

Article

Image Completion with Large or Edge-Missing Areas

Jianjian Ji and Gang Yang *

School of Information Science & Technology, Beijing Forestry University, Beijing 100083, China; kakacnn@163.com

* Correspondence: yanggang@bjfu.edu.cn; Tel.: +86-136-812-58163

Received: 3 December 2019; Accepted: 26 December 2019; Published: 31 December 2019



Abstract: Existing image completion methods are mostly based on missing regions that are small or located in the middle of the images. When regions to be completed are large or near the edge of the images, due to the lack of context information, the completion results tend to be blurred or distorted, and there will be a large blank area in the final results. In addition, the unstable training of the generative adversarial network is also prone to cause pseudo-color in the completion results. Aiming at the two above-mentioned problems, a method of image completion with large or edge-missing areas is proposed; also, the network structures have been improved. On the one hand, it overcomes the problem of lacking context information, which thereby ensures the reality of generated texture details; on the other hand, it suppresses the generation of pseudo-color, which guarantees the consistency of the whole image both in vision and content. The experimental results show that the proposed method achieves better completion results in completing large or edge-missing areas.

Keywords: image completion; generative adversarial networks; large missing area; edge-missing area

1. Introduction

Image completion technology is designed to synthesize the missing or damaged areas of an image and is a fundamental problem in low-level vision. The technology has attracted widespread interest in the field of computer vision and graphics because it can be used to complement occluded image areas or to repair damaged photographs [1,2]. In addition, it can erase distracting scene elements or adjust the object position in the image for better composition, or restore the image content in the occluded image area [3–5]. These and many other editing operations require automatic completion of missing areas, but due to the inherent ambiguity of the problem and the complexity of natural images, it is still a challenging task to synthesize content with reasonable detail for any natural image. At present, although many methods for image completion have been proposed, such as sample-based [6–9] and data-driven [4,10,11] image completion, problems remain. Not only does the texture pattern need to be completed, but the structure of the scene and the object to be completed must be understood.

In the past, the main method of image completion was to copy the existing image block in the uncorrupted area to the missing area. However, this method can achieve effective results only when the image to be complemented has a strong structure, the texture information such as the color of each region has strong similarity, and the missing region has a regular shape [12–16]. This approach has significant limitations for image completion work with complex structures and contents. Therefore, the method of intercepting an image block directly from surrounding information is not versatile.

Deep learning has advanced substantially and been used in image completion due to the strong learning ability of deep neural networks. The usual deep learning-based image completion methods are completed by the generation adversarial network (GAN) [17]. The generator and discriminator in GAN can generate clear and reasonable texture content [18–20]. However, the existing GAN-based image completion methods have important limitations. GAN adopts an encoder-decoder structure in

which two fully connected networks (FCNs) are used as intermediate layers connecting two structures. However, the existing image completion methods using GAN cannot maintain the consistency of the image space structure and context information. Therefore, most of the methods proposed above can only synthesize content with a small missing area, and the synthesis method depends on the low-level features of the image. Additionally, the input can only be a fixed-size picture. When the corrupted image has a large missing area, a complicated structure, or is located at the border of the image, the result is not ideal.

To solve the above problems, Iizuka et al. [21] proposed a globally and locally consistent image completion method. However, when the areas of the missing regions are large, it will generate distortion and blurred results. Notably, when the missing areas are located at the edge of the image, the results will have blanking and color distortion due to the lack of surrounding information and the instability of training (Figure 1).

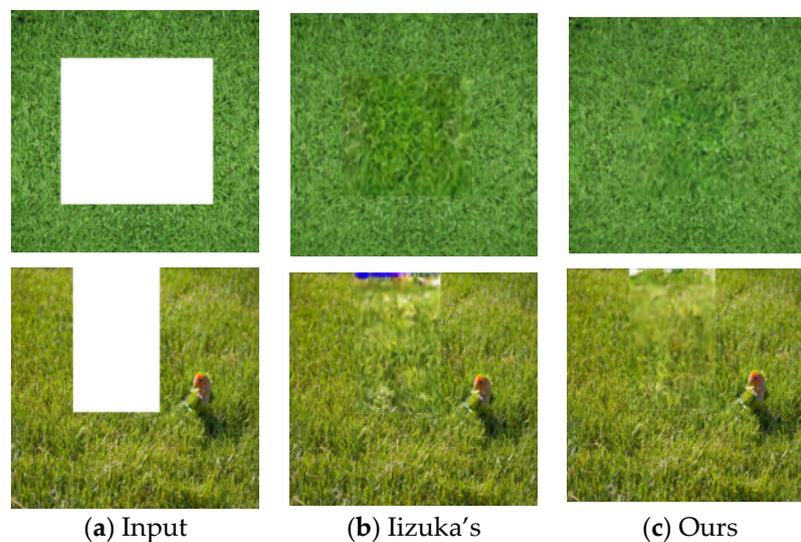


Figure 1. (a) A corrupted image to be completed; (b) The result of Iizuka's method; (c) The result of our method.

To solve the problems of completing large missing areas or regions located at the border of the image, and overcome the problem of unstable training of adversarial network, this paper proposes an image completion method with large or edge-missing areas, and makes improvements of the network structure used in Iizuka's method. The contributions can be summarized as follows:

- First, by separating the middle region of the preliminary complemented area and inputting it into the local discriminator for adversarial training overcomes the problems of the existing method, such as ambiguity and distortion when the missing areas are large. It not only ensures the authenticity and clarity of the content of the central area, but also guarantees the local consistency between the central area and surrounding area.
- Second, based on the existing network structure, we add a local discriminator 2 to solve the pseudo-color problem caused by the inconsistent training speed of the generator and the discriminators in the original network structure. In addition, we have made another two improvements to the original network structure, one is to remove the sigmoid activation function in the last layer of the original discriminator network, and the other is to replace the ReLU layer (Rectified Linear Unit layer) with the combination of batch normalization layer (BN layer) and Leaky_ReLU layer, so that the completion results are more realistic and the edges are more fused.

