Identity Preserving Face Completion with Landmark based Generative Adversarial Network

Zou Xinyi[1], **Jiang Runqing**[1], **Hou Tianxiang**[1], **Yang Hao**[1]

[1] Fujian Key Laboratory of Sensing and Computing for Smart City,
School of Informatics, Xiamen University,
Xiamen, China

Abstract

Given an occluded or partial face image, face completion concentrates on filling the missing area with semantical-aware pixels. With the rapid development of deep learning, the learning-based method becomes a trend in the inpainting field. Though improve dramatically, the results still fail to capture the identity consistency with the original image. Hence, they are limited to actual scenarios. For the sake of tackling the aforementioned issue, we are dedicated to firstly taking a review of the existing relevant approaches and then introducing a novel identity preserving loss to generate visually pleasing and plausible face images on the premise of identity consistency. Our plan can not only be applied to face editing tasks, but also facilitate the accuracy of occlusion face recognition.

Keywords Face completion, Image synthesis, Generative adversarial network, Identity preserving

Introduction

In general, human beings are expert in hallucinating the unknown region and determining the identity given an incomplete face image. In contrast, it is rather challenging for automatic face completion. There is nothing ambiguous that a large occlusion will reduce the visual effect and damage the identity information simultaneously.

Numerous research has been conducted in this field. Traditional image completion methods are mainly based on low-level cues, such as the color or gradient of the nearest pixel or patch (Bertalmio et al. 2000; Esedoglu and Shen 2002; Barnes et al. 2009; Darabi et al. 2012; Simakov et al. 2008). Nevertheless, these approaches suffer from repetitive and blurry patterns and are limited to only small missing area.

Recently, the learning-based method, especially the GAN-based method plays the dominant role in the image completion area. Pathak et al. proposed an encoder-decoder network called Context Encoder (Pathak et al. 2016), which for the first time introduced an adversarial loss (Goodfellow et al. 2014) to generate more realistic results without supervision. Subsequently, effective architectural components (e.g. global and local discriminator (Iizuka, Simo-Serra, and Ishikawa 2017), contextual attention layer (Yu et al. 2018)) have been adopted to various GAN-like networks to enhance the fidelity and details of the completion results.

However, those aforementioned methods are not sufficient to face completion task since the face images usually contains unique topology structure and attribute consistency. Specific to face completion, face-related prior knowledge like face parsing, edge or landmark is introduced and achieve great success (Li et al. 2017; Zhang et al. 2019; Nazeri et al. 2019; Song et al. 2019). Motivted by (Yang et al. 2019), we also adopt facial landmark to guarantee the overall structure consistency of the generated images.

Though the results improve substantially, we notice that there exists two problems remaining to be solved. Firstly, the inpainting results depend heavily on the size of the missing area. Many methods don't work when it comes to the large occlusion. Secondly, they fail to preserve the identity information of the original image so that they are limited to the actual scenarios. For example, the criminals may deliberately obscure the key facial feature to avoid being recognized by the monitor. Plus, during the CONVID-19 epidemic period, people are forced to wearing mask and many face recognition systems become invalid. Under those circumstances, reconstructing the deteriorated region with the original identity information is of great significance.

Related Work

Image Inpainting A multitude of image inpainting works have been proposed recently. They are widely used in many intriguing circumstances, such as image restoration and object removal (Criminisi, Pérez, and Toyama 2004; Xie, Xu, and Chen 2012). The most representative traditional branches are diffusion-based (Bertalmio et al. 2000; Esedoglu and Shen 2002) and patch-based methods (Barnes et al. 2009; Darabi et al. 2012; Simakov et al. 2008). However, these non-parametric approaches can not work properly in large occlusion, especially those containing unique patterns.

