Symmetry-Aware Face Completion with Generative Adversarial Networks

Jiawan Zhang[1], Rui Zhan[1], Di Sun[1,2(✉)], and Gang Pan[1]

[1] Tianjin University, Tianjin, China
dsun@tju.edu.cn
[2] Tianjin University of Science and Technology, Tianjin, China

Abstract. Face completion is a challenging task in computer vision. Unlike general images, face images usually have strong semantic correlation and symmetry. Without taking these characteristics into account, existing face completion techniques usually fail to produce a photo-realistic result, especially for the missing key components (e.g., eyes and mouths). In this paper, we propose a symmetry-aware face completion method based on facial structural features using a deep generative model. The model is trained with a combination of a reconstruction loss, a structure loss, two adversarial losses and a symmetry loss, which ensures pixel faithfulness, local-global contents integrity and symmetrical consistency. We conduct a dedicated symmetry detection technique for facial components and show that the symmetrical attention module significantly improves face completion results. Experiments show that our method is capable of synthesizing semantically valid and visually plausible contents for the missing facial key parts from random mask. In addition, our model outperforms other methods for detail completion of facial components.

Keywords: Face completion · GAN · Symmetry · Image inpainting

1 Introduction

Face completion, also known as face inpainting, is an important topic in computer vision and image processing. It refers to the task of filling missing pixels or removing unwanted parts from a face image. It is often used in conjunction with face recognition [31–33] and face editing [4,24].

The core challenge of face completion lies in synthesizing visually realistic and semantically plausible pixels for the missing areas that are coherent with the existing contents in a face image. However, the semantics of facial components are related to each other and do not exist independently. This makes the face completion task significantly more difficult than general image inpainting or completion.

This work was supported by Tianjin Philosophy and Social Science Planning Program under grant TJSR15-008.

C. V. Jawahar et al. (Eds.): ACCV 2018, LNCS 11364, pp. 289–304, 2019.
https://doi.org/10.1007/978-3-030-20870-7_18

(a) Input (b) GFC[15] (c) GntIpt[30] (d) Our (e) Target

Fig. 1. From left to right: (a) input image with mask. (b) inpainted result of GFC [15]. (c) inpainted result of GntIpt [30]. (d) inpainted result of our method. (e) target image.

Early face completion methods [5,6,10,19,34] are mainly based on texture synthesis or patch matching. These approaches based on texture synthesis work well on images with small missing or meshed occlusion. But they are unsuitable for situations with large areas occlusion. Besides, the approaches based on patch matching often require a reference image or database to complete the repair, which makes it impossible to obtain better repair results when similar semantic patches are not found in the reference image or database. Recently, great progress has been made in deep learning, especially in generative adversarial networks (GANs) [9]. At the same time, learning-based face completion methods [7,11,15, 17,28,30] have also achieved remarkable results. These methods are suitable for images with small or large missing areas which may include individual semantic components.

However, the above methods mostly ignore the property of symmetry for the facial components. The symmetry does not only refer to the shape similarity between the symmetric components. It refers to the overall feature similarity which includes many factors, such as color, texture, shape, etc. Neglecting the symmetry of the face in face completion methods may cause great inconsistency in the completion result. For example, when half the face is occluded (see Fig. 1(a)), the symmetrical components, such as eyes, eyebrows and nose, are all missing a part. In this case, the existing generative face completion (GFC) method [15] can only maintain the shape feature of the semantic components but cannot guarantee the consistency of the texture and other details between symmetrical semantic components. As shown in Fig. 1(b), there is a great asymmetry between the completed left eye and the existing right eye. The generative image inpainting method with contextual attention (GntIpt) [30] can borrow or copy texture and color information from known background patches to generate the missing parts, but it cannot ensure the shape similarity of symmetrical components. In Fig. 1(c), it still causes a huge gap between eyes. Therefore, the consideration of the symmetry in face image is the key to achieve better face completion.

In this paper, we propose a symmetry-aware face completion method based on deep learning. We adopt a generative adversarial networks as our completion model. The input is a face image with a random mask. First, the encoder of the generator is utilized to extract the structural features of the input image,

and then a decoder is employed to generate a completion result based on the structural features. To optimize the completion of the generator, we design three adversarial losses and a structure loss to regularize the training process. Among these, commonly used global adversarial loss and local adversarial loss [11] are adopted to constrain the global semantic integrity and local semantic consistency with surrounding pixels of the completion result. The symmetry adversarial loss is posed to constrain the symmetry of the completion result and the structure loss is used to guide the generation of completion image. By adopting the above mechanism, the proposed model can not only repair large missing area, but also can well maintain the symmetry of facial components, as shown in Fig. 1(d).