2. Related Work

At present, many methods have been introduced to solve the problems of image completion. Image completion based on diffusion is the first digital repair method. In this method, missing pixels are filled by diffusing image information from a known region to a missing region at the pixel level. These algorithms are based on variational methods and the partial differential equation (PDE) theory. The PDE algorithm was proposed by Bertalmio et al. [22]. The algorithm is an iterative algorithm, and its primary goal is to propagate the gradient direction and the grey value of the image to the interior of the region to be filled and to solve the demand for high-order partial differential equations in image processing. If the missing areas are small, the PDE algorithm will produce good results. However, when the missing areas are large, the algorithm will take a long time to process and will produce bad results. Vague areas will be produced by the algorithm that make the filling area unnatural. Another more traditional approach is to fill a missing image area with a large external database in a data-driven manner. These methods assume that areas surrounded by similar environments may have similar content [4,10,11], such as Zhu et al. assume that the input image is taken at a well-known landmark, so similar images taken at the same location can be easily found on the Internet. This method is very effective in finding sample images with enough visual similarity to the query but may fail when the query images are not of good quality in the database. These methods also require access to an external database, which greatly limits the possible application scenarios.

Compared with data-driven technology, the sample-based method can perform more complex image filling task and fill large areas of missing regions in natural images. Sample-based image completion method was initially used for texture synthesis by resampling either pixels [23] or whole patches [24] of the original texture to generate a new texture. This method is then extended to image mosaic [25] by image segmentation and texture generation based on energy optimization. In image completion applications, this method has been improved [12,22,26], such as the best patch search method. In particular, Wexler et al. [27] and Simakov et al. [8] proposed a method based on global optimization to obtain a more consistent filling. These techniques were later accelerated by a random block search algorithm (PatchMatch) [5], allowing real-time advanced image editing. However, sample-based completion method works well only when the missing region is composed of simple structure and texture.

In recent years, deep learning has also shown outstanding results in image completion. It can fill large missing areas while preserving semantic and contextual details and capturing the high-level features of images. Recently, deep neural networks have been used in texture synthesis and image stylization [28–32]. In particular, Phatak et al. [18] trained an encoder-decoder convolutional neural network (context encoder) with L2 loss and adversarial loss [17] to directly predict lost image regions. This work predicts a reasonable image structure and requires only one forward propagation, which is very fast. Although the results are encouraging, sometimes it lacks fine texture details and creates visible artifacts around the boundaries of the missing areas. In addition, this method is also not suitable for processing high-resolution images. The adversarial network needs to adjust the network parameters inversely according to the loss between the generated image and the real image, so it is difficult to train against the loss when the input is large. In view of the shortcomings of the above methods, Chao Yang et al. [20] proposed using the encoder-decoder convolutional neural network as the global content constraint and the similarity between the missing regions and the original regions as the texture constraint, the two combined to perform image completion. Chao Yang divided the high-resolution image into several steps to enhance the authenticity of the texture information, but it was difficult to ensure the global consistency of the complemented images.

To make sure the consistency of the complemented images, Iizuka uses a global and local consistency method to complete the images [21]. First, the method generates missing regions and the corresponding binary channel mask (1 indicates the region to be completed and 0 indicates the intact region) on the original image randomly. Second, the images with missing regions and corresponding binary channel masks are input into the completion network with the mean square error loss (MSE Loss) of the

missing regions in the original image and the complementary regions in the generated image to train the completion network. Third, the generated image and the region centered on the complemented region are input into the global discriminator and the local discriminator respectively while keeping the completion network fixed. The two networks are trained by using the generated adversarial loss. At last, the three networks are trained as a whole with the combination of the three losses. Iizuka's method is better than the existing image completion methods, but there are still problems: (1) when the missing areas are large, the final completion results are blurry; (2) when the missing areas are located at the border of the images, the lack of context information and the inconsistency training speed between the completion network and the discriminator network leads to the problem of blurring and color distortion.

In addition, the networks structure adopted in Iizuka's method is the basic GAN model. One of the most obvious shortcomings of the basic GAN is that the training process is difficult to converge and has great instability. The reasons are that (1) different from the general training process by the gradient descent method, which has a clear objective function, GAN's training is a process of finding Nash equilibrium points. The decrease in the error of either side of the generator and the discriminator during the training process may causes the error of the other party to rise. It is difficult for GAN to reach a relatively balanced state, and always oscillate between various modes of generated samples, one of the most common case is that GAN maps different input samples to the same generated sample, i.e., repeatedly generates the same data. This phenomenon is also called model collapse. (2) In general, since the iterative speed of the discriminator is quicker than that of the generator, the inconsistent iteration speed of the two will leads to the instability of the GAN model, therefore it is difficult to obtain an optimal model. (3) It is prone to the phenomenon that the gradient disappears during the backpropagation process, which makes the model more difficult to train. In response to the shortcomings of GAN, the researchers derived a series of new models based on the basic GAN model, such as Conditional GAN (cGAN) [33], Deep Convolutional GAN (DCGAN) [34], Wasserstein GAN (wGAN) [35] and so on. Among them, wGAN is mainly to improve the stability of the basic GAN's training. The generation problem can be approximated as a regression problem; however, the last layer of the discriminator in the basic GAN's network structure uses the sigmoid activation function, which approximates the generation problem to the 0–1 two classification problem that causes problems, such as mode collapse and difficulty in convergence, and difficulty for the generator and the discriminator to reach an equilibrium. WGAN improves the stability of the training by removing the last sigmoid function in the last layer of the discriminators.

Combined with wGAN, this paper improves the image completion method and network structure based on the Iizuka's method. (1) On the one hand, by using the central block of the complemented region as the input of the local discriminator 2 added in this paper, the synthesis results are more realistic because the training process is back-propagated in the loss function of the central region and the corresponding region of the real image; (2) On the other hand, the training of the network structure used in Iizuka's method is unstable and difficult to converge. By "widening" the network, i.e., adding the local discriminator 2 in the original network, removing the last sigmoid activation layer of the discriminators and replacing each ReLU layer with the combination of the BN layer and Leaky_ReLU layer, this paper avoids the problems of gradient disappearance, mode collapse, and difficulty in convergence, which make the network structure more stable and the completion results more realistic and clear.