Recently, deep learning based algorithms have played a dominant role in image inpainting. With the ability of capturing semantic features, the generated results are typically plausible. The first remarkable work is an encoder-decoder based network named Context Encoder (Pathak et al. 2016). Together with the standard reconstruction loss, an adversarial loss (Goodfellow et al. 2014) is adopted to ensure that

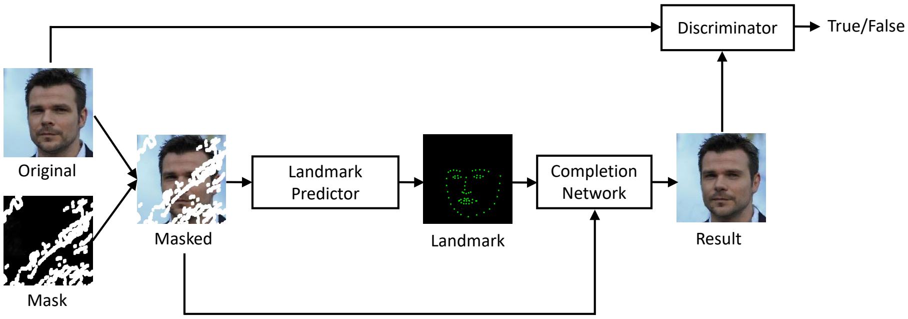

Figure 1: The architecture of our model.

the learned image representation is semantically meaningful. However, it fails to cope with arbitrary input and generate novel objects, which is tackled by Iizuka et al (Iizuka, Simo-Serra, and Ishikawa 2017). They use dialted convolution to build a fully convolutional network and add a local discriminator along with the global discriminator.

Deep Face Completion With special topological structure and complex texture, face images are much sensitive to artifacts. Therefore, the general inpainting algorithms often result in blurry results and semantic inconsistency. Specific to face completion, face prior knowledge or structure-relevant restrictions are deployed.

For example, Li et al. (Li et al. 2017) added a pretrained parsing network to regularize the generation networks. In order to share parsing information during the training process, Zhang et al. (Zhang et al. 2019) construct a two-stage network, which regards face parsing as a guidance to image completion. Similarly, edge mask, as a better representation is introduced to reconstruct more reasonable structure in (Nazeri et al. 2019). However, the performance descends dramatically when the prediction itself is inaccurate on large occlusion. Hence, landmark is an alternative due to its compactness, sufficiency and robustness (Yang et al. 2019). Other geometry-aware properties like reflectional symmetry are also investigated in face completion (Li et al. 2018).

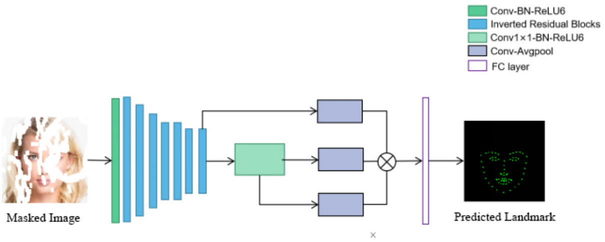

Figure 2: Landmark Predictor

Method

Given an occluded face image or only part of a human face, we aim at hallucinating the missing area with the corresponding identity. For the sake of the special geometric features, landmark is introduced as a guidance for the completion task due to its robusty and compactness. Overall, our model consists of two parts, a landmark-based generator and a discriminator that verifies the fidelity and consistency of the generated results. We will elaborate the architecture details in the following subsections.

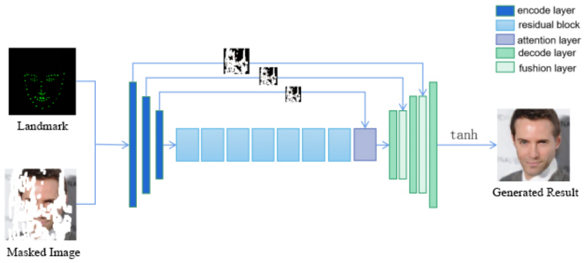

Figure 3: Completion Network

Generator

Differing from the natural scenery, faces are strictly constrained by its special geometry structure. Therefore, face prior knowledge is recommended to serve as a guidance of the completion task. Motivated by (Yang et al. 2019), we adopt landmark to act as the structural indicator. Considering the expensive cost of manual annotation, we first predict the landmark automatically. On account of the fact that inaccurate prediction will deteriorate the completion results, we use the pre-trained landmark prediction module of (Yang et al. 2019), which is built upon the MobileNet-V2 (Sandler et al. 2018) to extract the image feature. The deteailed architecture is shown in Figure. 2. After the prediction module, the masked input I^M is then mapped into its corresponding landmark, denoted as $\hat{L}$.