To summarize, our contributions are as follows: (1) We propose a deep generative face completion model. A U-Net network is utilized to generate the completion result and three additional networks are adopted to optimize the completion model. (2) Considering the symmetry characteristics of human face, we introduce a strategy for symmetrical components detection of face images. (3) We design a new loss function, of which a symmetry loss is added to constrain the symmetry and consistency between symmetrical facial components, and a structure loss is adopted to guide the structure generation of the completion result.

2 Related Work

A variety of different approaches have been proposed for the image completion task. Traditionally, image completion mainly uses diffusion-based or patch-based methods. Diffusion-based inpainting [3,8] uses smoothness priors via parametric models or partial differential equations (PDEs) to propagate the low-level features around the target holes to the interior of the hole. It can be naturally applied to small regions inpainting, but are not suitable for large missing areas cases. Patch-based methods [2,10,34] differ from the diffusion-based methods, they repair the missing regions by searching similar patches from the background or image database. Since the patch-based methods require external patch to fill in the unknown regions, it fails if similar patches do not exist in the surrounding region or database.

Recently, with the rapid development of deep learning, many learning-based approaches [14,16,22,26] have emerged and produced remarkable achievements. At first, learning-based methods can only work on small and thin damage. Later, Goodfellow et al. propose a new network architecture - GANs [9], which can be effectively applied to image completion. The completion methods based on GANs [11,17,20,27,28] achieve good completion results.

There are also some researchers who focus on face completion and have achieved remarkable research results [5–7,11,17,19,28,34]. Here we'd like to highlight two representative work. One is proposed by Li et al. [15] which adopts an encoding-decoding generator and two discriminators (global discriminator and local discriminator) to generate the completion result with truly global and local content. Motivated by this, we also adopt this similar network architecture in our work. In addition, in order to ensure the semantic structure similarity

between completion result and ground-truth, it also utilizes a semantic parsing network. This method can effectively generate completion results with similar structure to the ground-truth, but it lacks the detailed constraints on semantic components. Hence, there is still a great inconsistency of color or texture for symmetrical facial components.

Later, Yu et al. [30] pose a generative image inpainting method with contextual attention (GntIpt). This method integrates the advantages of the traditional patch-based methods and the current learning-based methods. It can solve the inconsistency of color and texture between the completed patches and the background pixels, but it lacks structure and shape constraints for facial components. Therefore, there is also a great asymmetry when completing symmetrical facial components.

To sum up, the above methods can achieve good performance for face completion, but they do not fully consider the symmetry of the face, so there are some distortions or inconsistencies when processing face images. Hence, in our work we need to take symmetrical facial components into account and specialize in face completion technique.

3 Proposed Method

Given a masked image, our goal is to synthesize the missing contents that are both semantically consistent with the whole object and visually realistic. Our method is based on generative adversarial networks trained for the face completion task. A generator is used for the image completion. Three additional networks, the global, local and symmetry discriminator networks, are used in order to train this network to realistically complete images. An overview of our method can be seen in Fig. 2.

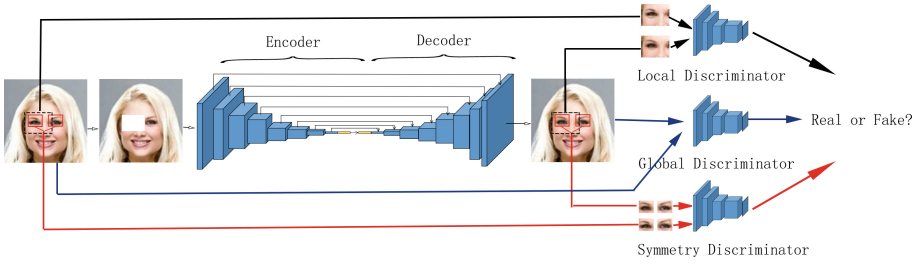


Fig. 2. Network architecture. It consists of one generator and three discriminators. The global discriminator and local discriminator are learned to distinguish the whole generated image and synthesize contents in the mask as real and fake. The symmetry discriminator is to further ensure the newly generated contents symmetric and encourage consistency between new and old pixels. (Color figure online)

3.1 Generator

Our approach is motivated by Generative Adversarial Networks (GANs) [9]. It trains two networks, a generator G, and a discriminator D. G is designed as an autoencoder to construct new contents given input images with missing regions, while D plays an adversarial role, discriminating between the image generated from G and the real image.

In this work, we adopt a "U-Net" network [23] as the generator which follows an encoder-decoder structure similar to the one used in [12]. The input of the generator is an RGB image with a binary mask (1 for a pixel to be completed), and the output is an RGB completed image. The encoder employs a down sampling process to extract the structural features of the masked input image, and then the decoder utilizes an up sampling process to gradually fill the contents of the missing area according to the structural features extracted by the encoder. It is worth mentioning that the skip connection of the U-Net network can keep more information for the generation of unmasked part. This eliminates the need to replace the unmasked area with original image in subsequent processing. In order to get more spatial support to generate the masked area, we add two dilated convolution layer [29] between the encoder and decoder which are represented by the yellow rectangle in Fig. 2.