3. Method

Based on Iizuka's method [21], we randomly generate blank areas with arbitrary shape on the original image *orig* to get the input image *x* and the corresponding binary mask (Figure 2). In the training process, the MSE loss between the complementary region *c* and the corresponding region in the original image *orig* is used to train the completion network (Equation (1)). Next, an image block *d* centered on the complemented region is cropped, the entire generated image and *d* are input into the

global discriminator and the local discriminator respectively and the two networks are trained with generate adversarial loss (GAN loss) (Equation (2)). The preliminary completion result preserved the structural information of the image. The global network discriminator is used to ensure the global consistency of the image, and the local network discriminator is used to maintain the local consistency of the complementary region with other regions.

$$L(orig, M_c) = \|M_c \odot (C(orig, M_c) - orig)\|_2, \tag{1}$$

$$L_{adv} = \min_C \max_D E[\log D(orig, M_d) + \log(1 - D(C(orig, M_c), M_c))], \tag{2}$$

where C denotes the completion network, D denotes the discriminator networks, M_c denotes the input binary mask, and M_d denotes the area corresponding to the local area d centered on the missing area in the binary image.

However, when the missing areas are large or located at the edge of the image, due to the lack of context information and the instability of the training, the result of the completion of the above three networks may be blurred and the color distorted. Therefore, as shown in Figure 2, we also generate an image block which located at the center of the missing region. By inputting d' into the local discriminator network 2 and calculating the GAN loss (Equation (3)) between d' and the original image, the completion result of the central position d' of the large missing area has global and local consistency with the surrounding information.

Finally, the entire network structure is trained synchronously with the joint loss of three networks (Equation (4)), where α denotes coefficient of the proportion of the adversarial loss.

$$L_{advd'} = \min_C \max_D E[\log D(orig, M_{d'}) + \log(1 - D(C(orig, M_c), M_c))], \tag{3}$$

$$L' = \min_C \max_D E[L(orig, M_c) + \alpha \log(1 - D(C(orig, M_c), M_c))] + \alpha \log D(orig, M_d) + \alpha \log D(orig, M_{d'}) \tag{4}$$

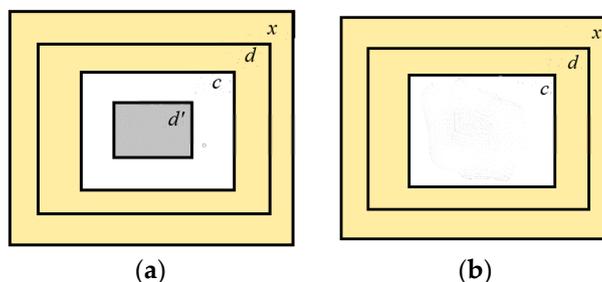


Figure 2. (a) Iizuka’s method; (b) Our method. x denotes the input image, d denotes the image block centered on the missing area in the image, c denotes the missing area, and d' denotes the image block located at the center of the missing area and having a size of 1/4 of the missing area.

The network structure adopted by our method is shown in Figure 3. The network includes a completion network, a global discriminator, a local discriminator, and a local discriminator 2. The completion network is used to generate content of the missing area, the global discriminator views global consistency of the generated image, and the local discriminator only looks at small areas centered on the complemented region to ensure local consistency of the generated image blocks. Our added local discriminator network (local discriminator 2) is used to ensure the content accuracy of the image generated in the central region of the missing region and the local consistency between the central area and the outer area.

The inputs of the completion network are the entire image with the missing areas and the corresponding binary channel mask. The output is the combined image of the intact area and the completed areas of the input image. Then, the entire generated image is input into the global network discriminator to judge the accuracy of the generated content and to maintain the global consistency of

the image. The input of the local network discriminator is a local region centered on the completed region in the generated image, which is used to maintain local consistency of the generated area with other areas. The input of the local network discriminator 2 added herein is to generate a central region of the completed region in the image for determining the accuracy of the generated content of the central region of the large-area missing regions and the local consistency with other regions.

Among them, the completion network adopts an encoder-decoder structure, which includes ten convolution layers, four dilated convolution layers, two deconvolutional layers, and one output layer. Except that the first convolutional layer uses a 5×5 convolution kernel, all other convolutional layers and dilated convolutional layers use a 3×3 convolution kernel, and the deconvolutional layer uses a 4×4 convolution kernel. In Iizuka's method, except for the last layer in the completion network, there is a ReLU layer behind each of the convolutional layers. Since the ReLU activation function ignores the effect of a negative value, the gradient of the neuron is set to 0 when its input is a negative value, causing a "neuron death" phenomenon. For this defect of ReLU, this paper replaces ReLU layers with the combination of the BN layers and the Leaky_ReLU layers. The network backpropagation process is carried out by biasing the error between the actual output and the expected output obtained during the forward propagation of the network, and then adjusting the parameters of each layer of the network. When the network layers are deep, the partial derivative approaches zero when the backpropagation is close to the input layer, which is easy to fall into local optimum. The BN operation keeps the input of each layer the same distribution in the network training process, avoiding above problems. Leaky_ReLU solves the problem that when the input is negative, the output is 0 in the ReLU activation function, avoiding the problem of the gradient disappearing.

In the original network structure, the discriminators train much faster than the generator, which causes the discriminator to easily distinguish which one is a natural image and which one is the generated image, but the gradient that is passed to the completion network in the backpropagation can only make it generates strange textures, which easily cause color distortion problems. In this paper, by "widening" the network structure, i.e., adding a local discriminator, the training speed of the two can be balanced, and the generated texture is more realistic, and pseudo-color is not generated. In addition, the last layer of the discriminators in the Iizuka's method uses the sigmoid activation function to obtain a continuous value in the interval $[0, 1]$ to represent the probability that the image is from the real sample. Since the sigmoid function approximates the image generation problem to a two-class problem, which is easy to cause the gradient disappearance, the mode collapses, and the difficulty to reach equilibrium between generator and discriminators. This paper draws on the idea of wGAN method and removes the last sigmoid activation function, making network training more stable and easier to converge, which resulting in more realistic and clear results.