Together with the exploited prior $\hat{L}$, the non-occulusion image I and its binary mask M (the known region is set to 0 while the unknown region is set to 1) are fed into the generator G to synthesis a completed face image. The final output $\hat{I}$ is obtained by replacing the generated unmasked region with the known pixel in the original image:

$$\hat{I} = G(I, M, \hat{L}) \odot M + I \odot (1 - M) \tag{1}$$

As shown in Figure 3, the generator G is designed in an encoder-decoder architecture with three encoding blocks, a bottleneck of seven residual blocks with dilated convolutions, a long-short term attention block and three gradually up-sampled decoder blocks. Notice that the corresponding decoder and encoder blocks are also connected to utilize the past information.

Discriminator

The generator is trained to fill the content of the occluded area, but the results are usually blurry and artificial. To better enhance the credibility and fidelity of the generated results, a discriminator is deployed. The model finally converges when the discriminator cannot distinguish the synthesized image from the ground truth.

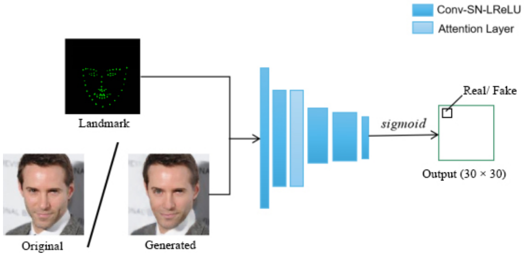

Figure 4: Discriminator Network

More specifically, we set up a local discriminator to judge the authenticity of the generated image. Since landmark can well control the overall structure of the generated image, we do not use another global discriminator like (Iizuka, Simo-Serra, and Ishikawa 2017). In particular, to reduce the obvious artifacts around the boundary of the filled region, we input the mask simultaneously to pay more attention to the junction between the generated and original image. With the help of the input mask, more plausible results are synthesized without any post-processing methods. As demonstrated in Figure 4, the discriminator is built upon the same architecture as PatchGAN (Isola et al. 2017) to better enhance the local fidelity of the generated result.

Objective Function

The training loss is composed of a reconstruction loss, a perceptual loss, an identity preservation loss, and an adversarial loss. Constrained by all these four losses, the model is capable to generate plausible completion results with the same identity.

Reconstruction Loss First, the mean square error (MSE) is adopted as a standard of the per-pixel wise difference between completed image and the real one. The equation is defined as bellow:

$$L_r = \frac{1}{n_m} \sum_{i=1}^{n_m} (I - \hat{I})^2 \tag{2}$$

where n_m denotes the input mask size other than the entire image. Since the difference only relevant to the unknown region, the penalty should also be adjusted by the input mask automatically.

Reconstruction loss is of significance to stabilize the training procedure and limit the filled region, but often suffers from producing ambiguous results.

Perceptual Loss Perceptual loss (Johnson, Alahi, and Fei-Fei 2016) represents the distinction of high-level information between images. It is believed to ensure the semantic similarity, such as texture or category resemblance. By minimizing the difference between the generated and real image feature map, which is extracted from a pre-trained network, high-frequency details can be generated. The function is defined as:

$$L_p = \sum_l \left\| \phi_l(G(I, M, \hat{L})) - \phi_l(I) \right\|_2^2 \tag{3}$$

where $\phi_l(x)$ denotes the output of the l-th layer in the pre-trained network. Here, we choose the ... layer in VGG-19 model (Simonyan and Zisserman 2014) trained on ImageNet (Deng et al. 2009).

Identity Preservation Loss For identity consistency, we use the pre-trained identity authentication network to obtain the feature vector first and define the identity loss as:

$$L_{id} = \frac{1}{n} \sum_{i=1}^{n} \left\| f(\hat{I}) - f(I) \right\| \tag{4}$$

where $f(x)$ is the FaceNet (Schroff, Kalenichenko, and Philbin 2015), which directly maps the face images to a compact Euclidean space. It captures the salient features and facial structure of the face image and is proved efficient in several face relevant tasks. Therefore, the embedded feature is feasible to calculate identity similarity.