3.2 Discriminator

The generator can be trained to fill the masked area with small reconstruction and structure losses. However, it does not ensure that the filled area is visually realistic and natural. As shown in Fig. 4(c), the generated pixels are quite fuzzy and only capture the coarse shape of missing face components. To encourage more photorealistic results, we design three discriminators, that is, global discriminator, local discriminator and symmetry discriminator. These three discriminators' structure is similar to [12] which uses batch normalization layer and LeakyReLU layer after each convolution layer.

The global discriminator is designed to determine the faithfulness of an entire image, so the inputs of it are the whole completion result and the ground-truth. The local discriminator follows the same pattern, except for the purpose of checking whether the generated contents of the missing area are real or not. It takes the 8-pixels dilation of the masked areas from the completion result and the ground-truth as inputs. The inputs of 8-pixels dilation around the masked area can make our model fully exploit the spatial correlation between adjacent pixels. With the global and local discriminators' further optimization, the completion result can be more realistic with less fuzzy artifacts.

In order to maintain the symmetry of the completion result, we further utilize the symmetry discriminator to optimize the generation of symmetrical face components. The inputs of the symmetry discriminator are the symmetrical components detection results taken from the completion image and the corresponding ground-truth as inputs. With the unmasked part of the symmetrical components as a condition [18], the completion of corresponding components

can be further optimized. Here the key of this issue is twofold: one is what these so-called symmetrical components should be, and the other is how to detect these symmetrical components for face images. We will introduce them in next subsection.

3.3 Symmetry Detection for Face Components

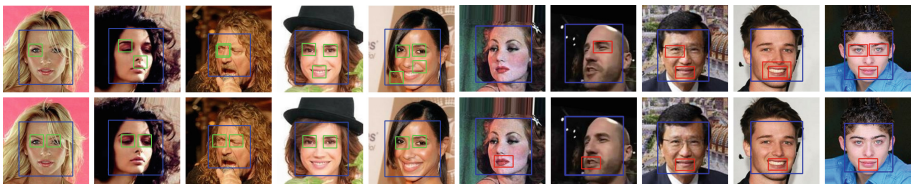


Fig. 3. Results of symmetrical components detection. The first row are the eye and mouth detection results by using the Haar cascade method [25] directly. The second row are the corresponding detection results after optimization using our detection strategy.

The face contains five components, eyes, eyebrows, nose, mouth and ears. The geometrical position of these components is relatively fixed and each can be seen as symmetrical, for example, the eyes are in pairs, and the nose itself is also bilaterally symmetrical. In this paper, we collectively refer to them as symmetrical components. Since the ears are always unpaired in most of the face images, we ignore the ears during the actual detection and symmetric optimization process. In addition, we also treat the eyebrows as parts of the eyes. To sum up, during the detection and optimization process, the symmetrical components involve only the eyes (eyebrows), nose and mouth.

As symmetrical components are important feature description of the human face, we need to detect them at first. We use the Haar feature cascade classifiers proposed in [25] as our detection method. Although the Haar cascade method can improve its detection accuracy by adjusting the "minNeighbors" parameters, there are still some false detections, see the first row in Fig. 3. Therefore, we adopt an optimization method to improve the detection accuracy. Firstly, we recognize the face and detect the symmetrical components to be optimized within the facial region, this can improve the detection rate efficiently. Next, we exclude the wrong detections based on the detection number and the relative position of the symmetrical components in the face image. Finally, we set a default value by averaging the past accurate detections for the wrong detection.

As the wrong detections exclusion is the most important part, we give more details about this filtering strategy. For each individual symmetrical component (mouth, nose), the correct detection number is one. For paired symmetrical components (eyes, eyebrows, ears), the correct detection number is two. We set four position parameters [x1, x2, y1, y1] to define the position of each symmetry component. x1, x2 represent the minimum and maximum horizontal positions,

and y1, y2 represent the minimum and maximum vertical positions. Based on the detection number and position range of components, we can exclude three types of errors. The first one is that the detection number is less than the correct detection number. (See in the first and sixth columns of Fig. 3). The second one is that the detection number is equal to the correct detection number, but the component is not within the correct position range. (See in the second and seventh columns of Fig. 3). For both cases, we use the default value to replace the detection value. The third one is that the detection number is greater than the correct detection number (See in the fourth, fifth and ninth and tenth columns of Fig. 3). In this case, we exclude the detections that the components are not within the correct position range and randomly select the one that the components are within the correct position range as the final detection result. In this detection process, the position parameters need to be manually set by experience and the others are all automatic.

With our optimization method, the detection accuracy can be significantly improved. The second row of Fig. 3 show the detection results after optimization.