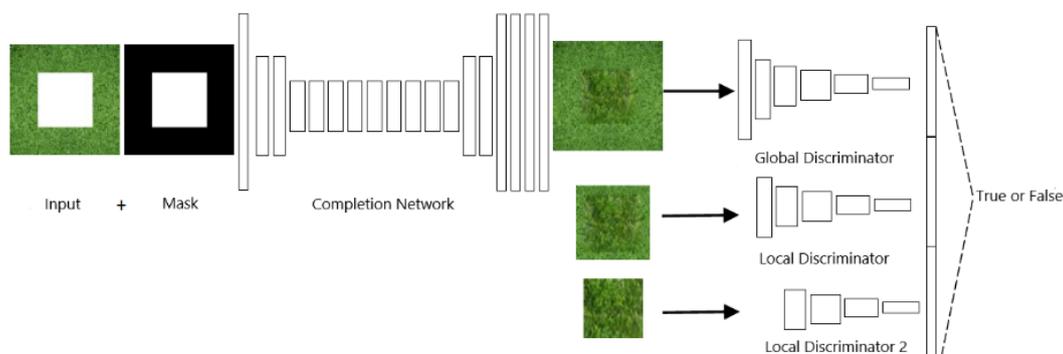


Figure 3. Overview of the network architecture we used.

The experiment is based on TensorFlow framework and implemented in python. The training process is similar to Iizuka's method, i.e., the completion network is iteratively trained T_c times; then, the three discriminator networks are trained T_d times; finally, the entire network is iteratively trained T_{train} times until the network is stable. The data set used in the experiment contains pictures of grass

downloaded from Baidu and Google Chrome image search engine, where the training set contains 1000 images, and the test set contains 120 images. The training set used in this paper is not large, only contains 1000 pictures, but for the grass-like texture with local similarity, good experimental results can be obtained, which can reflect the advantages of this method. In our experiment, the batch size is set to 16. The input of the completion network is a 256×256 picture with missing areas and a corresponding mask picture. The missing area is set to an area of 80×80 to 128×128 randomly generated at the original images. The number of iterations of the completion network is 4000 times; the input of the global discriminator which maintains the global consistency of the synthesis result is a 256×256 picture of the completion network output. The input of the local discriminator is the center of the synthesis result of the completion network. The area of 128×128 maintains the local consistency of the missing area and the context. The input of the local discriminator 2 is the 64×64 area in the center of the missing area in the synthesis result. The number of iterations of the discriminators is 1500. The completion network and the discriminators are iterated 8000 times together. The iteration required a week to process on the NVIDIA 1080Ti graphics card.

4. Results

To compare the experimental results, we tested the Iizuka method using the same parameters as our method. The experimental results are shown in Figure 4.

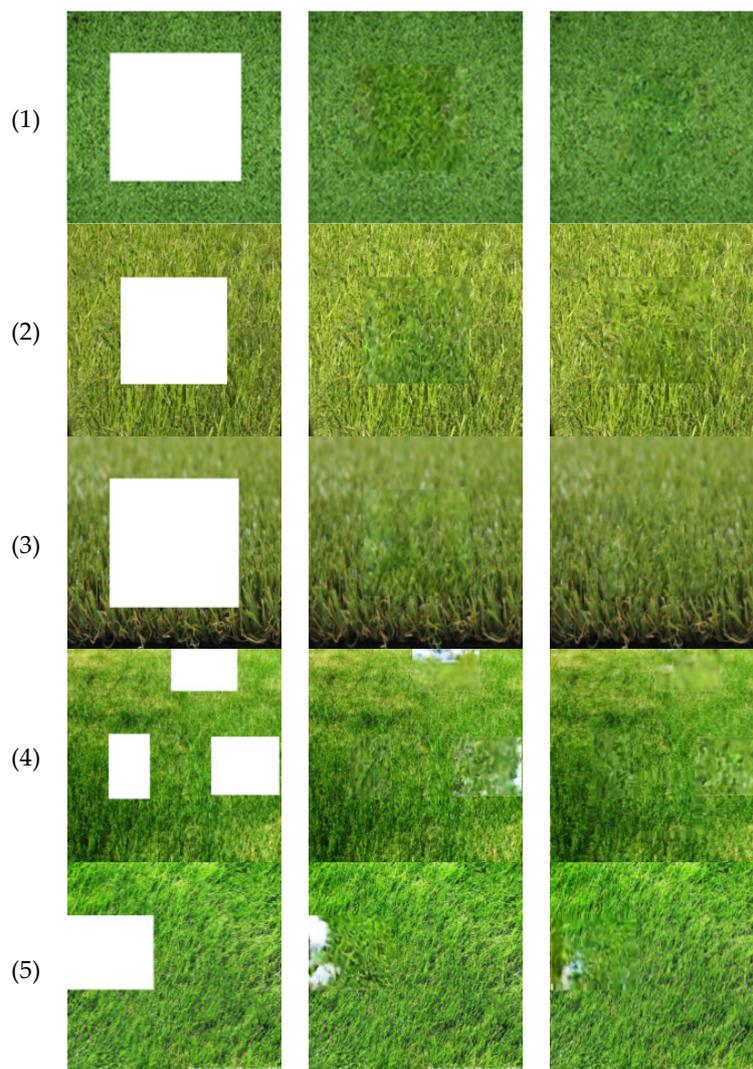


Figure 4. Cont.