Adversarial Loss As a two-player game, the generator does it utmost to deceive the discriminator, while the discriminator concentrates on distinguishing the generated image from the real one. The initial adversarial loss (Goodfellow et al. 2014) is:

$$\begin{aligned} L_{adv} = \min_G \max_D & E_{I \in P_r(I)}[(log(D(I, M, I))] \\ & + E_{\hat{I} \in P_g(\hat{I})}[log(1 - D(\hat{I}, M, I))] \end{aligned} \tag{5}$$

Experiments

Dataset and Settings

We plan to conduct our experiment on the public available CelebA (Liu et al. 2015) and the high-quality CelebA-Hq (Karras et al. 2017) dataset. It was provided by the Chinese University of Hong Kong and has been widely used in face-related computer vision tasks. Considering the time consumption and inpainting quality of our model, we trained it on the progressive CelebA-Hq dataset which has 30,000 images with higher resolution. We also tested out model on the original CelebA dataset to display various inpainting results.

The aim of the project is to complete the occluded image with the original semantic and identity information. Therefore, we divide the evaluation metric into two categories, namely the image quality evaluation, and the identity consistency evaluation. To measure the quality of the generated image, we choose the peak signal to noise ratio (PSNR) and structural similarity (SSIM). In terms of identity consistency, similarity distance of ArcFace toolbox is used as an objective index.

In the selection of the mask, the irregular random mask and random block mask provided in the part conv paper (Liu et al. 2018) are used in training. In the test process, in order to simultaneously measure the applicability of the model for image completion and its external expansion, the experiment selected three types of masks, namely standard masks, irregular masks, and real world occlusion simulations.Among them, the standard masks include the half mask for left and right occlusion (Mask 1,2), and the quarter mask (Mask 3-7) is covering the four corners of the image (Mask 7-12 represents the corresponding occlusion without damage the background). The irregular masks represent the random graffiti mask, and the real mask mainly simulates the occlusion of the face mask.

In breif, we compare our results under different iterations and compare the the original model. The description of the comparison models are exhibited in Table 1.

Table 1: Model description and their numbers

Model number	Iteration	Model description
1	54500	Our model
2	230900	Our model
3	600000	Our model
4	899800	Our model
5	970000	Our model
6	970000	w/o identity loss

Qualitative Results

Figure 5 shows part of the result of our model.

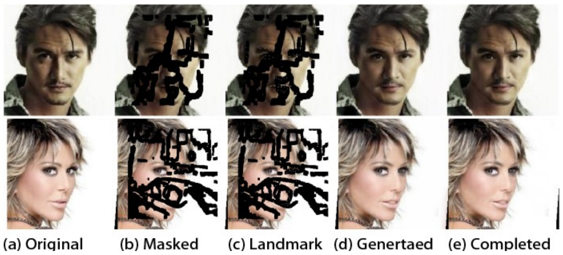
(a) Original (b) Masked (c) Landmark (d) Genertaed (e) Completed

Figure 5: Our Result

It is vividly shown that as the iteration increases, the quality of the generated images has been significantly improved. We also compare our results with the state of art method by using the open source repository or the online demo website. The results can be seen in Figure 6. Considering the identity and details, our model is better to some extent.

Table 2: Changes in the accuracy of face recognition on the standard mask under different iteration times

	Mask1	Mask2	Mask3	Mask4	Mask5	Mask6
1	58.35%	56.95%	99.37%	96.76%	98.85%	95.99%
2	74.77%	76.64%	99.57%	98.31%	99.37%	98.47%
3	80.23%	81.29%	99.71%	98.51%	99.55%	98.78%
4	82.34%	82.34%	99.75%	98.79%	99.64%	98.92%
5	**83.15%**	**83.28%**	99.75%	**98.99%**	**99.77%**	**98.95%**
6	83.04%	81.53%	**99.82%**	98.81%	99.59%	98.85%

