Article

Efficient Face Region Occlusion Repair Based on T-GANs

Qiaoyue Man and Young-Im Cho * [ID]

Department of Computer Engineering, Gachon University, Seongnam-si 13120, Republic of Korea; manqiaoyue@gachon.ac.kr
* Correspondence: yicho@gachon.ac.kr; Tel.: +82-31-750-5800

Abstract: In the image restoration task, the generative adversarial network (GAN) demonstrates excellent performance. However, there remain significant challenges concerning the task of generative face region inpainting. Traditional model approaches are ineffective in maintaining global consistency among facial components and recovering fine facial details. To address this challenge, this study proposes a facial restoration generation network combined a transformer module and GAN to accurately detect the missing feature parts of the face and perform effective and fine-grained restoration generation. We validate the proposed model using different image quality evaluation methods and several open-source face datasets and experimentally demonstrate that our model outperforms other current state-of-the-art network models in terms of generated image quality and the coherent naturalness of facial features in face image restoration generation tasks.

Keywords: face detection; convolutional neural network; generative adversarial networks; image fusion

1. Introduction

Image restoration is a challenging long-term research task in the field of computer vision, which aims to fill in the missing parts of an image by using intact regions and semantic content. High-quality image restoration can be used in many fields, including old photo restoration, image reconstruction, and image denoising, etc., where the face image restoration generation task is more challenging. The facial restoration generation task aims to fill in the missing parts of face features with visually plausible hypotheses. The task focuses on how to recover the semantic face structure and fine details. Face images are feature-rich and have fewer feature repetitions, making the restoration task more difficult and challenging compared to other landscape images that have feature repetitions, for example. In addition, face complement generation tasks have a wide range of scenarios, such as public security [1], face editing, masking faces [2], and removing unwanted contents (e.g., glasses, masks, and scarves), among others. In recent years, with the development of deep learning, especially in the excellent performance of convolutional neural networks in computer vision tasks [3,4], significant progress has been made in the research of image restoration and face complementation.

Image restoration requires the algorithm to complete the missing areas of the image to be repaired according to the image itself and to maintain the continuity of the image features, making the repaired image look very natural and hard to distinguish from the undamaged image. According to Valley Theory, it is very obvious whenever there is an image feature dissonance between the generated content and the undamaged region. Therefore, high-quality image restoration requires not only that the generated content is semantically sound but also that the texture of the generated image features is clear and realistic. Early image restoration studies tried to solve the problem using ideas similar to texture synthesis by matching and replicating background patches with holes propagated from low to high, and these methods are particularly suitable for background restoration tasks and are widely used in practical applications. However, this approach also has significant shortcomings, such as cases where the missing patch cannot be found somewhere in the background

region, where the restored region involves complex non-repetitive structures (e.g., facial objects), etc.; the method also fails to produce new image content.

The mainstream image restoration methods currently fall into two main categories: The first is the traditional texture synthesis method [5,6], which samples similar blocks of pixels from the undamaged regions of an image to fill the area to be completed and is suitable for relatively simple texture image restoration [7,8]. The second is a neural network-based generative model that encodes the image into features in a high-dimensional latent space and then decodes the features into the restored full image [9,10]. However, it is difficult to utilize the intact region information in the convolution operation of missing regions, which often causes the restored image to be visually blurred. Both methods have their limitations in ensuring reasonable semantics and clear texture requirements to generate a globally uniform image. To solve these thorny problems in image remediation and restoration tasks, more learning-based approaches have been proposed, also thanks to the continuous development of neural networks and deep learning. It uses codecs as the main network for image restoration [11,12], encodes the corrupted image, and then generates the missing regions through decoders and uses adversarial loss to make the generated image as realistic as possible to better restore the missing regions. However, this method suffers from the problem of discontinuity between missing region features and non-missing region features, and the final generated image has obvious repair traces. The generative adversarial network (GAN) [13] is a new approach to image restoration, and the contextual conditional adversarial network (CC-GAN) [14] is one of the methods. This network uses a codec as a generator and a VGG [15] network as a discriminator to achieve image restoration. The context loss term is also added to make the restoration marks of the restored images less obvious. The deep generative model proposed by Yeh [16] also utilizes adversarial networks for image restoration. Unlike the previously mentioned models that use adversarial networks for restoration, this network uses undamaged images as a dataset to train the depth volumes. After training, the trained model is used to find the closest encoding to the missing image for image restoration.