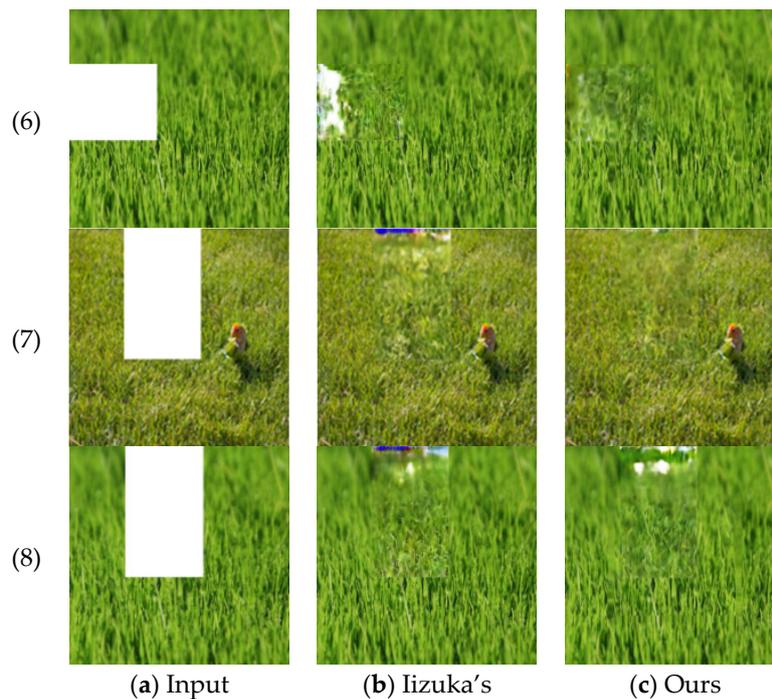


Figure 4. The result of Iizuka's (b) method and ours (c). (a) Input.

The results show that when the missing area is large (the width and length are both more than 50% in this case) or located at the border of the image, the results of the completion of the Iizuka's method are ambiguous. In Figure 4, the size of the input images of (1) to (3) is 256×256 pixels, wherein the missing area is an area of 128×128 size located at the center of the image, i.e., 1/4 size of the input image. It can be seen from the figures that the completion results of Iizuka's method will be blurred, distorted, etc., especially the difference at the boundary of the missing area is obvious. In this paper, both the global and local consistency of the completed image are considered. At the same time, the central region of the missing region in the preliminary completion result is input into the local discriminator 2 for confrontation training to guarantee the authenticity of the texture information, so the results of this paper are more clear and realistic; The input image of (4) contains a plurality of missing regions, wherein the missing regions are randomly distributed at different positions of the image. Due to the lack of context information, the completion result of the Iizuka's method will have obvious blank areas.

In addition, when the missing area is large and located at the edge of the image, as shown in (5) to (8), the Iizuka's method's completion results contains pseudo-color in addition to the large-area blank. This is because the iteration speed of the discriminator in the Iizuka's method is quicker than the generator, which causes the instability training of generated adversarial network. We measured the running time of the discriminators and generator in the Iizuka's method and found that the generator running time is 50% slower than the discriminators average, which causes the discriminator to easily distinguish which one is a natural image and which one is the generated image. The gradient only makes the generator generates strange textures that produce pseudo-colors during backpropagation. In this paper, the local discriminator 2, i.e., the "widening" network, is added to balance the training speed of the generator and discriminators. In addition, we also replace the ReLU layer behind each convolution layer with the combination of BN layer and Leaky_ReLU layer, and remove the sigmoid activation function of the last layer of the discriminator, avoiding gradient disappearance and pattern collapse and other problems during network training, which makes the network training more stable. Experiments show that the difference between the running speed of our network and Iizuka's method network is only 10%. This improvement makes the network training more stable and easier to converge, and overcomes the defects of generating pseudo-color in Iizuka's method, making the result more

realistic. In Figures 4–7, our synthesis results are clearer and more realistic than Iizuka’s method, but due to the lack of context information, blank areas still appear. This is a problem that we still need to improve.

We use a combination of MSE and structural similarity (SSIM) to evaluate the similarity of the complemented images and the original images. Among them, the MSE is used to compare the absolute error between image pixels, i.e., the difference between pixels, and the SSIM is used to evaluate the internal dependence between image pixels, i.e., human visual perception on images. The smaller the MSE and the greater the SSIM are, the more similar the images are. Figure 5 shows the MSE result pairs and Figure 6 shows the SSIM result pairs of the Iizuka’s method and our method on images of Figure 4. The red line denotes our method’s results and the green line denotes Iizuka’s method’s results.

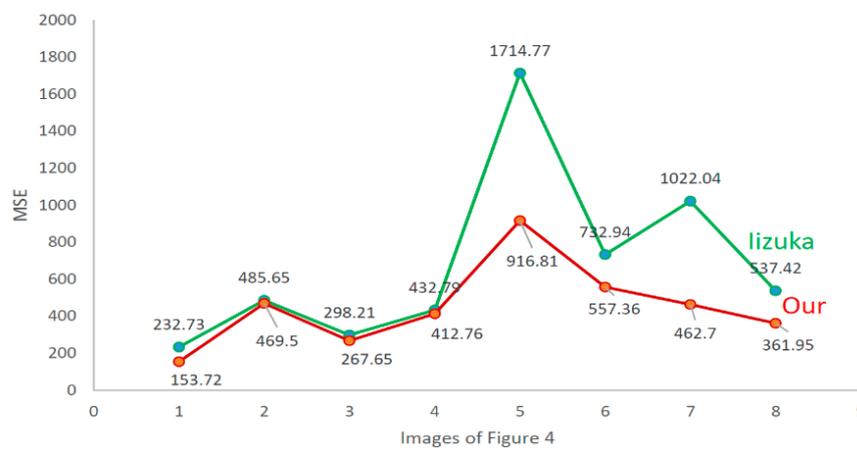


Figure 5. The MSE result pairs of our method and Iizuka’s method on images of Figure 4.

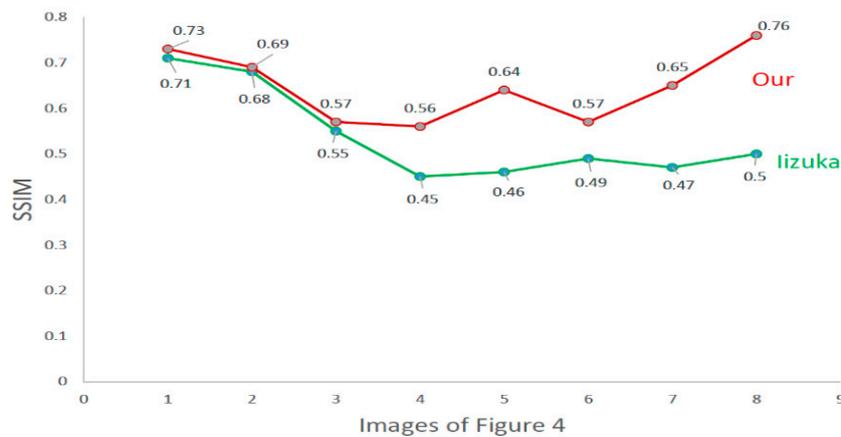


Figure 6. The SSIM result pairs of our method and Iizuka’s method on images of Figure 4.