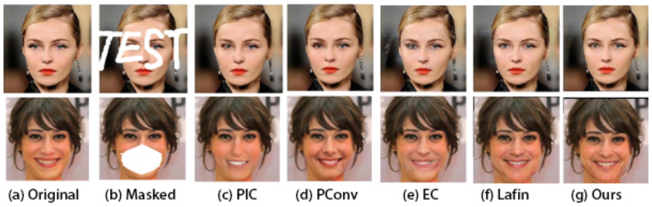
(a) Original (b) Masked (c) PIC (d) PConv (e) EC (f) Lafin (g) Ours

Figure 6: Comparing results

Quantitative Comparison

Table 3: PSNR with different models on different masks

	Mask1	**Mask2**	**Mask3**	**Mask4**	**Mask5**	**Mask6**
1	13.922	13.427	20.637	20.323	20.81	20.597
2	14.339	14.267	21.519	21.056	21.387	21.185
3	15.319	14.238	22.176	21.868	21.615	21.513
4	15.707	15.528	22.33	21.932	22.603	21.994
5	**16.396**	**16.353**	22.34	**22.173**	**23.098**	**22.142**
6	13.876	14.602	**22.481**	21.374	22.375	21.66
	Mask7	**Mask8**	**Mask9**	**Mask10**	**Mask11**	**Mask12**
1	24.031	23.417	26.218	28.91	25.569	28.782
2	24.76	24.141	27.115	30.131	26.599	30.069
3	24.295	23.741	26.794	29.833	26.419	29.615
4	24.961	24.185	27.674	30.682	26.932	30.506
5	**25.082**	24.398	**27.704**	**30.905**	26.919	**30.622**
6	24.849	**24.45**	27.584	30.598	**27.162**	30.453

First of all, we evaluate the image quality of the generated result with the metric PSNR in Table3. It can be seen from the table that the quality of the generated image has been significantly improved with the increase of the number of iterations. Even in the case of 50% occlusion, the PSNR of the generated image can still have an average value of 24.74, and the completion result is better than that without identity loss training.

It is obvious that our method has the ability of identity preserving. As shown in Table 4, our method can reach 99.91% face recognition under Mask 11. Ablation study between experiment setting 5 and 6 demonstrates that our method considers the identity information which leads to better performance.

In order to eliminate the possible influence of the same pixel on the identity features, the comparison image is modified to another reference identity image . Further experiments are conducted on the accuracy of face recognition. Table 5 show the ratio of accurate recognition of the generated image after replacing the comparison image. Under the condition of 25%, 50%, the average probability of face recognition is 73.06%, 38.17% respectively. This data more truly reflects the control of the model in this paper on the problem of identity preservation. The identity is well preserved under quarter masks while lost obviously under half masks. However, it is still improved compared with the original model.

Table 4: Face recognition accuracy with different iterations

	Mask1	Mask2	Mask3	Mask4	Mask5	Mask6
1	58.35%	56.95%	99.37%	96.76%	98.85%	95.99%
2	74.77%	76.64%	99.57%	98.31%	99.37%	98.47%
3	80.23%	81.29%	99.71%	98.51%	99.55%	98.78%
4	82.34%	82.34%	99.75%	98.79%	99.64%	98.92%
5	**83.15%**	**83.28%**	**99.75%**	**98.99%**	**99.77%**	**98.95%**
6	83.04%	81.53%	99.82%	98.81%	99.59%	98.85%
	Mask7	**Mask8**	**Mask9**	**Mask10**	**Mask11**	**Mask12**
1	31.51%	48.02%	99.62%	84.13%	99.57%	78.88%
2	52.65%	68.65%	99.91%	92.25%	99.82%	88.82%
3	59.39%	73.76%	99.82%	94.16%	99.84%	91.99%
4	62.09%	76.57%	99.93%	94.57%	99.86%	92.14%
5	**63.77%**	78.85%	99.94%	**94.81%**	**99.91%**	**92.22%**
6	62.72%	**76.30%**	**99.84%**	94.50%	99.89%	91.80%

Table 5: Face recognition accuracy on standard mask under different iterations (refer to identity picture)

