electronics

MDPI

Article

Face Image Completion Based on GAN Prior

Xiaofeng Shao [1], Zhenping Qiang [1,*], Fei Dai [1], Libo He [2] and Hong Lin [1]

[1] College of Big Data and Intelligent Engineering, Southwest Forestry University, Kunming 650224, China; shaoxf1996@gmail.com (X.S.); daifei@swfu.edu.cn (F.D.); linh1226@swfu.edu.cn (H.L.)
[2] Information Security College, Yunnan Police College, Kunming 650221, China; LiboHe2022@gmail.com
* Correspondence: qzp@swfu.edu.cn

Abstract: Face images are often used in social and entertainment activities to interact with information. However, during the transmission of digital images, there are factors that may destroy or obscure the key elements of the image, which may hinder the understanding of the image's content. Therefore, the study of image completion of human faces has become an important research branch in the field of computer image processing. Compared with traditional image inpainting methods, deep-learning-based inpainting methods have significantly improved the results on face images, but in the case of complex semantic information and large missing areas, the completion results are still blurred, and the color of the boundary is inconsistent and does not match human visual perception. To solve this problem, this paper proposes a face completion method based on GAN priori to guide the network to complete face images by directly using the rich and diverse a priori information in the pre-trained GAN. The network model is a coarse-to-fine structure, where the damaged face images and the corresponding masks are first input to the coarse network to obtain the coarse results, and then the coarse results are input to the fine network with multi-resolution skip connections. The fine network uses the a priori information from the pre-trained GAN to guide the network to generate the face images, and finally uses the SN-PatchGAN discriminator to evaluate the completion results. The experiment is performed on the CelebA-HQ dataset. Compared with the latest three completion methods, the qualitative and quantitative experimental analysis shows that our method has obvious improvement in texture and fidelity.

Keywords: face image completion; generative adversarial network; SN-PatchGAN

check for updates

Citation: Shao, X.; Qiang, Z.; Dai, F.; He, L.; Lin, H. Face Image Completion Based on GAN Prior. *Electronics* **2022**, *11*, 1997. https://doi.org/10.3390/electronics11131997

Academic Editor: Byung Cheol Song

Received: 21 May 2022
Accepted: 23 June 2022
Published: 26 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 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 (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Image completion, also often referred to as image inpainting, is the process of using information from undamaged areas of the original image or intact images of the same type to fill in the contents of missing areas of the image. The core problem of image completion is to synthesize semantically reasonable and realistic pixels in the region of missing information with the original image.

Image inpainting methods can be roughly divided into traditional methods and deep-learning-based methods. There are two main types of traditional image inpainting methods, including diffusion-based and patch-based methods. Although diffusion-based and patch-based traditional inpainting methods are very effective in texture completion, they are not capable of face completion with semantics. With the rapid development of deep learning, the emergence of feature-learning-based image completion methods exactly compensates the shortcomings of traditional image inpainting methods. Although the learning-based approach is capable of generating reasonable content, many existing image completion techniques generate over-smoothed and/or blurry regions, failing to reproduce fine details.

In particular, Goodfellow et al. proposed Generative Adversarial Networks (GAN) [1] in 2014, which again provided new ideas for image completion. This is because GAN models trained on large-scale natural images can obtain rich texture and shape priors. For example, the progressive growth generative adversarial networks PGGAN [2] and

StyleGAN [3] proposed by karras et al. have achieved outstanding results in the work of generating human faces, and they can generate realistic images from a random noise (latent vector). GAN has been widely used in image inpainting [4], image recognition [5], image denoising [6], super resolution [7] and other image processing fields.

Inspired by GLEAN [7], in an encoder–decoder architecture, a pre-trained GAN can provide a priori information to guide the image to recover texture and details. In this paper, we propose a face image completion model based on GAN priors. The model uses a two-stage generative network from coarse to fine, and uses gated convolution instead of vanilla convolution in the generative network to distinguish valid pixels from invalid pixels of the image and improve the quality of the completion results. The coarse network is a typical encoder–decoder structure, in which the dilated convolution is added to enhance the perceptual field of the features and output to get the coarse result, and then the coarse result is input to the fine network. The fine network uses the a priori information from the pre-trained GAN to guide the network to generate the face image, and finally, the SN-PatchGAN discriminator is used to evaluate the completion result. Figure 1 shows the completion results of our method from coarse to fine. The coarse result in Figure 1c generates reasonable content but lacks details, while the fine network fills in the details using a priori information, as shown in Figure 1d.

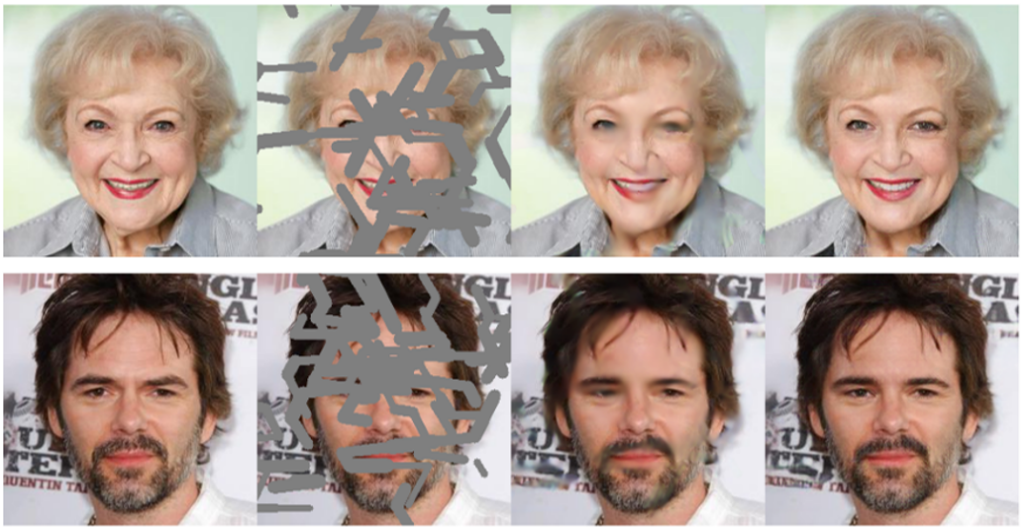


Figure 1. Completion results from coarse to fine.