As can be seen from Figure 5, when the missing area are large, the MSE results of Iizuka’s method are higher than our method, which indicate that our method guarantees the realistic of texture information when completing large missing areas. As shown in 4 to 8 of the horizontal coordinate in both Figures 5 and 6, when the missing area are located at the border of the image, both difference of MSE and SSIM between the results of Iizuka’s method and our method with the original images have significant difference, which indicate that the method of this paper has better completion results than Iizuka’s method in context information and global and local consistency, and the completed images are closer to the original images.

Figure 7 shows the operation of each loss in the training process of the network structure using the ReLU layer and the sigmoid activation function in the last layer of the discriminator in Iizuka’s method. Figure 8 shows the operation of each loss in the training process of the network structure

using the combination of BN layer and leak_relu layer which replaces the ReLU layer, and the sigmoid activation function of the last layer is removed, also, the local discriminator 2 is added.

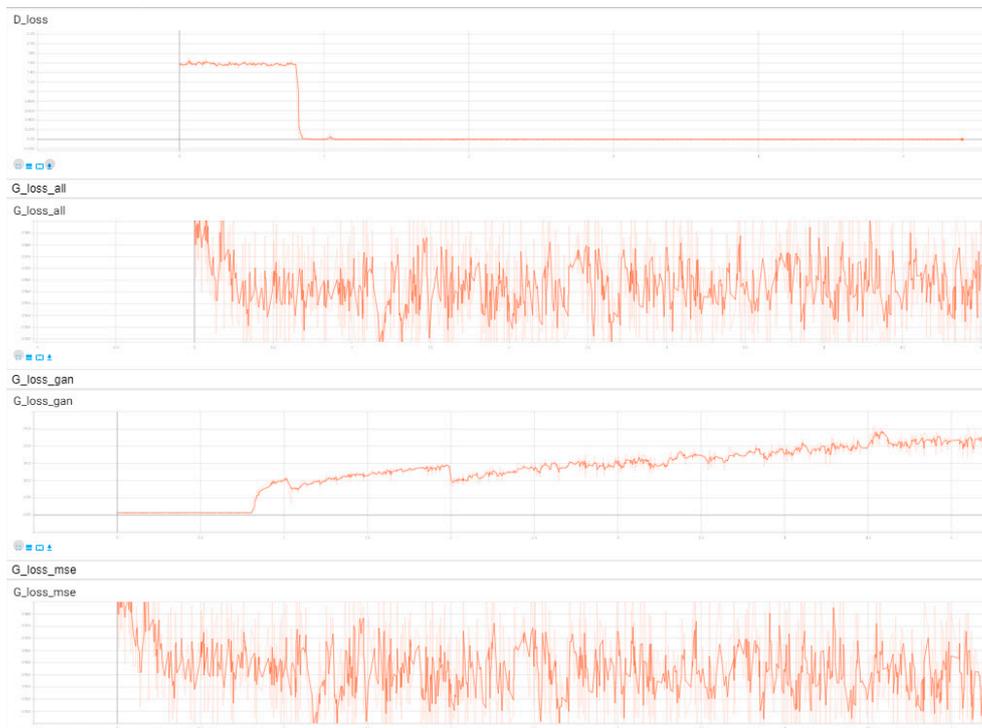


Figure 7. The losses graph of the Iizuka’s network structure.

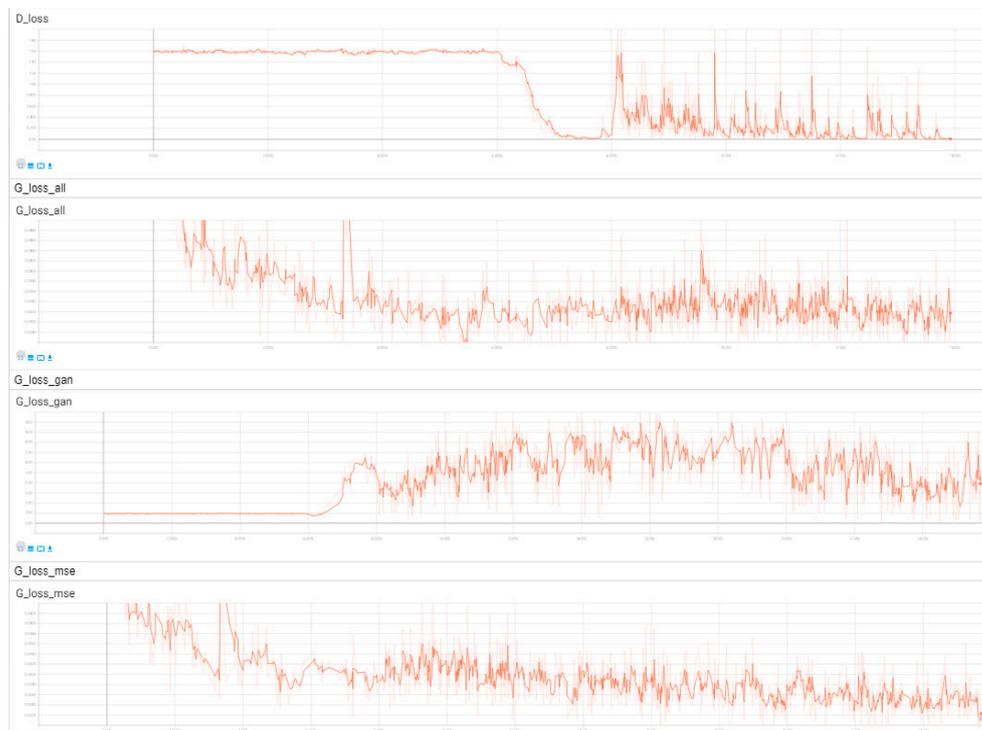


Figure 8. The losses graph of this method’s network structure.