	Mask1	Mask2	Mask3	Mask4	Mask5	Mask6
1	23.40%	22.90%	63.87%	59.83%	64.61%	59.93%
2	33.48%	34.18%	71.00%	66.66%	71.71%	65.95%
3	38.08%	36.54%	73.73%	69.07%	75.10%	68.47%
4	39.29%	38.82%	75.03%	69.16%	75.86%	69.12%
5	**41.00%**	**39.29%**	**76.99%**	**71.30%**	**76.52%**	**71.40%**
6	40.45%	38.88%	76.92%	69.93%	74.61%	70.00%
	Mask7	**Mask8**	**Mask9**	**Mask10**	**Mask11**	**Mask12**
1	15.36%	19.78%	61.71%	53.65%	66.57%	54.56%
2	22.02%	29.02%	65.32%	62.90%	74.99%	63.04%
3	25.57%	30.96%	71.69%	64.77%	78.06%	67.04%
4	27.07%	32.71%	72.57%	65.01%	78.36%	67.60%
5	**27.58%**	**34.81%**	**74.33%**	**65.63%**	79.94%	**68.28%**
6	25.94%	33.64%	73.80%	64.80%	**80.07%**	67.81%

Visualization System

The user interface of our system is shown in the Figure 7. User is asked to open a image file and choose a desired mask. Drawing a free mask is also supported. After the mask is confirmed, click the generate button to view the result.The face distance between the result and the original image is also displayed. Generally, when the distance is less than 0.5, it can be considered as the same person. In addition, the compare button can be clicked to compare the result with the original image.

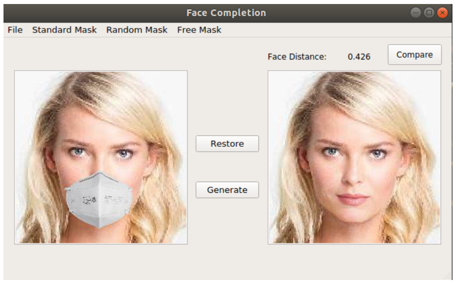

Figure 7: The user interface of our system

Failure Case

Although our system works well in most cases, it fails in some cases.First,the inpainting of the background is not good enough because our models focus more on the face. Second,our system don't perform well at extreme pose or angle, which is shown in the first two rows of Figure 8. Plus, the completion result depend highly on the predicted landmark. Third, we ignore the symmetry feature of face. As shown in the last two rows of the Figure 8 , neither the man's beard nor the woman's glasses are well completed.

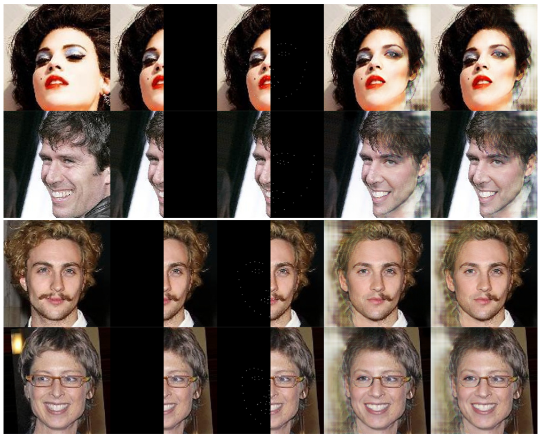

Figure 8: The failure case of our system

Conclusion

Face completion technique is to recover a occulsion or partial face image. A lot of excellent work has emerged, but most of them only pay attention to the authenticity of the generated image. The identity information is ignored under most circumstances. Besides, most of GAN-based inpainting relevant work is not suitable for the face completion task, which is more sensitive to image details. Taken the aforementioned problem into account, we present a novel framework which preserves the identity information as well as the face details simultaneously. Our method adds an identity loss to preserve the image identity, and introduce the landmark as a prior knowledge to constrain the face topology. Extensive experiments show that our method obtains state-of-the-art results on the CelebA dataset.

References

Barnes, C.; Shechtman, E.; Finkelstein, A.; and Goldman, D. B. 2009. PatchMatch: A randomized correspondence algorithm for structural image editing. In *ACM Transactions on Graphics (ToG)*, volume 28, 24. ACM.

Bertalmio, M.; Sapiro, G.; Caselles, V.; and Ballester, C. 2000. Image inpainting. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, 417–424.