The contributions of this paper are follows:

1. A coarse-to-fine face completion network based on pre-trained GAN prior to improve the texture and fidelity of the completion results.
2. Integration of a priori information from a pre-trained GAN into an encoder-decoder architecture with multi-resolution skip connections.

The rest of the paper is structured as follows: In Section 2, related works including multiple image inpainting methods are discussed. This is followed in Section 3 by details about our methods. Section 4 describes our experiments in detail. Section 5, ablation studies. Finally, conclusions and future works are given.

2. Related Work

2.1. Traditional Image Inpainting Methods

There are two main traditional methods for image compeltion, diffusion-based methods and patch-based methods. Diffusion-based [8,9] methods fill the missing region by smoothly propagating the image content from the boundary to the interior of the missing region. This type of method only fills the texture of the image, and it is difficult to capture

the semantic information and global structure of the image, so this type of method cannot inpainting face images with complex information semantics. In addition, this approach does not work in the case of large missing areas.

On the other hand, the patch-based approach [10] can fill in the missing regions in an image by searching for similar patches in the image. Patch-based methods can synthesize plausible textures, but they usually fail badly in cases such as complex scenes and human faces. Additionally, these methods do not work in complex scenes with large missing areas, especially when the texture to be filled is not present in the image. Moreover, patch-based inpainting methods are slow, computationally expensive, and not based on semantic understanding of the image.

Therefore, traditional image completion methods are not competent for face image completion.

2.2. Deep-Learning-Based Image Inpainting Methods

In recent years, image inpainting methods based on deep learning have been widely studied. Pathak et al. [4] introduced adversarial loss into the image inpainting task, and they used adversarial loss to make the generated images as realistic as possible. Based on this work, Yeh et al. [11] used DCGAN [12] for image inpainting, which was able to successfully generate missing parts of the image and fill them. Although these methods achieved satisfactory results, there were obvious stitching traces in the border areas of the completed images. He et al. [13] successfully achieved the completion of high-resolution face images using a pre-trained stylegan.

Iizuka et al. [14] added a dual discriminator (local discriminator and global discriminator) to the image completion model, and this model ensured the local consistency and global consistency of the generated images. Li et al. [15] proposed a face image completion model on this structure and further ensured that the generated content was more realistic and enhanced the consistency between old and new pixels by using a pre-trained Parsing network. Yu et al. [16] proposed a coarse-to-fine network structure based on this structure and added a contextual attention layer to copy or borrow feature information from known backgrounds to generate real pixels. This dual discriminator structure is valid for the central rectangular mask, but in practice, there may be multiple arbitrarily shaped masks at any location.

Liu et al. [17] argued that vanilla convolution indiscriminately convolves valid and masked pixels, which usually leads to situations such as color differences and blurring, and thus proposed partial convolution to improve the quality of image completion and proposed a free-form mask generation method. Yu et al. [18] proposed a gated convolution, generalized partial convolution by providing a learnable dynamic feature selection mechanism for each channel at each spatial location across all layers.

Nazeri et al. [19] proposed a two-stage adversarial image completion model that used an edge generator to predict the edge map of an image in the first stage, and then generated the image in the second stage based on structural information using the edge map. Yang et al. [20] also used a similar structure to achieve face image completion by predicting the landmarks of faces. Such methods rely heavily on the accuracy of the prediction, and the completion results are often unsatisfactory when the mask area is large.

Deep-learning-based methods have achieved remarkable success in image completion tasks. These methods use the learned data distribution to fill in the missing pixels. They are able to generate coherent structures in the missing regions, a feat that is almost impossible to achieve with traditional techniques. Although these methods are able to generate meaningful structures in the missing regions, the generated regions sometimes suffer from blurring, lack of context, and discontinuous color of the completion boundaries.

3. Our Method

To solve these problems, in this paper we design a coarse-to-fine architecture that allows the user to output completion results using a priori information in just one pass

forward. The coarse network is a typical encoder–decoder structure. Adding a gated dilation convolution operation to this coarse network increases the range of the feature perception field, which allows the model to see a larger input image area and more global information, and outputs a coarse completion result. The coarse results are then fed into the fine network, and the prior information in the pre-trained GAN is used to guide the network to generate face images, and finally the completion results are evaluated using the SN-PatchGAN discriminator. An overview of the architecture is depicted in Figure 2.