In this study, we solve the problem that the traditional generative confrontation network cannot preserve the consistency and naturalness of the global feature of the image when the local image is generated. For the task of face restoration generation, we have designed the novel T-GANs (Transformer-Generative Adversarial Networks) image generation network model. In the network, a transformer module with an attention mechanism [17] is employed to detect the missing features of the face image and perform feature generation verification. In the generation network, a double discriminator is used to improve the ability to sense local and global features of the image. The following is a brief summary of the contributions of this work:

1. We studied and proposed T-GANs for face image restoration generation tasks, which solved the inability of most GANs-based image generation networks in image local feature generation to achieve uniformity and natural fidelity of the generated image as a whole.
2. In the network, add multiple transformer modules to confirm the missing features of the face before generating the local features of the face, and perform feature prediction. In the generative network, a discriminator is combined to perform local and global feature detection on the generated image.
3. In the generation network, the generator uses convolution and dilated convolution to generate facial features. In the discriminator, global and local dual discriminators are used for feature discrimination of the generated images.
4. The proposed network is validated using several open-source datasets such as VGG Face, Celeba, FFHQ, and EDFace-Celeb as well as various image quality evaluation methods such as FID, PSNR, and SSIM.

2. Related Work

Traditional image restoration methods

Traditional image restoration methods can be roughly divided into two categories: diffusion-based methods and sample patch-based methods. The diffusion-based methods [18,19] use the edge information of the missing area to try to slowly diffuse the structure and texture information of the area to fill in the internal information of the missing area. Bertalmio et al. First proposed a diffusion-based model. This model diffuses the known information of the edge of the missing area in the image to the interior of the missing area along the direction of the ISO illuminance line to achieve the goal of image repair. However, the model does not take into account the integrity of the image, resulting in poor repair effect. Shen and Chan improved the BSCB model, used the total variation (TV) model to repair images, and proposed a structural edge processing algorithm based on the variation principle, but the repair results showed the obvious fracture. Because image reconstruction is limited to locally available information, diffusion-based methods are usually unable to restore meaningful structures in the missing areas or adequately process images with large missing areas. The sample patch-based method [20,21] attempts to search for the best matching patch in the non-missing part of the image and copy it to the corresponding location to fill the missing area block by block. Drori et al. proposed an image inpainting algorithm based on samples, which uses the principle of self-similarity to get information on missing areas of images, but it runs slowly. Barnes et al. proposed a patch match algorithm to quickly find similar matches between image patches, which can repair images with large missing areas to some extent, but it requires manual intervention. Although these traditional methods can handle some simple instances, the computation of inpainting completion tasks with complex texture images appears unsatisfactory due to the lack of high-level understanding of image semantics.

Convolutional neural network-based image restoration completion

Image restoration can be regarded as a special image generation problem, which generates a part of the image (the area to be filled) and uses the texture of the known part of the image. The image inpainting completion method based on deep learning generally uses a mask to process the original image [22], confirms that the image is missing the area to be repaired, and generates a new repaired image after calculation. At present, the methods based on convolutional self-encoding networks and generative adversarial networks are widely used. Pathak et al. first proposed a deep-neural-network-based CE model for image inpainting of large missing areas and achieved impressive results. Since then, deep-learning-based image inpainting has been widely studied. Yu et al. [23] proposed a novel image restoration framework consisting of a coarse-to-fine two-stage network structure. The first-stage network uses the reconstruction loss to roughly predict the image content of the missing areas; the second-stage network refines the inpainting results of the first-stage blur through the reconstruction loss and adversarial loss, which solves the problem of images with complex textures to a certain extent. Liu et al. [24] added a coherent semantic attention layer to the two-stage model to solve the problem of discontinuous results produced in inpainting images. However, these coarse-to-fine two-stage network models often require a lot of computing resources during training, and the repair effect of the model is highly dependent on the output results of the first stage. Liu et al. [25] achieve high-quality image inpainting using partial convolutions (PConv) with automatic mask updates. Yan et al. [26] proposed a model based on the U-Net [27] architecture to accurately inpainting missing regions in images from both structure and detail.

GAN-based face image restoration completion

The emergence of the generation confrontation network generates high-quality samples with a unique zero-sum game and confrontation training ideas. It has more powerful feature learning and feature expression capabilities than traditional machine learning algorithms, has achieved significant success in the field of sample generation, and has become a research hotspot in the field of image generation [28]. Gao et al. [29] proposed a two-stream network for image inpainting, respectively, reconstructing texture structure and

structure-constrained texture synthesis. At the same time, the bidirectional gated feature fusion (Bi-GFF) module and the context feature aggregation (CFA) module are designed inside the network to enhance global consistency and generate finer detailed features. Christian et al. [30] proposed SRGAN to infer realistic natural images with a 4x upscaling factor for image super-resolution (SR). Zhang et al. [31] proposed a DeBLuRring Network (DBLRNet) by applying 3D convolutions to perform spatio-temporal learning in both spatial and temporal domains for restoring blurred images in videos. Zhang et al. [32] Designed a multi-facial prior search network (MFPSNet) to optimally extract information from different facial priors for blind face restoration (BFR) tasks [33,34]. Ge et al. [35] proposed Identity Diversity GAN (ID-GAN), which integrates the face recognizer of CNN into GAN, uses CNN for feature reconstruction, and GAN for visual reconstruction, generating realistic and preserving identity feature images. Xu et al. [36] proposed a GAN-based feedforward network to generate high-quality face images, coupling facial features such as identity and expression through 3D priors, and extracting each facial attribute information separately.