3.4 Loss Functions

We first introduce a per-pixel reconstruction loss $\mathcal{L}_r$ to the generator, which is the $L1$ distance between the network output and the ground-truth image. The reason we use $L1$ loss is that $L2$ loss penalizes outliers heavily and tends to generate blurry contents. The $\mathcal{L}_r$ loss is defined as below:

$$\mathcal{L}_r(x, M) = \|G(x, M) - x\|_1 \tag{1}$$

where x is the ground-truth, M is the mask image, G represents the generator.

In order to make the generated image have a high structural similarity with the ground-truth, we then adopt a structure loss $\mathcal{L}_s$. The structure loss is similar to the perceptual loss used in [16]. Instead of using an additional VGG-16 network, we directly use the generator to obtain the structure loss. We take both the masked image and the ground truth as input. Note that the feature maps of the masked image at each layer of the decoder should be consistent with the corresponding feature maps of the ground-truth. The structure loss $\mathcal{L}_s$ is defined as

$$\mathcal{L}_s(x, M) = \sum_{k=1}^{N} \alpha_k \|\Psi_k(x, M) - \Psi_k(x)\|_1 + \beta \|G(x) - x\|_1 \tag{2}$$

Here, the former item computes the $L1$ distances between the feature maps of the raw output image and the ground truth image. It can improve the structural similarity between the output and the ground truth. Ψ_k is the activation map of the kth selected layer, α_k is the corresponding weight. The latter item is a constraint to guarantee the output of the ground-truth is consistent with itself. β is the weight of the constraint.

We further employ the global and local adversarial loss. They both adopt the sigmoid cross-entropy loss but have different inputs. The inputs for the global

discriminator are the overall completion result and the ground-truth, and for the local discriminator, the inputs are a patch with 8-pixel dilation around the completed region and the corresponding ground truth. The global adversarial loss $\mathcal{L}_g$ and local adversarial loss $\mathcal{L}_l$ are defined by:

$$\mathcal{L}_g(x, M) = \min_G \max_D \mathbb{E}\left[\log D(x) + \log(1 - D(G(x, M))\right] \tag{3}$$

$$\mathcal{L}_l(x, M) = \min_G \max_D \mathbb{E}\left[\log D(x, M) + \log(1 - D(G(x, M), M)\right] \tag{4}$$

The symmetry discriminator network also works as a kind of loss, called the symmetry loss. The symmetry loss is used to optimize the completion of symmetrical components, which includes symmetry adversarial loss and symmetry pixel loss. The symmetry adversarial loss adopt the sigmoid cross-entropy adversarial loss, but it takes the symmetrical components of the ground-truth and the completion result as inputs. The symmetry pixel loss is the L1 distance between the symmetrical components of the ground-truth and completion result. If the symmetrical components are in pairs, such as eyes, eyebrows and ears, we use P_l to represent the left part of the symmetrical components and P_r to represent the right part. When P_l is masked and P_r is unmasked, the symmetry loss function is defined as:

$$\mathcal{L}_{sym}(x, M) = \gamma \mathcal{L}_{sym_a}(x, M) + \mathcal{L}_{sym_b}(x, M) \tag{5}$$

$$\mathcal{L}_{sym_a}(x, M) = \min_G \max_D \mathbb{E}\left[\log D(P_l, P_r) + \log(1 - D(\bar{P}_l, P_r)\right] \tag{6}$$

$$\mathcal{L}_{sym_b}(x, M) = \left\|\bar{P}_l - P_l\right\|_1 \tag{7}$$

Here, we set $\gamma = 0.1$ in our experiments. $\bar{P}_l$ is the left part of the generated symmetrical components in completion result. If P_r is unmasked while P_l is masked, the loss function needs to replace $\overline{P_l}$ with $\overline{P_r}$. If the missing component is a single symmetrical component, such as nose or mouth, we simply input the single component as a whole and have no need to divide it into two parts. The simplified formulation is changed as below:

$$\mathcal{L}_{sym_a}(x, M) = \min_G \max_D \mathbb{E}\left[\log D(P) + \log(1 - D(\bar{P})\right] \tag{8}$$

$$\mathcal{L}_{sym_b}(x, M) = \left\|\bar{P} - P\right\|_1 \tag{9}$$

P and $\bar{P}$ represent the single symmetrical component of the ground truth and the completion result, respectively. By using the symmetry loss, we can use the known part of the symmetrical components as a constraint to generate the unknown part, this will make the completion result have better symmetry and consistency.

