Research on Efficient Image Inpainting Algorithm Based on Deep Learning

Tao Qin[1], Juanjuan Liu[2(✉)], and Wenchao Xue[3]

[1] National Computer Network Emergency Response Technical Team/Coordination Center of China, Beijing 100020, China
[2] School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing 100083, China
[3] Engineering Agent Construction Management Office of Navy Logistics Department, Beijing 100036, China

Abstract. The rapid development of deep learning has brought a new development direction for image inpainting, changing the traditional image inpaiting algorithm, which can only repair the problem of small area damage based on the structure and texture of the damaged image. In recent years, image inpainting algorithm based on deep learning has received widespread attention from industry and academia so that it has made great progress. However, the current image inpainting algorithm based on deep learning still has the problem of consuming so much time. In order to solve the above problem, an end-to-end image inpainting algorithm suitable for real-time scene was proposed. The mask of the damaged image generated by D-linkNet network, the edge information of the damaged image, and damaged image were used to control the network input, which avoided the damage to the existing semantics of the image and retained the intact image outside the damaged. On this basis, in order to improve the performance of image inpainting, Convolutional Block Attention Module (CBAM) was used in the residual network. Experimental results show that, compared with edge information-based deep learning algorithm Edge Connect, the repair speed is twice as fast as Edge Connect while the repair results are similar.

Keywords: Deep learning · Image inpainting · D-LinkNet · CBAM attention

1 Introduction

Image inpainting belongs to a branch of computer vision, which is to repair the defect area in the image according to the existing information in the image. With the rapid development of deep learning, image inpainting has become a research hotspot and has been widely used in the fields of image redundant target removal, criminal investigation facial reconstruction, aerospace engineering and bioengineering. There are two traditional image inpainting methods, one is based on sample block [1, 2], and the other is based on sparse representation [3]. These two methods can only repair small and medium-sized defects in the image. With the rapid development of deep learning theory, this has

X. Sun et al. (Eds.): ICAIS 2021, LNCS 12736, pp. 109–120, 2021.
https://doi.org/10.1007/978-3-030-78609-0_10

brought new opportunities for image inpainting technology, and gradually evolved a new image inpainting method – image inpainting method based on deep learning [4].

Although image inpainting using deep learning has achieved good results, the model of deep learning algorithm takes too much time, and it is difficult to be applied to scenes with high time requirements such as criminal investigation and edge calculation. In 2019, the Edge Connect [5] method based on edge information has achieved great repair effect. Edge Connect algorithm constructs a two-stage image inpainting network. The first stage is responsible for repairing the edge of the image, and the second stage is for filling the image content. The algorithm needs to train two image inpainting networks, which takes too much time. At the same time, the repair effect of the first stage has a great impact on the second stage, which largely determines the repair effect of the second stage.

In view of shortcomings of Edge Connect method, this paper proposes an end-to-end generative adversarial network, through which faster and natural inpainting results can be obtained. Our paper makes the following contributions:

(1) This paper proposes an end-to-end image inpaiting method to improve the network repair speed;
(2) CBAM (Convolutional Block Attention Module) attention is added in the network to improve the inpainting effect;
(3) D-LinkNet network is used to generate the mask to avoid damage to the normal area.

The work arrangement of this paper is as follows: the second part mainly summarizes the related work of this paper and introduces the development of image inpainting algorithms based on deep learning. The third part mainly introduces the network module of this paper, including segmentation network, image inpainting network and attention module. The fourth part mainly analyzes the experimental results as well the advantages and disadvantages of this method. The fifth part is the conclusion, which summarizes the methods of this paper and puts forward the development direction.

2 Related Work

Most of the image inpainting methods based on deep learning use GAN (Generative Adversarial Networks), which is generated by the Ian J. Goodfellow et al. [6]. GAN is composed of the generative network and the discriminant network. The generative network is mainly responsible for generating data information, and the discriminant network is responsible for determining which data information is generated by the generative network, that is, false data. In the process of confrontation, the generated information is closer to the real data, so as to "cheat" the discriminant network. However, the ability of the discriminant network to identify real or false data is getting stronger and stronger. The generative network and the discriminant network game each other in this process, and reach a state of dynamic equilibrium. So that the discriminating network cannot distinguish between true and false, and obtain a better generator. Since GAN was put forward, a large number of scholars have started to apply it to image inpainting. Deepak Pathak et al. [7] firstly used GAN in image inpainting in 2016, they proposed a context-based adaptive encoder. The generative network adopts the structure of encoder and