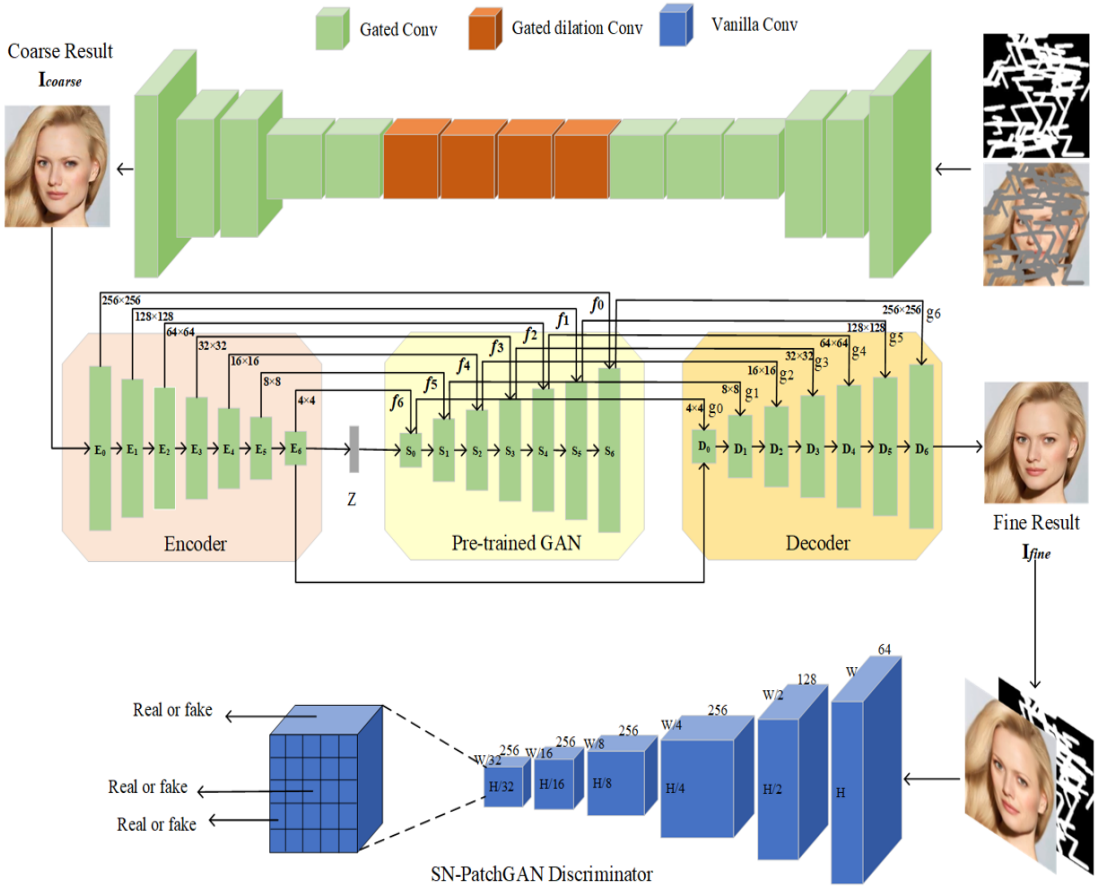


Figure 2. Face image completion model based on GAN prior.

3.1. Coarse Network

The coarse network is a encoder–decoder structure with a gated dilation convolution block, because the global information is crucial for image completion, and through the dilation convolution, the model can see a larger area of the input image than when using the standard convolution layer, so the coarse network can effectively learn better feature representation. With the input damaged image and corresponding mask, a coarse result can be obtained by the coarse network.

3.2. Fine Network

The fine network consists of three components, including a multi-resolution encoder, a pre-trained GAN, and a decoder.

3.2.1. Multi-Resolution Encoder

In order to extract features and latent vectors of the images and provide a priori information for the pre-trained GAN to generate features, we use gated convolution blocks to gradually reduce the resolution of the features.

$$f_i = \begin{cases} E_0(I_{coarse}), & if \ \ i = 0 \\ E_i(f_{i-1}), & else \ \ i \in 1, \ldots N \end{cases} \tag{1}$$

where E_i denotes a stack of a stride-2 gated convolution and a stride-1 gated convolution. N is the number of downsampling in the encoder stage of the fine network, and in our work, N is 6. To generate the latent vector, we first downsample the input image using multiple and multi-resolution features, and finally obtain the latent vector through the fully connected layer.

3.2.2. Pre-Trained Generator

Given convolutional features f_i and latent vectors Z, we utilize a pre-trained PG-GAN as a generator to provide a priori information for texture and detail generation. Since PGGAN uses a progressive generation approach and was originally designed for image generation tasks, it cannot be directly integrated into the proposed fine-grained network framework. In this work, we adapt PGGAN to the image completion network of our work with some modifications:

1. The dimensionality of the latent vector output from the fully connected layer is the same as the input dimensionality of the pre-trained GAN
2. In order to adjust the feature S_i of the encoder, we use an additional gated convolution in each module for feature fusion:

$$g_i = \begin{cases} S_0(Z, f_N) \\ S_i(g_{i-1}, f_{N-i}) \end{cases} \tag{2}$$

   S_0 initializes Z into a 4×4 feature map using a pre-trained generator, and then fuses it with the features F_N to g_i . The S_i module first upsamples the S_{i-1} features, then convolves the features with a priori information using the pre-trained generator, and finally fuses the features F_{N-1} from the encoder to generate g_i .
3. Instead of generating the output directly from the generator, we output the features g_i and pass them to the decoder to better fuse the features from the generator and the encoder.

3.2.3. Decoder

We use an additional decoder with progressive fusion to integrate the features of the encoder and the generator to generate the output image. It takes the encoder's features E_N input and progressively fuses the multi-resolution features g_i generated by the pre-trained generator.

$$d_i = \begin{cases} D_0(f_N), & if \ \ i = 0 \\ D_i(d_{i-1}, g_i), & otherwise \end{cases} \tag{3}$$

where D_i and d_i denote the 3×3 convolution and its output, respectively. In addition to the final output layer, each gated convolution is followed by a PixelShuffle [21] layer. With a skip connection between the encoder and decoder, the information captured by the encoder can be enhanced so that the generator can focus more on the generation of textures and details.