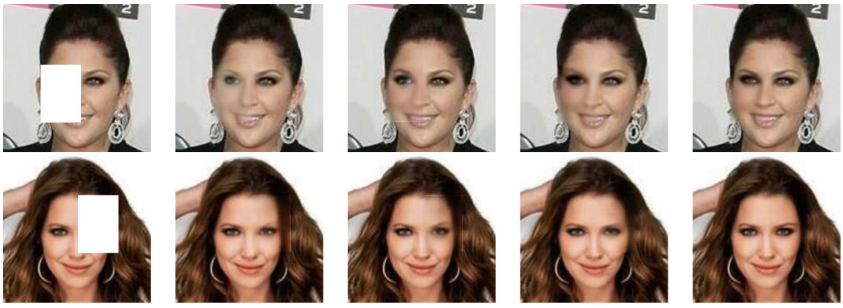
The overall loss function is defined by:

$$\mathcal{L} = \omega_1 \mathcal{L}_r + \omega_2 \mathcal{L}_s + \omega_3 \mathcal{L}_g + \omega_4 \mathcal{L}_l + \omega_5 \mathcal{L}_{sym} \tag{10}$$

where ω_i is the weight to balance the effects of different losses. We set $\omega_1 = 100$, $\omega_2 = 5$, $\omega_3 = 1$, $\omega_4 = 1$, $\omega_5 = 1$ in our experiments.

3.5 Training

The training process can be scheduled in three stages. Firstly, we train the network using the reconstruction loss and structure loss to get the initial result. Secondly, we utilize the global adversarial loss and local adversarial loss to fine-tune the network so as to obtain the completion results with less fuzzy. Finally, the symmetry loss is incorporated to optimize the completion of symmetrical components. For the third stage, we need to optimize all the symmetrical components in turn.



(a) Input (b) L1 (c) Stage1 (d) Stage2 (e) Stage3

Fig. 4. Comparison results of different training stages. (a) is the input, (c) shows the first stage with both reconstruction loss and structure loss, (d) displays the second stage with additional global and local adversarial loss, (e) indicates the last stage with additional symmetry loss, (b) are the results obtained by only using reconstruction loss.

Figure 4 shows the experimental results of different stages. Figure (a) are inputs, Figure (c–e) are respectively the outputs of each progressive stage with different optimization. Each stage is an optimization of its previous stage. As shown in Fig. 4, the first stage produces a fuzzy initial result with a white border. While the second stage, after the optimization of the global discriminator and the local discriminator, the blurring is reduced and the white border is eliminated, but the paired symmetrical components such as eyes and eyebrows are still asymmetrical. In the third stage, the completion results look more symmetrical and natural by using the symmetry loss. Figures (b) shows the results obtained by only using reconstruction loss with the same training number as the first phase. The comparison of the (b) and (c) reveals that the structure loss can guide the generator to produce completion results with similar structure to the ground-truth. In order to improve the stability of training, we adopt the training procedure proposed by [21] and use Adam [13] to optimize. The initial learning rate is 0.00005.

4 Experimental Results

4.1 Datasets

We use the CelebA dataset to train and test our model. The CelebA dataset contains 202,599 face images and each image is cropped and rescaled to $256 \times 256 \times 3$ pixels for data normalization. We randomly select 2599 images for testing and 200,000 images for training. To ensure our network can complete missing contents with different sizes and positions, we generate a random mask for the input image with the mask size between $[30 \times 30, 130 \times 130]$ during training. Besides, the position of the mask is also randomly selected.

4.2 Qualitative Results

Figure 5 shows some results on the test dataset. In each test image, the mask covers at least one key facial components. The first row shows the examples with large square-like mask from which more than two facial components are occluded. In these examples, any of the facial components covered by the mask are relatively complete. For example, both eyes or the entire nose is blocked. For this case, the model can generate pleasing results even without the symmetrical components optimization.

In the second and third row, only one part of the paired symmetrical components is missing, then the symmetry optimization is required so as to obtain

Fig. 5. Face completion results on the CelebA test dataset. In each panel from left to right: masked inputs, our completion results, ground truth images.

symmetrical and natural completion results. It is worth noting that the symmetry optimization not only has a better optimization effect on the geometrically symmetric facial components, but also has a better optimization effect on the facial components with a certain view angle. The fourth row displays some results for individual symmetrical component. Taking lips as examples, our method is capable of maintaining both the shape symmetry and the texture consistency. Note that the color of the lips is well recovered. The fifth row shows the completion results with multiple masks, and the sixth row exhibits the results with irregular masks. For both cases, Our method can get photo-realistic and pleasing results.

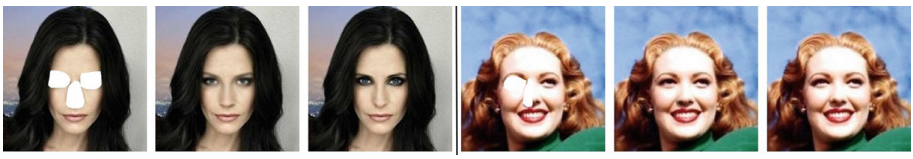


Fig. 6. Face completion results with irregular masks. In each panel from left to right: masked inputs, our completion results, ground truth images.