decoder, and the discriminant network adopts the structure of five-layer convolution to judge the generated or real image. This method can only deal with the damaged image of fixed damaged area, without adding texture details consistent with the real image, making people look particularly fuzzy. However, it is the first method of GAN network image inpainting and subsequent researchers have made improvements on the basis of this paper. Chao Yang et al. [8] proposed a new network structure in 2017, which added a texture generative network on the basis of the structure of encoder and decoder. They propose to divide the damaged area into fixed size patches and divide other parts of the image into patches of the same size. They then extract their texture information by the texture generation network and find the information of the closest two patches to match to ensure the visual consistency of the image. This method can process the damaged images of higher resolution, however, it can only repair the images of fixed areas and cannot meet people's needs for image inpainting. In the same year, Satoshi iizuka [9] et al. proposed a new network structure, which improved the discriminant network of the structure of encoder and decoder. They construct a global discriminant network and a local discriminant network. They then input the repaired complete image and repaired damaged areas into the discriminant network respectively. On the basis of keeping the original structure unchanged, the author added expansion convolution in the generative network to increase the sensor field. This method can produce fine texture details, but it cannot repair large areas of the image. Nvidia [10] proposed a brand new image inpainting method in 2018, in which they proposed the idea of partial convolution to repair the image with irregular holes. The author convolves the mask updated from the previous layer with the image in each layer and obtains great inpainting results. However, the author also pointed out in the paper that although this method can reconstruct the missing area of the image well, it cannot guarantee that the generated image is exactly the same as the real image.

Kamyar Nazeri et al. [5] proposed an image inpainting method combining edge information in 2019. The author constructs a two-stage image inpainting network in this paper. They firstly use the edge generative network to supplement the edge information of the damaged image. They then use the content generation network to carry out the semantic repair of the image information. Finally, the author verifies that this method can achieve good repair results through experiments. The work of our paper is based on the Edge Connect method proposed by Kamyar Nazeri et al. By simplifying the network structure and improving the training speed of the network, in order to achieve better repair results, by adding CBAM attention to the network, finally by using the split network to generate an image mask to avoid repairing other intact areas.

3 Image Inpainting Network Model

In this paper, shadow areas are generated randomly in the images of the Places2 dataset as the damaged image dataset. The network module of image inpainting is shown in Fig. 1: (1) Image segmentation network is necessary to produce a mask of the damaged image, which is used to limit the damaged area. (2) The Canny edge detector is useful to detect the edge information of the damaged image. (3) During the image inpainting, the damaged image, the mask of the damaged image generated by D-LinkNet [11] and the

edge of the damaged image are input into the image inpainting network. CBAM attention module is added into the image inpainting network for accurate repair of the damaged image. In this section, the network module used in the paper is briefly introduced.