3.3. SN-PatchGAN Discriminator

While a normal discriminator maps the input to a scalar, where the magnitude of the value indicates the probability that the input sample is real or fake, a Patch-GAN [22] discriminator maps the input to a feature map, where the real or fake of each pixel block on the feature map is evaluated. The SN-PatchGAN [18] discriminator uses spectral normalization (SN) [23], where the weight matrix is divided by the corresponding Lipschitz constant. Theoretical analysis shows that spectral normalization can control the generalization error and stabilize the training.

3.4. Loss Function

In this paper, we use L1 loss, perceptual loss [24], style loss [25] and adversarial loss to optimize the network. For a given damaged image and the corresponding binary mask map M (value 0 for the missing region and value 1 for the valid pixel region), define the output of the first stage of the network as I_{coarse}, the output of the second stage as I_{fine} and the real image as I_{gt}. Define L1 loss:

$$L_1 = \|(1-M)(I_{coarse} - I_{gt})\|_1 + \|(1-M)(I_{fine} - I_{gt})\|_1 \tag{4}$$

The perceptual loss and style loss compare the differences in the deep feature maps between the generated and real images. Such a loss function can effectively transfer the structural and textural information of the image to the model. The perceptual loss function is defined as follows:

$$L_{perc} = \mathbb{E}\left[\sum_i \frac{1}{N_i}\|\phi_i(I_{fine}) - \phi_i(I_{gt})\|_1\right] \tag{5}$$

where ϕ_i is the activation map of the ith layer of a pre-trained network. In this work, ϕ_i corresponds to the activation mappings of the $Relu1_1$, $Relu2_1$, $Relu3_1$, $Relu4_1$ and $Relu5_1$ layers of the VGG-19 [26] network pre-trained on the ImageNet dataset [27]. These activation feature maps are also used to calculate the style loss. Given the feature map of $C_j \times H_j \times W_j$, define the style loss as:

$$L_{style} = \mathbb{E}_j\left[\|G_j^{\phi}(I_{fine}) - G_j^{\phi}(I_{gt})\|_1\right] \tag{6}$$

where G_j^{ϕ} is a $C_j * C_j$ Gram matrix constructed from activation maps ϕ_j. Moreover, we use LSGAN [28] to train the face image completion network. The adversarial loss is defined as:

$$L_{adv} = \frac{1}{2}\mathbb{E}_{x\sim p_r}[D(x)-a]^2 + \frac{1}{2}\mathbb{E}_{z\sim p_z}[D(G(z)-b]^2 + \frac{1}{2}\mathbb{E}_{z\sim p_z}[D(G(z)-c]^2 \tag{7}$$

where D is the discriminator and G is the generator. The constants a, b denote the labels of the real image and the generated image, respectively. c is the value of the generator to make the discriminator think that the generated image is the real data. In our experiment, $a = c = 1$, $b = 0$. So, our overall loss function is:

$$L_{total} = \lambda_1 L_1 + \lambda_2 L_{perc} + \lambda_3 L_{style} + \lambda_4 L_{adv} \tag{8}$$

4. Experimental Results

In this paper, we use pre-trained PGGAN as generator and use publicly available codes of existing methods for comparison. The mask is generated by the method of Yu et al. [18], and a small number of small area rectangles are randomly added to increase the diversity of the mask. Classify masks in 10% steps according to their size relative to the whole image (e.g., 0–10%, 10–20%). To maintain fairness, both the method in this paper and the compared method are trained on the CelebA-HQ dataset to ensure that the difference in completion quality is mainly caused by the algorithm rather than the training distribution, while the test set is strictly excluded from the training set. The CelebA-HQ dataset contains 30k face images with a resolution of 256 × 256. We use 29k for training and the remaining 1k for testing. We use Adam [29] in our training to optimize the network and $\beta_1 = 0.5$, $\beta_2 = 0.99$. The learning rates of the generator and discriminator are $lr_g = 1 * e^{-4}$ and $lr_d = 4 * e^{-4}$, respectively. The weight values of the losses are $\lambda_1 = 1$, $\lambda_2 = 0.1$, $\lambda_3 = 250$ and $\lambda_4 = 1$, respectively. We compare the following completion methods: Gated Convolution(GC) [18], EdgeConnect (EC) [19], and Recurrent Feature Reasoning(RFR) [30].

4.1. Qualitative Analysis

The qualitative comparison results under free form and rectangular center masks are shown in Figures 3 and 4, respectively. The completion results of GC suffer from blurring and lack of details, as in Figure 3b. The completion result of EC depends heavily on the accuracy of the predicted edge map, and EC does not work well when the key information is obscured in a large area, and even two pairs of eyes appear in the completion result in row 5 of Figure 4c. RFR uses recurrent feature reasoning, which recursively infers the hole boundaries of the convolutional feature mapping, and then uses them as cues for further reasoning. This recurrent feature inference does not work well on human faces, and its completion results are often blurred and do not understand the semantic information of the input, as in Figure 3d. The experimental results under free-form and rectangular center masks show that the completion results of our method have better texture and detail information than those of other methods.