It can be seen from the figures that the gradient disappearance phenomenon is easy to occur during the training of the network structure of Iizuka’s method, and the training process is very

unstable. The training process of the network structure adopted in this method is relatively stable, and when the generation loss decreases, the adversarial loss shows an upward trend, and the two always maintain the confrontation state until the balance is reached.

The following is another set of experimental results. We used the same parameter settings as the grass to train and complete the image of the barbed wire with certain structural features. The result of the completion is shown in Figure 9.

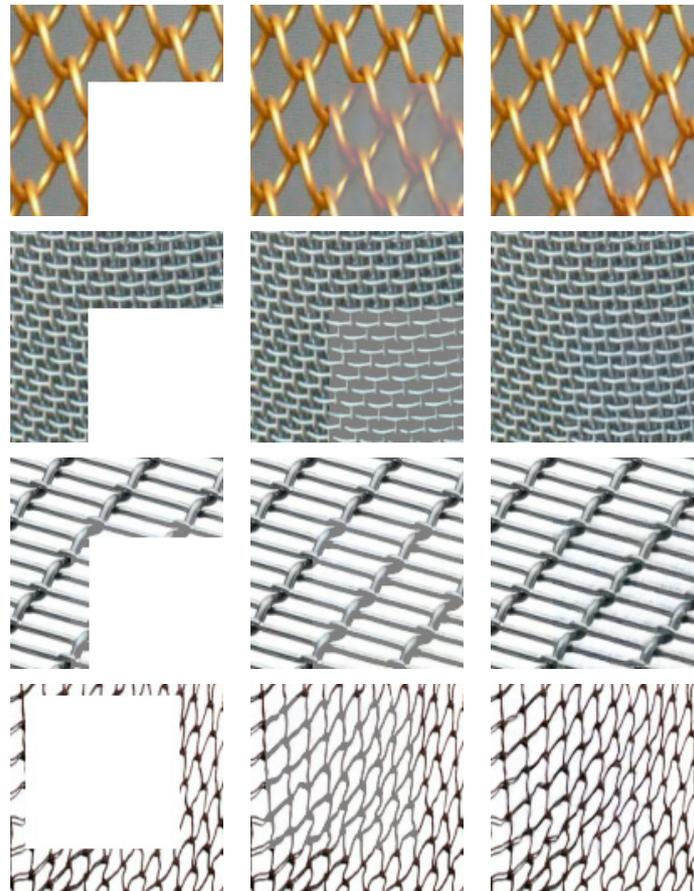


Figure 9. The results comparison of Iizuka's method and ours with barbed wire images.

In addition, to verify the generalization ability of the proposed method, we randomly extracted 15,000 images from the Places2 dataset for training. The test results are shown in Figure 10.



Figure 10. *Cont.*

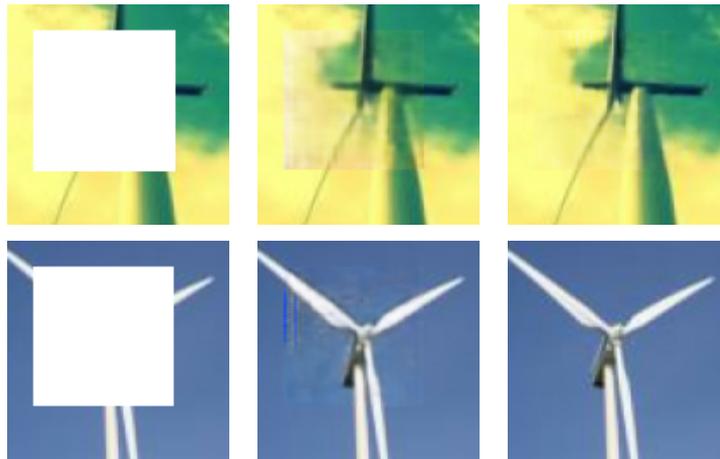


Figure 10. The results comparison of Iizuka's method and ours with Places2 dataset.

It can be seen from the figures that the complemented area of the image generated by Iizuka's method are very blurred, and the boundary areas are more obvious, which cannot maintain the authenticity and consistency of the image well. Our method has improved the network structure, so that the network avoids the gradient disappearing during the training process, and reduces the inconsistency of the generator and discriminator iteration speed, which prove that the method is able to generate clearer, more realistic images and guarantee the globally and locally consistency of the images.

5. Conclusions

In this paper, we propose an image completion method with large or edge-missing areas. The method divides the image into the region centered on the missing region and the region at the center of the missing region. The local discriminator 2 is added to complete the image, and the GAN model used in the Iizuka's method is improved. On the one hand, it overcomes the problem of lacking context information due to the large missing area which thereby generates more realistic texture details; on the other hand, it makes the adversarial generate network training more stable and suppresses the generation of pseudo-color. The method considers both the global consistency and local consistency of the image when the completion step is performed, and the authenticity and consistency of the texture information of the large-area and edge-missing regions are also considered.

However, there are still limitations of our method. Since there is a lack of context information, the results of our method still have small blank areas when the large missing area is located at the border of the image. Whether it is possible to match the appropriate context information in the image library will be an interesting way to explore.

Author Contributions: J.J. and G.Y. contributed to analysis and manuscript preparation; J.J. and G.Y. conceived and designed the experiments; J.J. and G.Y. performed the data analyses and wrote the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by [Fundamental Research Funds for the Central Universities] grant number [No. 2015ZCQ-XX] and [National Key Research and Development Project] grant number [No. 2017YFC0504404].

Conflicts of Interest: We declare that no conflict of interest exists in the submission of this manuscript, and manuscript is approved by all authors for publication.