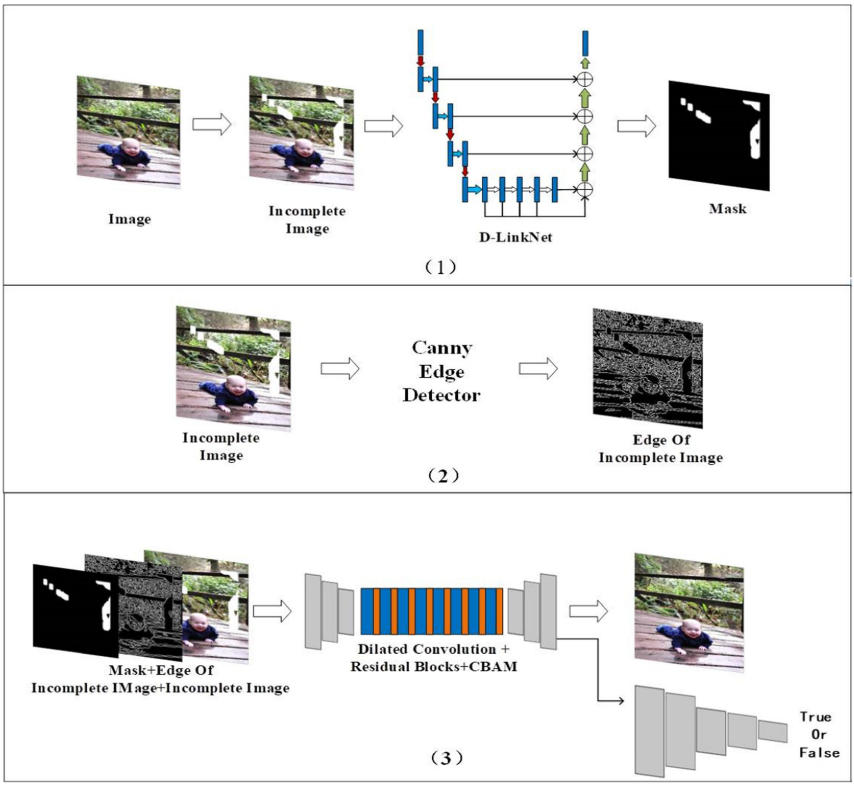

Fig. 1. Image inpainting network model

3.1 Segmentation Network

Image segmentation is an important part of image processing, which is widely used in medical diagnosis and treatment, image beautification and automatic driving. Image segmentation is to classify each pixel in the image into its own category, which is equivalent to a marking process. Each pixel is marked and then the pixels belonging to the same category are grouped together, so we can segment the image into different regions. For example, the nose, mouth, eyes and other parts can be segmented by using segmentation network in face image.

In this paper, while repairing the damaged area without changing the other intact part of the image, generating a mask is necessary to restrict the damaged region. The mask of the damaged image is that the part of the damaged is white, while the other part is black. The mask is added with the damaged part during the inpainting process, and

the damaged image is restricted. For a damaged image, its mask is not obtained directly, so it is important to produce the mask. The process of mask generation is to separate the damaged area (the pixel point of the damaged area is different from the intact area, the pixel point is 255) from the image, which is similar to the image segmentation. Therefore, D-LinkNet segmentation network is used in this paper to produce the image mask.

D-LinkNet [11] is an improvement on the basis of LinkNet network. Encoder adopts ResNet structure, the middle part adds dilated convolution to increase the sensor field of the network, and the last part adds the results obtained from the first two parts for feature fusion. This network is first used for road segmentation in high-resolution satellite images, which is a very simple and efficient segmentation method. The damaged image is input into the segmentation network to obtain a clear mask image.

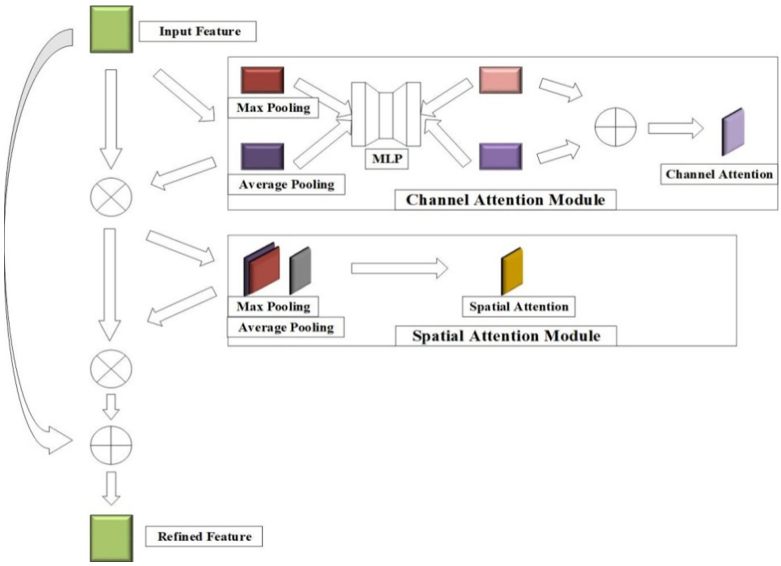

Fig. 2. CBAM attention mechanism

3.2 CBAM Attention Module

CBAM attention module is mainly used between the two feature layers of convolutional neural network [12]. Channel attention and spatial attention are mainly used. In the channel attention module, the main problem to be solved is which channel information is more important. There are two pooling methods used in the channel attention module: average pooling and maximum pooling. Three full connection layers are used and the two outputs are superimposed after two pooling methods. The spatial attention module mainly solves the problem of which position is more important in the two-dimensional plane. Two pooling methods are also used, and then the convolution layer is passed to obtain the convolution result. The operation process is shown in Fig. 2.