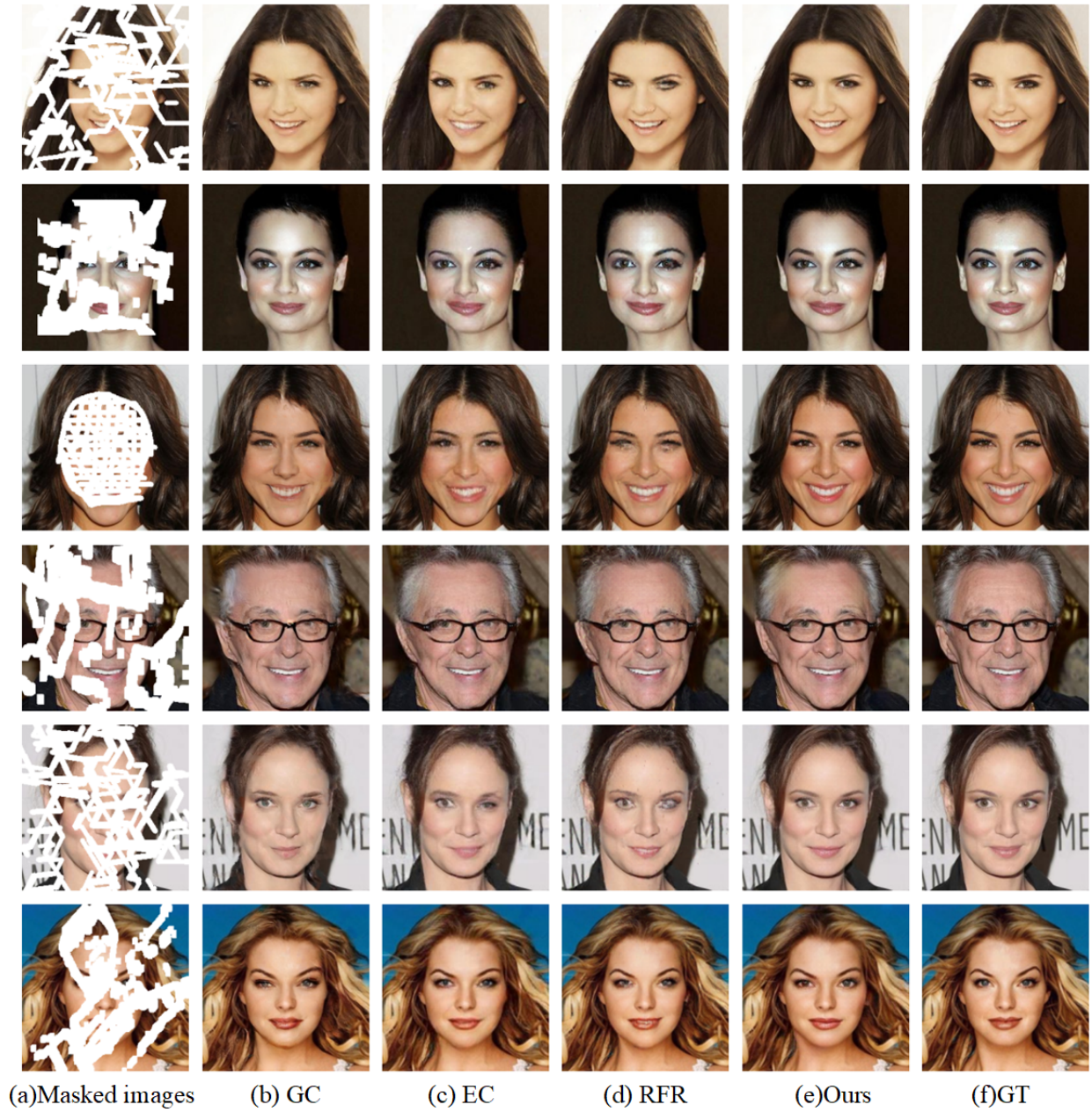


Figure 3. Qualitative comparison results under free-form masks.