Criminisi, A.; Pérez, P.; and Toyama, K. 2004. Region filling and object removal by exemplar-based image inpainting. *IEEE Transactions on image processing* 13(9): 1200–1212.

Darabi, S.; Shechtman, E.; Barnes, C.; Goldman, D. B.; and Sen, P. 2012. Image melding: combining inconsistent images using patch-based synthesis. *ACM Transactions on Graphics (ToG)* 31(4): 1–10.

Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; and Fei-Fei, L. 2009. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, 248–255. Ieee.

Esedoglu, S.; and Shen, J. 2002. Digital inpainting based on the Mumford–Shah–Euler image model. *European Journal of Applied Mathematics* 13(4): 353–370.

Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; and Bengio, Y. 2014. Generative adversarial nets. In *Advances in neural information processing systems*, 2672–2680.

Iizuka, S.; Simo-Serra, E.; and Ishikawa, H. 2017. Globally and locally consistent image completion. *ACM Transactions on Graphics (ToG)* 36(4): 1–14.

Isola, P.; Zhu, J.; Zhou, T.; and Efros, A. A. 2017. Image-to-Image Translation with Conditional Adversarial Networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, 5967–5976. IEEE Computer Society. doi:10.1109/CVPR.2017.632. URL https://doi.org/10.1109/CVPR.2017.632.

Johnson, J.; Alahi, A.; and Fei-Fei, L. 2016. Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision*, 694–711. Springer.

Karras, T.; Aila, T.; Laine, S.; and Lehtinen, J. 2017. Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196* .

Li, X.; Liu, M.; Zhu, J.; Zuo, W.; Wang, M.; Hu, G.; and Zhang, L. 2018. Learning Symmetry Consistent Deep CNNs for Face Completion. *arXiv preprint arXiv:1812.07741* .

Li, Y.; Liu, S.; Yang, J.; and Yang, M.-H. 2017. Generative face completion. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3911–3919.

Liu, G.; Reda, F. A.; Shih, K. J.; Wang, T.-C.; Tao, A.; and Catanzaro, B. 2018. Image inpainting for irregular holes using partial convolutions. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 85–100.

Liu, Z.; Luo, P.; Wang, X.; and Tang, X. 2015. Deep Learning Face Attributes in the Wild. In *Proceedings of International Conference on Computer Vision (ICCV)*.

Nazeri, K.; Ng, E.; Joseph, T.; Qureshi, F. Z.; and Ebrahimi, M. 2019. Edgeconnect: Generative image inpainting with adversarial edge learning. *arXiv preprint arXiv:1901.00212* .

Pathak, D.; Krahenbuhl, P.; Donahue, J.; Darrell, T.; and Efros, A. A. 2016. Context encoders: Feature learning by inpainting. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2536–2544.

Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; and Chen, L.-C. 2018. Inverted residuals and linear bottlenecks: Mobile networks for classification, detection and segmentation. *arXiv preprint arXiv:1801.04381* .

Schroff, F.; Kalenichenko, D.; and Philbin, J. 2015. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 815–823.

Simakov, D.; Caspi, Y.; Shechtman, E.; and Irani, M. 2008. Summarizing visual data using bidirectional similarity. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, 1–8. IEEE.

Simonyan, K.; and Zisserman, A. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* .

Song, L.; Cao, J.; Song, L.; Hu, Y.; and He, R. 2019. Geometry-aware face completion and editing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 2506–2513.

Xie, J.; Xu, L.; and Chen, E. 2012. Image denoising and inpainting with deep neural networks. In *Advances in neural information processing systems*, 341–349.

Yang, Y.; Guo, X.; Ma, J.; Ma, L.; and Ling, H. 2019. LaFIn: Generative Landmark Guided Face Inpainting. *arXiv preprint arXiv:1911.11394* .

Yu, J.; Lin, Z.; Yang, J.; Shen, X.; Lu, X.; and Huang, T. S. 2018. Generative image inpainting with contextual attention. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 5505–5514.

Zhang, Z.; Zhou, X.; Zhao, S.; and Zhang, X. 2019. Semantic Prior Guided Face Inpainting. In *Proceedings of the ACM Multimedia Asia on ZZZ*, 1–6.