3.3 Improving Inpainting Network

The image inpainting network used in this paper is generated countermeasure network. The basic idea of generative antagonism is to reach an equilibrium state through the dynamic game of generating network and discriminating network. The generating network is responsible for generating some missing things in the new image to "cheat" the network, and the discriminating network is responsible for identifying the authenticity of the output of the generated network. In this paper, the structure of the generated network is that three-layer convolution is used for down sampling, then eight residual blocks of dilation convolution are connected, and finally three-layer convolution is used for up sampling, as shown in Fig. 3. Different from edge connect, this paper only uses one generated countermeasure network, which greatly shortens the repair time. Meanwhile, in order to ensure the effect of network, CBAM module is added to the residual block.

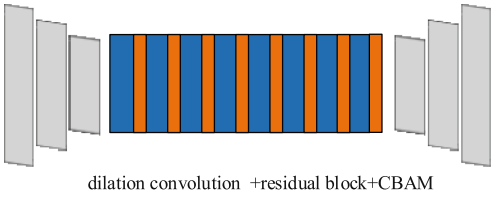

Fig. 3. Basic structure of generating network

Residual network combines the information of shallow network with that of deep network, which avoids the degradation of network caused by the shift of learning focus while deepening the network, and avoids the problems of gradient disappearance and explosion due to the increase of network depth. In this paper, based on the structure of residual network, expansion convolution and CBAM module are added to further improve the performance of residual network. The network structure is shown in Fig. 4. In this paper, a residual network with dilation convolution and CBAM module is used in the eight residual blocks in the middle of the generating network, and the ReLu activation function is not used in the latter layer of the residual block. Expanding convolution can increase the receptive field of the network and extract more features, while CBAM module gives more weight to the important features to obtain more important features. The combination of the two can obtain a wider range of effective features and improve the performance of the network.

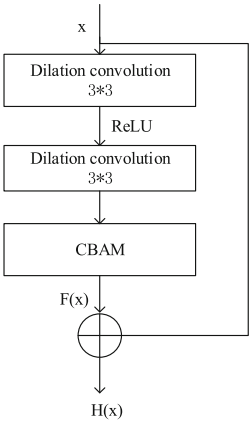

Fig. 4. Basic structure of residual block

4 Experiments

In this experiment, 5000 images are selected from the places2 [13] dataset for training and 2991 images are selected for testing. These images are images about the scene. In this paper, the official weight of Edge Connect is used for testing. The batchsize in the training process is 6, and the number of iterations is 300,000. The code is carried out in Pytorch.

4.1 Qualitative Analysis

This experiment is mainly compared with the current most advanced Edge Connect method. Comparison experiments of the two methods are performed on the same dataset for the reliability of the test. The experimental inpainting results are shown in Fig. 5 with the same configuration and the same number of iterations.

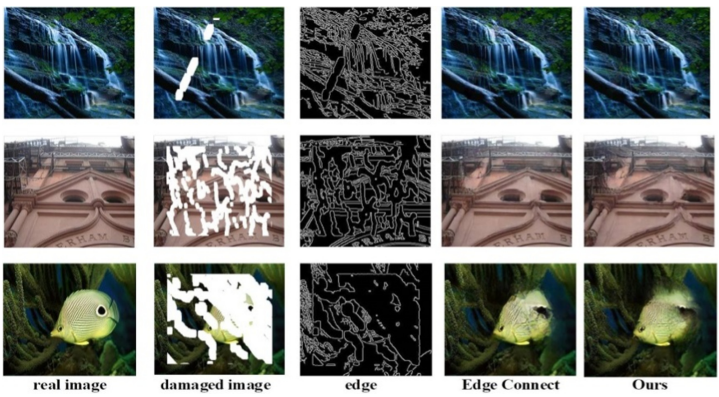

Fig. 5. Inpainting image effect contrast

4.2 Quantitative Analysis

For the objectivity of the experimental results, three indicators commonly used to evaluate the image quality are also calculated, namely L1 loss, SSIM (Structural Similarity Index Measure) and PSNR (Peak Signal to Noise Ratio), which are used to calculate the difference between the restored image and the real image.

L_1 loss, also known as the minimum absolute value error, represents the absolute value of the pixel error between the inpainting image and the real image. The smaller the value of L_1 loss, the smaller the absolute value error of the image when we calculate it.

SSIM calculate the degree of structural similarity between two images, and its value is between 0 and 1. The value of SSIM is 1 when two images are exactly the same. The calculation formula of SSIM is shown as Eq. (1) when the structural similarity of image x and y is calculated:

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

In Eq. (1), μ_x and μ_y represent the average value of image x and y separately. σ_x^2 and σ_y^2 represent the variance of image x and y separately. σ_{xy} represents the mean variance of image x and y. c_1 and c_2 represent two different constants. We want the result as large as impossible when the SSIM between the inpainting image and the real image is calculated.

