^\dagger$	CA	0.0241	0.0615	0.0991	0.0248	0.0564	0.0921	0.0210	0.0553	0.0906
	GLCIC	0.0266	0.0678	0.1096	0.0253	0.0695	0.1121	0.0220	0.0558	0.0902
	PIC	0.0161	0.0441	0.0944	0.0111	0.0314	0.0749	0.0140	0.0379	0.0799
	EC	0.0157	0.0408	0.0821	0.0088	0.0247	0.0572	0.0110	0.0286	0.0582
	RFR	0.0142	0.0381	0.0761	0.0075	0.0212	0.0470	0.0110	0.0275	0.0546
	MOPR(Ours)	0.0129	0.0359	0.0736	0.0082	0.0150	0.0383	0.0113	0.0269	0.0503

4.2 Quantitative Comparisons

We also performed quantitative comparisons from three aspects: 1) structural similarity index (SSIM), 2) peak signal-to-noise ratio (PSNR), and 3) mean $l1$ loss, to evaluate our model and compare the results with other methods. Table 1 lists the results of the six methods for different irregular mask ratios for the three standard datasets. It can be observed from the table that our method shows superior results under different irregular mask ratios on the Places2, CelebA, and Paris StreetView datasets in most cases, particularly for large holes. The data of CA [12] and RFR [13] in the table were obtained from their papers, and the remaining data were obtained using the pre-model provided by their author.

4.3 Ablation Studies

The preceding content illustrates the effectiveness of the overall architecture of our model. In this section, we present a verification of the validity of our proposed contributions. Here, we describe the proposed IOM and GAA.



Fig. 7. Qualitative comparison of the two methods on the Paris Streetview dataset: 1)without IOM, 2)with IOM.

Table 2. Quantitative comparison of the two methods on the Paris Streetview dataset with different irregular mask ratios: 1)without IOM, 2)with IOM.

Dataset	Without IOM		With IOM	
Mask ratio	20%–30%	40%–50%	20%–30%	40%–50%
SSIM	0.910	0.799	0.917	0.839
PNSR	28.14	23.80	28.30	24.72
Mean l1	0.0254	0.0375	0.0246	0.0332

Capabilities of the IOM. To demonstrate the function of the IOM, we compared the network repair effects of using and deprecating this module. It can be observed from Fig. 7 that it is lacking in details, although an almost complete image can be repaired without the IOM. The addition of the IOM can significantly improve the local performance of the image. From the perspective of a quantitative comparison, as shown in Table 2, we tested two cases with different sizes of irregular masks, and identified that IOM can improve the network performance significantly. Furthermore, the missing area is larger, the more pronounced the effect.



Fig. 8. The comparison results with different attention. (1) Masked Image; (2) Traditional Attention; (3) Existing Progressive Attention; (4)GAA(ours)

Capabilities of the GAA. As mentioned in Sect. 2.3, our GAA module is a progressive attention mechanism with equal emphasis on the structure and texture. We compared it with other existing attention mechanisms, and the results

are shown in Fig. 8. From this, we can determine that the progressive accumulation of the attention mechanism is more suitable with global consistency, which can make it more deceiving to the eye, particularly for large holes. Further, allowing the structure to constrain the details (shown in Eq. 5) can effectively prevent local artifacts and generate a more realistic image.

5 Conclusion

In this paper, we built a new GAN-based image inpainting framework (MOPR-GAN), which first repairs the missing areas from the outside to the inside step by step, and then, uses the proposed IOM to correct the details of the generated pre-repair images to obtain more accurate results. Moreover, we proposed a GAA module that acts on the entire network. Through qualitative and quantitative analyses on three standard datasets, plus several ablation experiments, the superiority of our network was proved.

References

1. Maas, A.L., Hannun, A.Y., Ng, A.Y.: Rectifier nonlinearities improve neural network acoustic models. In: Proceedings of ICML, vol. 30 (2013)
2. Antonio Criminisi, P.P., Toyama, K.: Region filling and object removal by exemplar based image inpainting. *IEEE TIP* **13**(9), 1200–1212 (2004)
3. Bolei Zhou, Agata Lapedriza, A.K.A.O., Torralba, A.: Places: a 10 million image database for scene recognition. *IEEE TPAMI* **40**(6), 1452–1464 (2018)
4. C. Ballester, M. Bertalmio, V.C.G.S., erdera, J.V.: Filling-in by joint interpolation of vector fields and gray levels. *IEEE TIP* **10**(8), 1200–1211 (2001)
5. Doersch, C., Singh, S., Gupta, A., Sivic, J., Efros, A.: What makes Paris look like Paris? *ACM TOG* **31**(4) (2012)
6. Zheng, C., Cham, T.J., Cai, J.: Pluralistic image completion. In: Proceedings of CVPR, pp. 1438–1447 (2019)
7. Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T., Efros, A.A.: Context encoders: feature learning by inpainting. In: Proceedings of CVPR, pp. 2536–2544 (2016)
8. Liu, H., Jiang, B., Song, Y., Huang, W., Yang, C.: Rethinking image inpainting via a mutual encoder-decoder with feature equalizations. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) *ECCV 2020*. LNCS, vol. 12347, pp. 725–741. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58536-5_43
9. Goodfellow, I., et al.: Generative adversarial nets. In: Proceedings of NIPS, vol. 27, pp. 2672–2680 (2014)
10. Huang, J.B., Kang, S.B., Ahuja, N., Kopf, J.: Image completion using planar structure guidance. *ACM TOG* **33**(4), 1–10 (2014)
11. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) *ECCV 2016*. LNCS, vol. 9906, pp. 694–711. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46475-6_43
12. Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X., Huang, T.S.: Generative image inpainting with contextual attention. In: Proceedings of CVPR (2018)

13. Li, J., Wang, N., Zhang, L., Du, B., Tao, D.: Recurrent feature reasoning for image inpainting. In: Proceedings of CVPR, pp. 7760–7768 (2020)
14. Nazeri, K., Ng, E., Joseph, T., Qureshi, F.Z., Ebrahimi, M.: Edgeconnect: generative image inpainting with adversarial edge learning. In: Proceedings of ICCV (2019)
15. Peng, J., Liu, D., Xu, S., Li, H.: Generating diverse structure for image inpainting with hierarchical VQ-VAE. In: Proceedings of CVPR (2021)
16. Song, L., Cao, J., Song, L., Hu, Y., He, R.: Geometry-aware face completion and editing. CoRR (2008)
17. Iizuka, S., Simo-Serra, E., Ishikawa, H.: Globally and locally consistent image completion. ACM TOG **36**(4), 1–14 (2017)
18. Min-cheol, S., Shin, Y.-G., Kim, S.-Wk., Park, S., Ko, S.-J.: Pepsi: fast image inpainting with parallel decoding network. In: Proceedings of CVPR, pp. 11360–11368 (2019)
19. Yu, F., Koltun, V.: Multi-scale context aggregation by dilated convolutions. In: Proceedings of ICLR, pp. 23–32 (2016)
20. Xiong, W., et al.: Foreground-aware image inpainting. In: Proceedings of CVPR (2019)
21. Liu, Z., Luo, P., Wang, X., Tang, X.: Deep learning face attributes in the wild. In: Proceedings of ICCV, pp. 3730–3738 (2015)
22. Wan, Z., et al.: Bringing old photos back to life. In: Proceedings of CVPR, pp. 2747–2757 (2020)