Figure 6 shows more results with irregular masks. In a real-world scenario, our method can perform face completion by replacing irregular masks with multiple rectangular masks, and also can give completion results that are consistent with their contextual semantics.

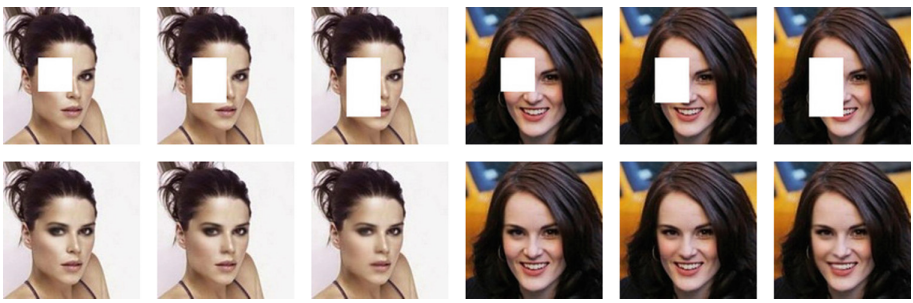
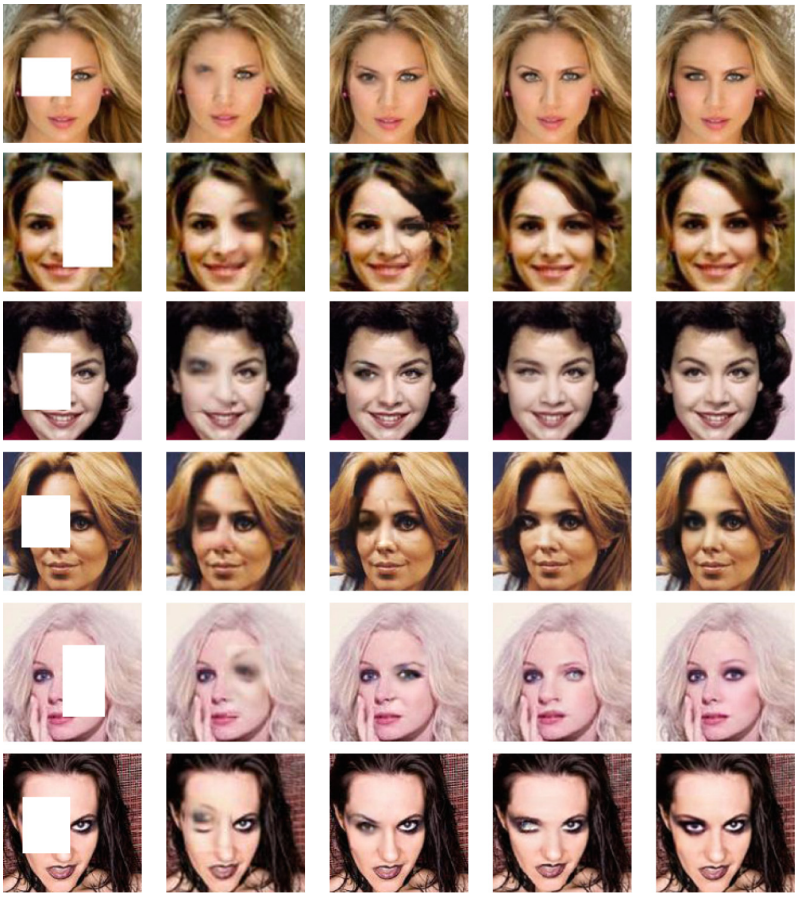


Fig. 7. Completion results with different mask sizes: 64×64, 84×64, and 120×64.

We also give some completion examples with different mask sizes in Fig. 7. In these examples, each image is processed by three different masks whose sizes are respectively 64×64, 84×64, and 120×64. As can be seen from the results, the completion effect does not change significantly as the mask increases.

Overall, our completion algorithm is competent for images with missing symmetrical facial components or large missing area, or partially/completely corrupted by the masks with different shapes and sizes.



(a) Input (b) GL[11] (c) GFC[15] (d) GntIpt[30] (e) Ours

Fig. 8. Qualitative comparison with different methods.

4.3 Comparison with the State of the Art

We compare our model with some current learning-based methods [11,15,30] in Fig. 8.

As can be seen from the results, the GL [11] method can identify the contents to be completed, but the completion results have large semantic distortion. The GFC [15] method can maintain the shape similarity of semantic components to some extent, but there are still inconsistencies in colors and textures for symmetrical components. The GntIpt [30] method can effectively alleviate the inconsistency of color and texture between symmetrical components, but it cannot guarantee the shape symmetry. Note that our method can make up for the above deficiencies and obtain natural and consistent results.