PSNR is an objective standard to evaluate an image. It is the signal-to-noise ratio of an image, whose unit is dB. The larger PSNR is, the smaller the image noise is and the less distortion is. The calculation formula of PSNR is shown in Eq. (2):

$$PSNR = 10 \times \log_{10}\left(\frac{(2^n - 1)^2}{MSE}\right) \tag{2}$$

In Eq. (2), all the calculations are based on the pixel points of the image. MSE represents the mean square error between the original image and the inpainting image, n represents the number of bits per pixel, and we take the value of n as 8. We want the result as large as impossible when we calculate the PSNR of the inpainting image.

The comparison results of the experiment are shown in Table 1. As can be seen from the comparison of results, when mask is 0–10% and 10%–20%, our method is similar to that in Edge Connect. But our method is slightly worse than that in Edge Connect when mask is 20%–60%.

Table 1. Comparison of experimental results

	Mask	Edge connect	Ours
L_1	0–10%	0.006	0.007
	10%–20%	0.016	0.017
	20%–30%	0.027	0.030
	30%–40%	0.040	0.043
	40%–50%	0.053	0.057
	50%–60%	0.079	0.084
SSIM	0–10%	0.979	0.978
	10%–20%	0.944	0.937
	20%–30%	0.892	0.878
	30%–40%	0.834	0.815
	40%–50%	0.762	0.737
	50%–60%	0.623	0.593
PSNR (dB)	0–10%	32.978	32.565
	10%–20%	27.602	26.982
	20%–30%	24.649	24.029
	30%–40%	22.589	22.047
	40%–50%	20.908	20.421
	50%–60%	18.448	17.998

4.3 Time Analysis

From the above data, we can see that our method is similar to the Edge Connect method. But from the network structure, we have improved our structure, so the time efficiency is higher than Edge Connect method.

The inpainting time of each image is calculated, and the results are shown in Table 2. Masks are not used directly in this paper because there are no masks in the natural damaged image. Our approach is closer to natural image inpainting than Edge Connect method.

Table 2. Time comparison of each image

	Edge connect	Ours
Time(s)	0.0230	0.0124

4.4 Comprehensive Evaluation Function

In order to synthesize each index, this paper constructs a weighted function P to comprehensively evaluate the performance of image inpainting. The larger the value of P, the better the quality of image inpainting. The function is represented as follows Eq. (3):

$$P = \frac{w_1 \times S_S + w_2 \times P_S}{w_3 \times L_1 + w_4 \times t} \tag{3}$$

Among them, S_S and w_1 represent the value and weight coefficient of SSIM, Ps and w_2 respectively represent the value and weight coefficient of PSNR, L_1 and w_3 respectively represent the value and weight coefficient of L_1, and t and w_4 represent the value and weight coefficient of inpainting time respectively.

From the experimental comparison in Table 1, it can be seen that other performance indicators are not significantly different, so we make w_1 take 0.3, w_2 take 0.01, w_3 take 2, observing the influence of time weight w_4 on image inpainting performance, and the results are shown in Fig. 6. The solid line depicts the algorithm we used in this paper, and the dotted line depicts the edge connect algorithm. It can be seen from the figure that when w_4 is about 0.70 (actually 0.69906, about 0.70), the comprehensive performance of the proposed algorithm is better than that of edge connect algorithm. The algorithm proposed in this paper is more suitable for scenes with larger time weight and smaller delay, such as the scene of edge computing.

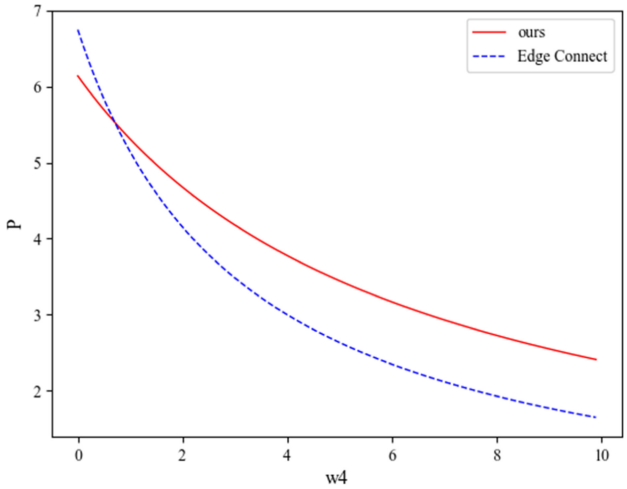

Fig. 6. w4 and P transformation curve

In order to further describe the influence of time on the comprehensive evaluation function, we take w_4 as 0.70 to draw the transformation curve between time t and comprehensive evaluation index P, as shown in Fig. 7. As can be seen from Fig. 7, when the effect P is the same, our algorithm consumes less time. In other words, under the same comprehensive performance, the speed of this algorithm is faster than that of Edge Connect algorithm.