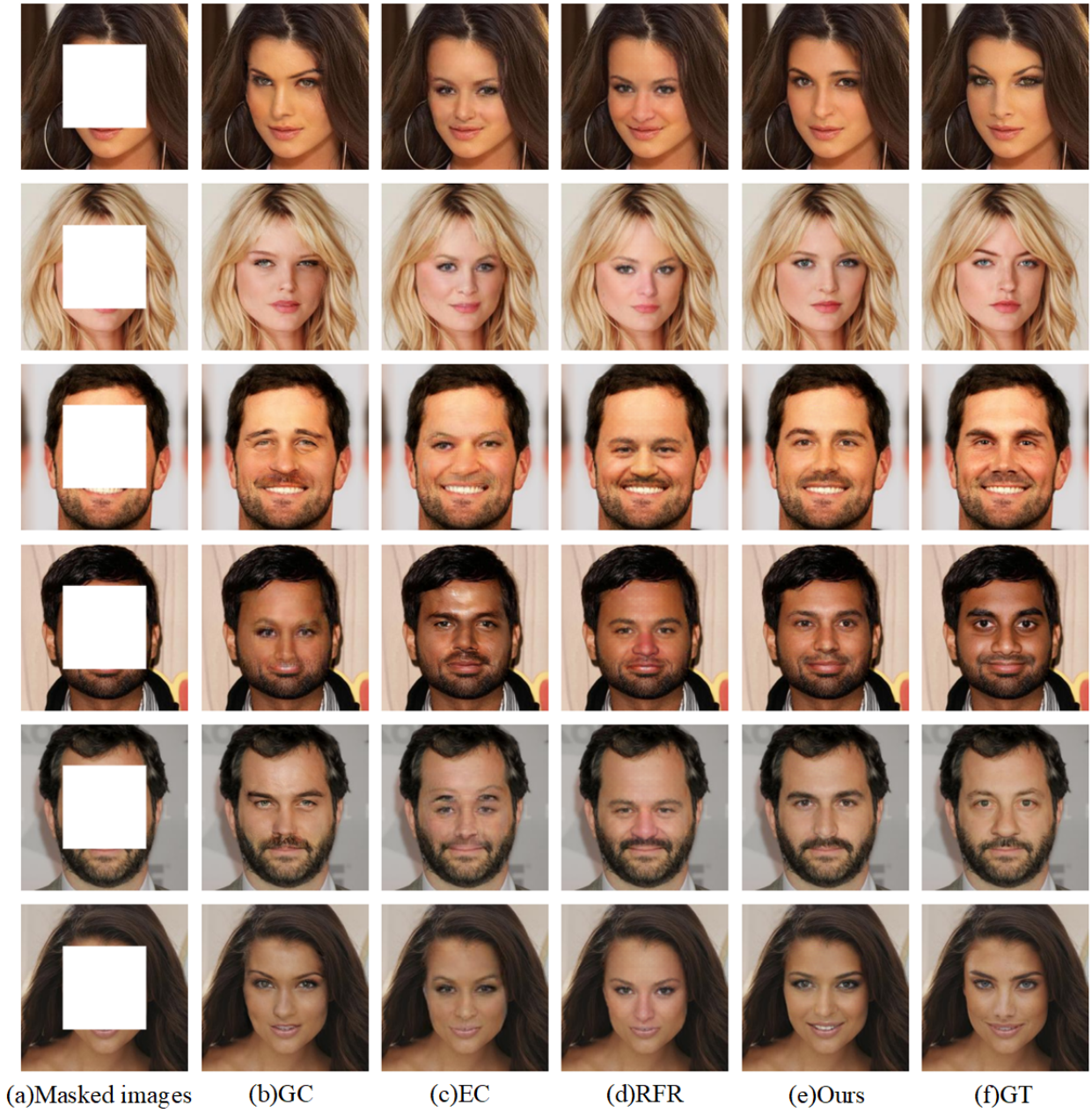


Figure 4. Qualitative comparison results under the center rectangle mask.

4.2. Quantitative Analysis

We provide a quantitative comparison of the different methods in Table 1. For each category, we selected 10k images and calculated their average Structural Similarity(SSIM) and Peak Signal to Noise Ratio (PSNR). The experiments show that the PSNR of this method is better than those of the compared methods in the case of 10–50% mask, and the average improvement over the RFR model is 8.6%. For a mask size of 10–40%, the structural similarity between our method and the other methods compared is not significant; for a mask size of 40–50% of the image, our method achieves the highest structural similarity. This demonstrates the effectiveness of our approach.

Table 1. Quantitative comparison results when the proportion of the mask to the image is 10–50%.

	Mask	GC	EC	RFR	Ours
SSIM	10–20%	0.973	0.975	**0.981**	0.966
	20–30%	0.938	0.932	**0.952**	0.943
	30–40%	0.914	0.915	**0.934**	0.923
	40–50%	0.859	0.731	0.886	**0.900**
PSNR	10-20%	32.56	32.48	33.56	**35.14**
	20–30%	29.25	28.92	30.89	**32.38**
	30–40%	26.72	26.62	27.76	**30.86**
	40–50%	24.31	24.18	25.46	**29.42**

5. Ablation Studies

In this section, we performed an ablation study in the CelebA-HQ dataset to show the effect of different sections on the complementation results.

5.1. Dilation Convolution

The global information of an image is crucial for image completion work. Dilated convolution can increase the perceptual field of the feature map. Adding dilated convolution to the network enables the network model to obtain a larger range of information. Figure 5 illustrates the effect of dilation convolution on the completion results. In the experiments, the four gated dilated convolutions in Figure 2 are replaced with gated convolutions. From Figure 5, it can be seen that the completion results of the model without the dilated convolutional products look reasonable but the quality of details is poor, while the model with the dilated convolutional products has a better quality because it obtains more global information.

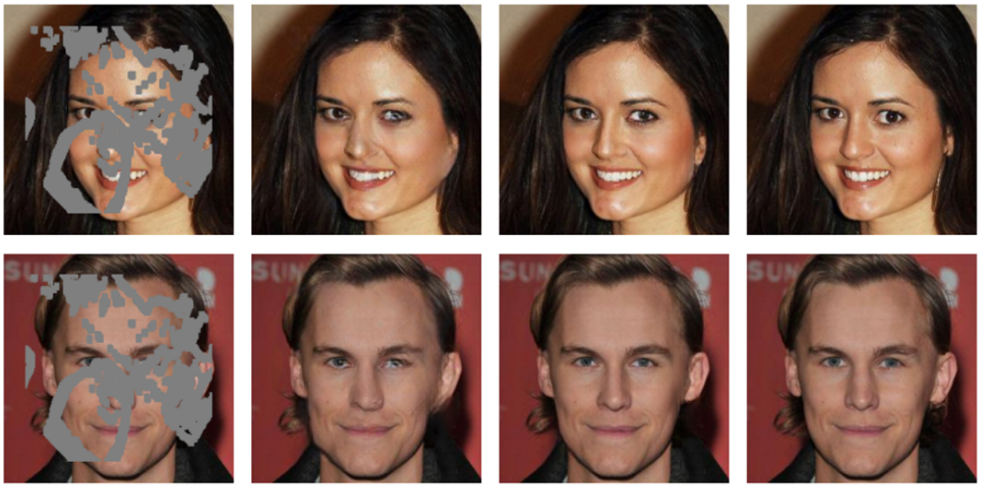


Figure 5. The effect of dilation convolution on the completion result.

5.2. Multi-Resolution Encoder

Figure 6 shows how the encoder features can be used to generate features with a priori information to help recover details and local structure. In the experiments, only the latent vectors are started and the effect of the features on the generated results is observed when they are gradually introduced into the generator. In order to avoid the interference of the decoder on the completion results, the experiments are tested using a variant of the decoder, and the output image is generated directly by the generator. When all convolutional features are dropped, the information of the image is poorly preserved and finer details are not retained. When coarse convolutional features (4×4 to 16×16 features) are passed to the pre-trained generator, more details are recovered and the output better approximates the original image. Further improvements in quality

and fidelity are observed when finer features are passed to the generator. The above observations confirm that multi-resolution convolutional features are essential in guiding the recovery of fine details and local structures that cannot be reconstructed using only latent vectors.