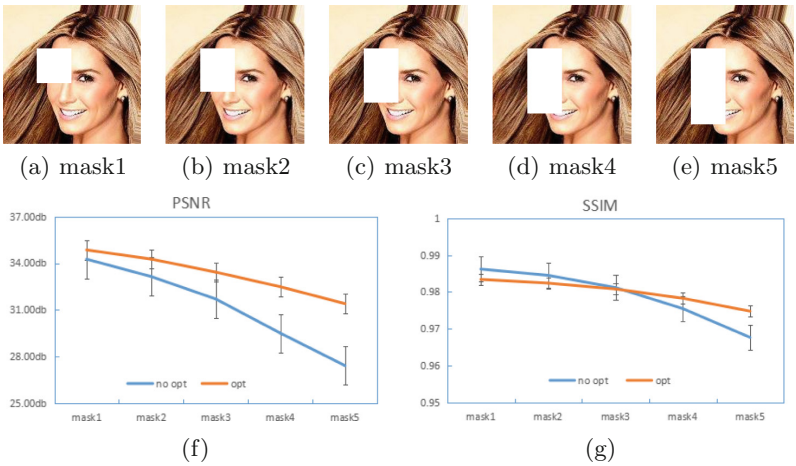


Fig. 9. Quantitative comparison. (a–e) represent five masks which size are 64×64, 64×80, 64×100, 64×120, 64×140. (f) is the result of the PSNR evaluation. (g) is the result of the SSIM evaluation. In (f) and (g), the blue line represents the evaluation curve obtained without symmetry optimization, the orange line represents the evaluation curve obtained using symmetry optimization. (Color figure online)

4.4 Quantitative Comparison

A quantitative comparison experiment is adopted to evaluate the effectiveness of the symmetry optimization. In this quantitative comparison experiment, two metrics, PSNR and SSIM are used for evaluation. In this experiment, we only focus on the completion of the symmetry components, so we use five masks as shown in Fig. 9(a–e) to occlude the symmetric area. The evaluation results of PSNR and SSIM are shown in Fig. 9(f–g). It can be seen that the completion results with symmetry optimization have higher PSNR values, and the PSNR curve decreases relatively slowly when the mask size increases. As for the SSIM evaluation, the results without optimization have a higher structural similarity when the mask size is small. But as the mask size increases, the value drops faster. When the input images have a large mask size, the symmetry optimization can make the completion results have higher structural similarity. Through the quantitative comparison, we can conclude that symmetry optimization has a better improvement effect on face completion.

4.5 Limitations and Discussion

Though our model performs well at completing various images, it has some limitations. The mask size used in this paper is between $[30 \times 30, 130 \times 130]$. When it is close to or less than the floor, the generated patch will be inconsistent with the surrounding pixels in color (see in Fig. 10(b)). To handle this problem, we use a modified Poisson blending method (MPB) proposed in [1] for post-processing. However, post-processing can only alleviate this problem but can

(a) Input (b) Output (c) MPB Result (d) Input (e) Output (f) target

Fig. 10. Model limitation. (a) is the input with small mask. (b) is the directly result of the proposed method. (c) is the MPB [1] optimization of (b). (d) is the input with large mask. (e) is completion result after optimization. (f) is the target image.

not eliminate it completely, as shown in Fig. 10(c). In addition, when the mask size larger than 130×130, the completion result will become more and more blurred (see in Fig. 10(e)). If the mask is too large to provide enough context information, it will be impossible to generate pleasant completion results. If we want to completely solve those problem, the lower and upper limits of the mask size should be set to $[0 \times 0, 256 \times 256]$, but meanwhile it will result in an increase in training time.

5 Conclusions

In this paper, we have proposed a deep generative network for face completion. The network is based on a GAN, with an structure-constraint autoencoder as the generator, two adversarial loss functions (local and global) and a symmetry correction as the discriminators. Our method automatically detects symmetrical facial components and completes them without any user interaction. We show that the symmetrical attention module significantly improves face completion results for the missing facial key parts from random mask. We provide in-depth comparisons with existing approaches and show semantically valid and visually plausible completion results.

References

1. Afifi, M., Hussain, K.F.: MPB: a modified poisson blending technique. Comput. Vis. Media **1**(4), 331–341 (2015)
2. Barnes, C., Shechtman, E., Finkelstein, A., Goldman, D.B.: PatchMatch: a randomized correspondence algorithm for structural image editing. ACM Trans. Graph. (ToG) **28**(3), 24 (2009)
3. Bertalmio, M., Sapiro, G., Caselles, V., Ballester, C.: Image inpainting. In: Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques, pp. 417–424. ACM Press/Addison-Wesley Publishing Co. (2000)
4. Bitouk, D., Kumar, N., Dhillon, S., Belhumeur, P., Nayar, S.K.: Face swapping: automatically replacing faces in photographs. ACM Trans. Graph. (TOG) **27**, 39 (2008)