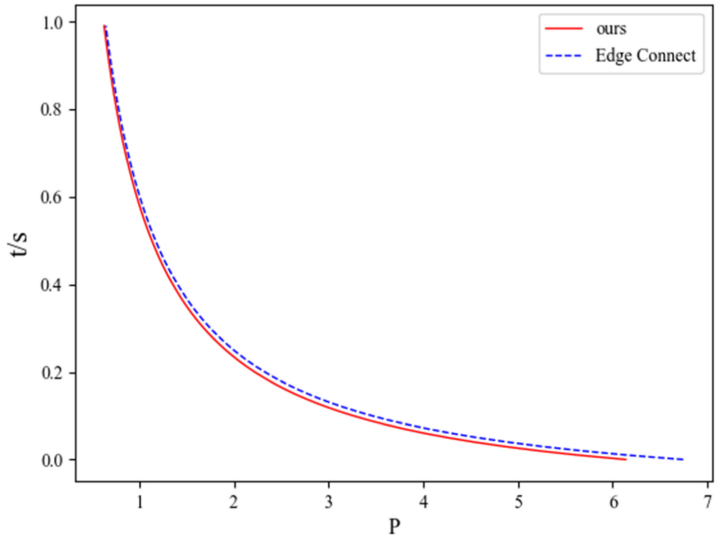

Fig. 7. Transformation curve of t and P

5 Conclusion

This paper proposes a one-stage image inpainting method based on the Edge Connect method. Our inpainting results are similar to Edge Connect network, but our approach has great potential in the future. Our approach is closer to the natural image inpainting and faster than Edge Connect method. Since the training resources are limited, we have not used all the training data. We believe that the next research work can bring about greater improvement.

References

1. Criminisi, A., Patrick, P., Kentaro, T.: Object removal by exemplar-based inpainting. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, USA, vol. 2 (2003)
2. Xu, Z., Sun, J.: Image inpainting by patch propagation using patch sparsity. IEEE Trans. Image Process. **19**(5), 1153–1165 (2010)
3. Gao, C.Y., Xu, X.E., Luo, Y.M.: Object image restoration based on sparse representation. Chin. J. Comput. (9), 4 (2019)
4. Yu, J., Lin, Z., Yang, J.: Generative image inpainting with contextual attention. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, New Zealand, pp. 5505–5514 (2018)
5. Nazeri, K., Ng, E., Joseph, T.: Edgeconnect: generative image inpainting with adversarial edge learning. arXiv preprint arXiv:1901.00212, 1901 (2019)
6. Goodfellow, I., Pouget-Abadie, J., Mirza, M.: Generative adversarial nets. In: Advances in Neural Information Processing Systems, pp. 2672–2680 (2014)
7. Pathak, D., Krahenbuhl, P., Donahue, J.: Context encoders: feature learning by inpainting. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, USA, pp. 2536–2544 (2016)

8. Yang, C., Lu, X., Lin, Z.: High-resolution image inpainting using multi-scale neural patch synthesis. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, USA, pp. 6721–6729 (2017)
9. Iizuka, S., Simo-Serra, E., Ishikawa, H.: Globally and locally consistent image completion. ACM Trans. Graph. (ToG) **36**(4), 107 (2017)
10. 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
11. Zhou, L., Zhang, C., Wu, M.: D-LinkNet: LinkNet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction. In: Proceedings of CVPR Workshops, New Zealand, pp. 182–186 (2018)
12. Woo, S., Park, J., Lee, J.-Y., Kweon, I.S.: CBAM: convolutional block attention module. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) ECCV 2018. LNCS, vol. 11211, pp. 3–19. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01234-2_1
13. Zhou, B., Lapedriza, A., Khosla, A.: Places: A 10 million image database for scene recognition. IEEE Trans. Pattern Anal. Mach. Intell. **40**(6), 1452–1464 (2017)