References

1. Jia, J.; Tang, C. Image repairing: Robust image synthesis by adaptive nd tensor voting. In Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Madison, WI, USA, 18–20 June 2003.
2. Rares, A.; Reinders, M.J.; Biemond, J. Edge-based image restoration. *IEEE Trans. Image Process.* **2005**, *14*, 1454–1468. [[CrossRef](#)] [[PubMed](#)]

3. Huang, J.; Kang, S.B.; Ahuja, N.; Kopf, J. Image Completion Using Planar Structure Guidance. *ACM Trans. Graph.* **2014**, *33*, 1–10. [[CrossRef](#)]
4. Whyte, O.; Sivic, J.; Zisserman, A. Get Out of my Picture! Internet-based Inpainting. *Br. Mach. Vis. Conf.* **2009**, *2*, 5.
5. Krähenbühl, P.; Koltun, V. Efficient inference in fully connected crfs with gaussian edge potentials. *Adv. Neural Inform. Process. Syst.* **2011**, *24*, 104–117.
6. Barnes, C.; Shechtman, E.; Finkelstein, A.; Goldman, D.B. Patch-Match: A Randomized Correspondence Algorithm for Structural Image Editing. *ACM Trans. Graph.* **2009**, *28*, 24.
7. Darabi, S.; Shechtman, E.; Barnes, C.; Goldman, D.B.; Sen, P. Image Melding: Combining Inconsistent Images using Patch-based Synthesis. *ACM Trans. Graph.* **2012**, *31*, 82. [[CrossRef](#)]
8. Simakov, D.; Caspi, Y.; Shechtman, E.; Irani, M. Summarizing visual data using bidirectional similarity. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA, 23–28 June 2008; pp. 1–8.
9. Wexler, Y.; Shechtman, E.; Irani, M. Space-Time Completion of Video. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 463–476. [[CrossRef](#)]
10. Hays, J.; Efros, A.A. Scene completion using millions of photographs. *ACM Trans. Graph.* **2007**, *26*, 4. [[CrossRef](#)]
11. Zhu, Z.; Huang, H.Z.; Tan, Z.P.; Xu, K.; Hu, S.M. Faithful Completion of Images of Scenic Landmarks using Internet Images. *IEEE Trans. Vis. Comput. Graph.* **2015**, *22*, 1945–1958. [[CrossRef](#)]
12. Drori, I.; Cohen-Or, D.; Yeshurun, H. Fragment-based image completion. *ACM Trans. Graph.* **2003**, *22*, 303–312. [[CrossRef](#)]
13. Roth, S.; Black, M.J. Fields of experts. *Int. J. Comput. Vis.* **2009**, *82*, 205–229. [[CrossRef](#)]
14. Hu, Y.; Zhang, D.; Ye, J.; Li, X.; He, X. Fast and accurate matrix completion via truncated nuclear norm regularization. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 2117–2130. [[CrossRef](#)] [[PubMed](#)]
15. Wilczkowiak, M.; Brostow, G.J.; Tordoff, B.; Cipolla, R. Hole filling through photomontage. *BMVC* **2005**, *5*, 492–501.
16. Komodakis, N.; Tziritas, G. Image completion using efficient belief propagation via priority scheduling and dynamic pruning. *IEEE Trans. Image Process.* **2007**, *16*, 2649–2661. [[CrossRef](#)]
17. Pouget-Abadie, I.J.G.; Mirza, M. Generative adversarial nets. *arXiv* **2014**, arXiv:1406.2661.
18. 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, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 2536–2544.
19. Yeh, R.; Chen, C.; Lim, T.Y.; Hasegawa-Johnson, M.; Do, M.N. Semantic image inpainting with perceptual and contextual losses. *arXiv* **2016**, arXiv:1607.07539.
20. 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, Honolulu, HI, USA, 21 July–26 July 2017.
21. Iizuka, S.; Simo-serra, E.; Ishikawa, H. Globally and Locally Consistent Image Completion. *ACM Trans. Graph.* **2017**, *36*, 107. [[CrossRef](#)]
22. Bertalmio, M.; Vese, L.; Sapiro, G. Simultaneous Structure and Texture Image Inpainting. *IEEE Trans. Image Process.* **2001**, *10*, 1200–1211.
23. Efros, A.A.; Leung, T.K. Texture synthesis by non-parametric sampling. In Proceedings of the IEEE Computer Society, Kerkyra, Greece, 20–27 September 1999.
24. Efros, A.A.; Freeman, W.T. Image quilting for texture synthesis and transfer. In Proceedings of the Conference on Computer Graphics and Interactive Techniques, Los Angeles, CA, USA, 12–17 August 2001; pp. 341–346.
25. Kwatra, V.; Schödl, A.; Essa, I.; Turk, G.; Bobick, A. Graphcut Textures: Image and Video Synthesis Using Graph Cuts. *ACM Trans. Graph.* **2003**, *22*, 277–286. [[CrossRef](#)]
26. Criminisi, A.; Perez, P.; Toyama, K. Region Filling and Object Removal by Exemplar-based Image Inpainting. *IEEE Trans. Image Process.* **2004**, *13*, 1200–1212. [[CrossRef](#)]
27. Li, C.; Wand, M. Combining markov random fields and convolutional neural networks for image synthesis. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 2479–2486.

28. Gatys, L.A.; Ecker, A.S.; Bethge, M. A neural algorithm of artistic style. *arXiv* **2016**, arXiv:1508.06576. [[CrossRef](#)]
29. Gatys, L.A.; Ecker, A.S.; Bethge, M. Texture synthesis and the controlled generation of natural stimuli using convolutional neural networks. In Proceedings of the Bernstein Conference, Heidelberg, Germany, 14–17 September 2015.
30. Champandard, A.J. Semantic Style Transfer and Turning Two-Bit Doodles into Fine Artworks. *arXiv* **2016**, arXiv:1603.01768.
31. Ulyanov, D.; Lebedev, V.; Vedaldi, A.; Lempitsky, V.S. Texture networks: Feed-forward synthesis of textures and stylized images. *Int. Conf. Mach. Learn.* **2016**, *1*, 4.
32. Johnson, J.; Alahi, A.; Fei-Fei, L. Perceptual losses for real-time style transfer and super-resolution. In *European Conference on Computer Vision*; Springer: Cham, Switzerland, 2016; pp. 694–711.
33. Mirza, M.; Osindero, S. Conditional generative adversarial nets. *Comput. Sci.* **2014**, *21*, 2672–2680.
34. Radford, A.; Metz, L.; Chintala, S. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. *arXiv* **2015**, arXiv:1511.06434.
35. Arjovsky, M.; Chintala, S.; Bottou, L. Wasserstein GAN. *arXiv* **2017**, arXiv:1701.07875.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).