Figure 6. The effect of multi-resolution encoder on completion results.

5.3. Pre-Trained GAN

In order to analyze the contribution of pre-trained generators to the completion network, we investigated the impact of generating features from pre-trained generators. In the experiments, all pre-trained models are discarded at first, and then these features are gradually passed to the decoder. As Figure 7 gives two examples of the experimental results, the network is responsible for both generating realistic details and textures and maintaining the fidelity of the output results due to the lack of proper prior information. Such a demanding goal eventually leads to completion results that are deficient in both structure recovery and texture generation (as shown in Figure 7b). Since the pretrained generator already captures rich image priors, it can reduce the burden of generating textures and details. Therefore, when passing finer features (4×4, 8×8, $\ldots$, 256×256 features) to the pre-trained generator, improvements in structure and texture can be observed (e.g., Figure 7c–e by gradually adding finer features, the completion results have more reasonable structure and finer texture).

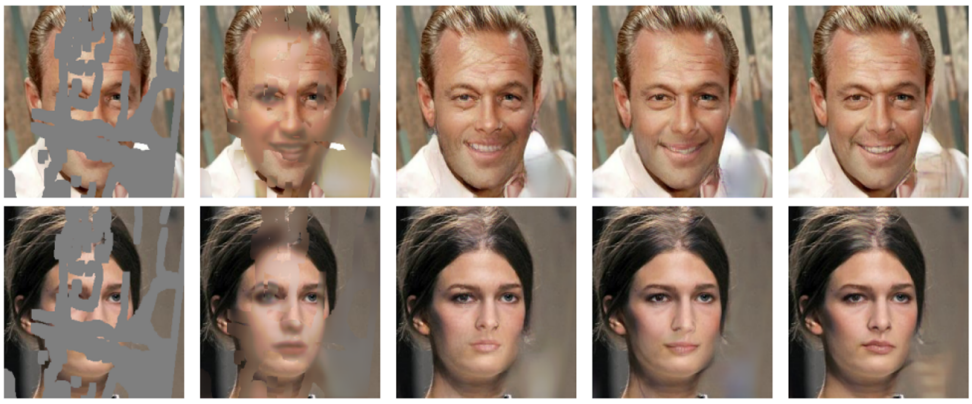


Figure 7. The effect of pre-trained GAN on completion results.

5.4. Decoder

In the case of no decoder, although the overall perception is convincing, the enlarged output image clearly loses a lot of details. As shown in Figure 8, the completion result in row 1 (with decoder) has much more detail in the eyes and teeth than in row 2 (without decoder). This is because the decoder fused more feature information to produce more natural details. In addition, the multi-resolution skip connection between the encoder and decoder enhances the spatial information captured in the encoder features and makes the generator pay more attention to detail generation, which further improves the output quality.

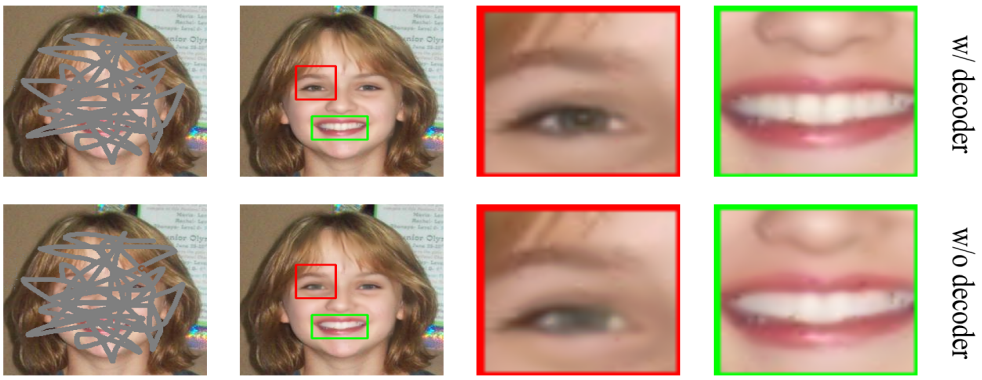


Figure 8. The effect of the decoder on the completion result.

5.5. SN-PatchGAN

In this paper, we consider the problem of free-form masks. Such masks can appear anywhere in the image, whereas previous work, for a single mask, often used global and local discriminators to discriminate the generated region and global information, respectively, which is not applicable to this work. In Figure 9, we show the difference between SN-PatchGAN discriminators and global discriminators. Although the global discriminator can also produce reasonable results, there is still a significant difference in details between the SN-PatchGAN. For example, the eyes generated in Figure 9b lose details compared to the completion results in Figure 9c, which illustrates the effectiveness of SN-PatchGAN in this work.



Figure 9. Comparison of SN-PatchGAN discriminator and global discriminator completion results.

6. Conclusions

We propose a method to perform face image completion tasks by adding a pre-trained GAN to the encoder–decoder network structure. It is experimentally demonstrated that a pre-trained GAN can improve the quality of completion results by using its rich prior information. Compared with existing methods, the completion results in this paper are significantly improved in terms of texture and plausibility.

However, a priori-based methods have a prerequisite that a GAN with superior performance needs to be trained in advance. Such methods will fail when the experimental data cannot be trained to generate a GAN with excellent performance. In addition, recent research has shown that multiple reasonable results can be achieved [31,32], and we will study multiple-result tasks in the follow-up work.

Author Contributions: Conceptualization, Z.Q. and X.S.; methodology, X.S.; software, X.S.; validation, L.H.; formal analysis, Z.Q. and F.D.; investigation, L.H.; resources, Z.Q. and X.S.; data curation, X.S.; writing—original draft preparation, X.S.; writing—review and editing, H.L., Z.Q. and X.S.; visualization, X.S.; supervision, Z.Q.; project administration, Z.Q.; funding acquisition, Z.Q. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the projects of Natural Science Foundation of China (Grant No. 12163004), the Yunnan Fundamental Research Projects (Grant No. 202001AT070135, No. 202002AD080002), the key scientific research project of Yunnan Police College (19A010).