5. Deng, Y., Dai, Q., Zhang, Z.: Graph Laplace for occluded face completion and recognition. IEEE Trans. Image Process. **20**(8), 2329–2338 (2011)
6. Deng, Y., Li, D., Xie, X., Lam, K.M., Dai, Q.: Partially occluded face completion and recognition. In: 2009 16th IEEE International Conference on Image Processing (ICIP), pp. 4145–4148. IEEE (2009)
7. Dolhansky, B., Ferrer, C.C.: Eye in-painting with exemplar generative adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7902–7911 (2018)
8. Elad, M., Starck, J.L., Querre, P., Donoho, D.L.: Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA). Appl. Comput. Harmonic Anal. **19**(3), 340–358 (2005)
9. Goodfellow, I., et al.: Generative adversarial nets. In: Advances in Neural Information Processing Systems, pp. 2672–2680 (2014)
10. Hays, J., Efros, A.A.: Scene completion using millions of photographs. ACM Trans. Graph. (TOG) **26**, 4 (2007)
11. Iizuka, S., Simo-Serra, E., Ishikawa, H.: Globally and locally consistent image completion. ACM Trans. Graph. (TOG) **36**(4), 107 (2017)
12. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. arXiv preprint (2017)
13. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
14. Köhler, R., Schuler, C., Schölkopf, B., Harmeling, S.: Mask-specific inpainting with deep neural networks. In: Jiang, X., Hornegger, J., Koch, R. (eds.) GCPR 2014. LNCS, vol. 8753, pp. 523–534. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-11752-2_43
15. Li, Y., Liu, S., Yang, J., Yang, M.H.: Generative face completion. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), vol. 1, p. 3 (2017)
16. Liu, G., Reda, F.A., Shih, K.J., Wang, T.C., Tao, A., Catanzaro, B.: Image inpainting for irregular holes using partial convolutions. arXiv preprint arXiv:1804.07723 (2018)
17. Liu, P., Qi, X., He, P., Li, Y., Lyu, M.R., King, I.: Semantically consistent image completion with fine-grained details. arXiv preprint arXiv:1711.09345 (2017)
18. Mirza, M., Osindero, S.: Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784 (2014)
19. Mo, Z., Lewis, J.P., Neumann, U.: Face inpainting with local linear representations. In: BMVC, vol. 1, p. 2 (2004)
20. Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T., Efros, A.A.: Context encoders: feature learning by inpainting. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2536–2544 (2016)
21. Radford, A., Metz, L., Chintala, S.: Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434 (2015)
22. Ren, J.S., Xu, L., Yan, Q., Sun, W.: Shepard convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 901–909 (2015)
23. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28

24. Saito, Y., Kenmochi, Y., Kotani, K.: Estimation of eyeglassless facial images using principal component analysis. In: Proceedings of the 1999 International Conference on Image Processing, ICIP 1999, vol. 4, pp. 197–201. IEEE (1999)
25. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2001, vol. 1, p. I. IEEE (2001)
26. Xie, J., Xu, L., Chen, E.: Image denoising and inpainting with deep neural networks. In: Advances in Neural Information Processing Systems, pp. 341–349 (2012)
27. Yang, C., Lu, X., Lin, Z., Shechtman, E., Wang, O., Li, H.: High-resolution image inpainting using multi-scale neural patch synthesis. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), vol. 1, p. 3 (2017)
28. Yeh, R.A., Chen, C., Lim, T.Y., Schwing, A.G., Hasegawa-Johnson, M., Do, M.N.: Semantic image inpainting with deep generative models. In: CVPR, vol. 2, p. 4 (2017)
29. Yu, F., Koltun, V.: Multi-scale context aggregation by dilated convolutions. arXiv preprint arXiv:1511.07122 (2015)
30. Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X., Huang, T.S.: Generative image inpainting with contextual attention. arXiv preprint (2018)
31. Zhang, S., He, R., Sun, Z., Tan, T.: Multi-task convnet for blind face inpainting with application to face verification. In: 2016 International Conference on Biometrics (ICB), pp. 1–8. IEEE (2016)
32. Zhang, S., He, R., Sun, Z., Tan, T.: DeMeshNet: blind face inpainting for deep MeshFace verification. IEEE Trans. Inf. Forensics Secur. **13**(3), 637–647 (2018)
33. Zhang, W., Shan, S., Chen, X., Gao, W.: Local Gabor binary patterns based on Kullback–Leibler divergence for partially occluded face recognition. IEEE Signal Process. Lett. **14**(11), 875–878 (2007)
34. Zhuang, Y.T., Wang, Y.S., Shih, T.K., Tang, N.C.: Patch-guided facial image inpainting by shape propagation. J. Zhejiang Univ.-SCIENCE A **10**(2), 232–238 (2009)