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

GT	Ground True.
GAN	Generative Adversarial Network.
DCGAN	Deep Convolutional GAN.
PGGAN	Progressive Growing GAN.
VGG	Visual Geometry Group.
SN	Spectral normalization.
GC	Gated Convolution.
EC	Edge Connect.
RFR	Recurrent Feature Reasoning.
SSIM	Structural SIMilarity.
PSNR	Peak Signal-to-noise Ratio.
W/	With.
W/O	Without.

References

1. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **2014**, *3*, 2672–2680.
2. Karras, T.; Aila, T.; Laine, S.; Lehtinen, J. Progressive growing of gans for improved quality, stability, and variation. *arXiv* **2017**, arXiv:1710.10196.
3. Karras, T.; Laine, S.; Aila, T. A style-based generator architecture for generative adversarial networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–17 June 2019.
4. 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, 27–30 June 2016.
5. Yacine, K.; Benzaoui, A. A new framework for grayscale ear images recognition using generative adversarial networks under unconstrained conditions. *Evol. Syst.* **2021**, *12*, 923-934.
6. Khan, A.; Jin, W.; Haider, A.; Rahman, M.; Wang, D. Adversarial gaussian denoiser for multiple-level image denoising. *Sensors* **2021**, *21*, 2998. [CrossRef] [PubMed]
7. Chan, K.C.; Wang, X.; Xu, X.; Gu, J.; Loy, C.C. Glean: Generative latent bank for large-factor image super-resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Venice, Italy, 22–29 October 2021.
8. Li, H.; Luo, W.; Huang, J. Localization of diffusion-based inpainting in digital images. *IEEE Trans. Inf. Forensics Secur.* **2017**, *12*, 3050–3064. [CrossRef]

9. Chan, T.F.; Shen, J. Nontexture inpainting by curvature-driven diffusions. *J. Vis. Commun. Image Represent.* **2001**, *12*, 436-449. [CrossRef]
10. Barnes, C.; Shechtman, E.; Finkelstein, A.; Goldman, D.B. PatchMatch: A randomized correspondence algorithm for structural image editing. *ACM Trans. Graph.* **2009**, *28*, 24. [CrossRef]
11. Yeh, R.A.; Chen, C.; Yian Lim, T.; Schwing, A.G.; Hasegawa-Johnson, M.; Do, M.N. Semantic image inpainting with deep generative models. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.
12. Radford, A.; Metz, L.; Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv* **2015**, arXiv:1511.06434.
13. He, L.; Qiang, Z.; Shao, X.; Lin, H.; Wang, M.; Dai, F. Research on High-Resolution Face Image Inpainting Method Based on StyleGAN. *Electronics* **2022**, *11*, 1620. [CrossRef]
14. Iizuka, S.; Simo-Serra, E.; Ishikawa, H. Globally and locally consistent image completion. *ACM Trans. Graph. (ToG)* **2017**, *36*, 1–14. [CrossRef]
15. Li, Y.; Liu, S.; Yang, J.; Yang, M.H. Generative face completion. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.
16. Yu, J.; Lin, Z.; Yang, J.; Shen, X.; Lu, X.; Huang, T.S. Generative image inpainting with contextual attention. In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018.
17. Liu, G.; Reda, F.A.; Shih, K.J.; Wang, T.C.; Tao, A.; Catanzaro, B. Image inpainting for irregular holes using partial convolutions. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.
18. Yu, J.; Lin, Z.; Yang, J.; Shen, X.; Lu, X.; Huang, T.S. Free-form image inpainting with gated convolution. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019.
19. Nazeri, K.; Ng, E.; Joseph, T.; Qureshi, F.Z.; Ebrahimi, M. Edgeconnect: Generative image inpainting with adversarial edge learning. *arXiv* **2019**, arXiv:1901.00212.
20. Yang, Y.; Guo, X.; Ma, J.; Ma, L.; Ling, H. Lafin: Generative landmark guided face inpainting. *arXiv* **2019**, arXiv:1911.11394.
21. Shi, W.; Caballero, J.; Huszár, F.; Totz, J.; Aitken, A.P.; Bishop, R.; Rueckert, D.; Wang, Z. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016.
22. Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.
23. Miyato, T.; Kataoka, T.; Koyama, M.; Yoshida, Y. Spectral normalization for generative adversarial networks. *arXiv* **2018**, arXiv:1802.05957.
24. Johnson, J.; Alahi, A.; Li, F.F. Perceptual Losses for Real-Time Style Transfer and Super-Resolution. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016.
25. Gatys, L.A.; Ecker, S.A.; Bethge, M. Image style transfer using convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016.
26. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
27. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. Imagenet large scale visual recognition challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [CrossRef]
28. Mao, X.; Li, Q.; Xie, H.; Lau, R.Y.; Wang, Z.; Paul Smolley, S. Least squares generative adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Nashville, TN, USA, 20–25 June 2017.
29. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
30. Li, J.; Wang, N.; Zhang, L.; Du, B.; Tao, D. Recurrent feature reasoning for image inpainting. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020.
31. Liu, H.; Wan, Z.; Huang, W.; Song, Y.; Han, X.; Liao, J. Pd-gan: Probabilistic diverse gan for image inpainting. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Venice, Italy, 22–29 October 2021.
32. Wang, J.; Deng, W.; Hu, J.; Tao, X.; Huang, Y. High-fidelity pluralistic image completion with transformers. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021.