3. Methods

In this section, we describe the proposed facial restoration generative networks T-GANs. The network contains two parts, the transformer module and the ResNet-based image generation network, as shown in Figure 1. It is well known that convolutional computing is weak in perceiving global feature information, and the generated image cannot achieve the desired effect if the missing face image is computed and generated directly. We use the transformer module to extract global feature information from its original image before image restoration generation and confirm the missing feature part of the face at the same time. In the face restoration generation network, we use ResNet as a generator for image restoration generation. In the discriminator, global and local discriminators are used to improve the detail discrimination performance of the generated images, and the transformer module is added to further discriminate the quality of the generated face-restored images.
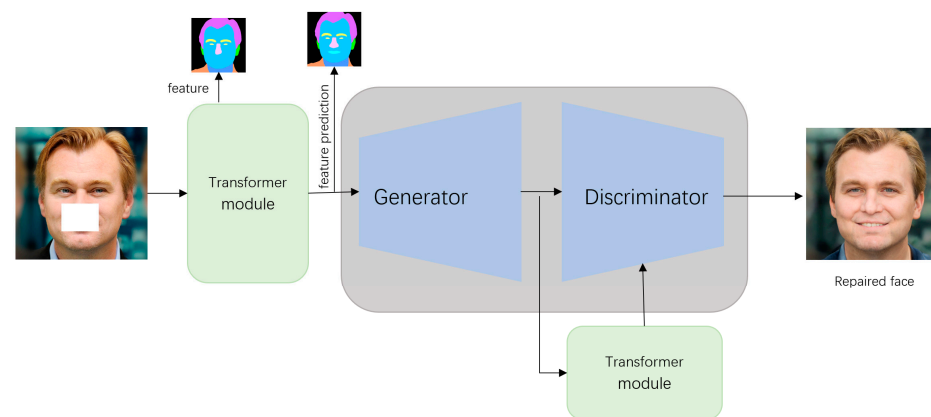

Figure 1. T-GANs framework.

3.1. Transformer Module

In the image generation task, compared to the global feature image generation from nothing, the face part feature generation has to generate both the missing feature image parts and maintain the uniformity and naturalness of the overall face image features, and the simple use of GANs-based generative network models cannot achieve the desired results. In the network, we use the transformer module of the original framework to extract the global features of the face and detect the missing feature parts under the multi-layer perceptron and multi-head self-attention calculations, which are used to repair and generate the missing parts of the face, as shown in Figure 2. The module contains an MLP layer, multi-head self-attention layer, layer-norm layer, etc., which divide the input missing face image into multiple image patches, where each patch can be regarded as a language 'word'

similar to NLP (Natural Language Processing), through the multi-head attention feature extraction of each patch, after multi-layer calculation obtains the global feature information and confirms the missing part of the feature.
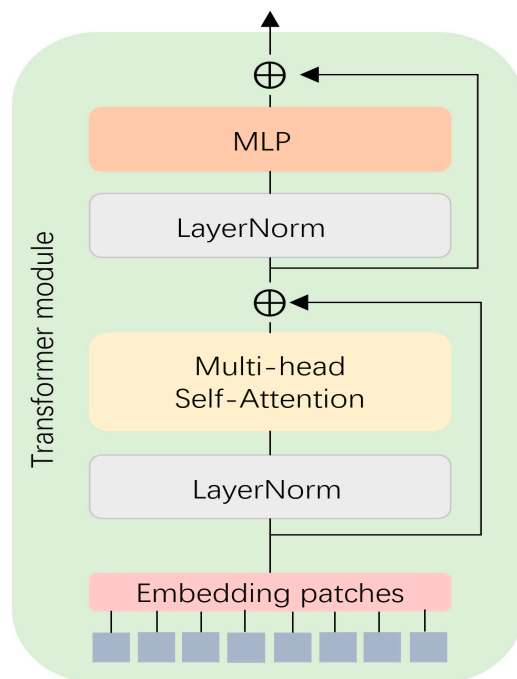

Figure 2. Transformer module architecture.

Where the multi-head attention is denoted as:

$$\text{head}_i = \text{Attention}\left(QW_i^Q, KW_i^K, VW_i^V\right)$$
$$\text{multi} - \text{head}(Q, K, V) = \text{Concat}(\text{head}_1, \ldots, \text{head}_h)W^O \tag{1}$$

In the equation, K, V, and Q represent the key, value, and query of the matrix packing sets. W^o is the projection matrix of multi-head output; W^Q, W^K, and W^V represent their corresponding different subspace correspondence matrices.

3.2. Facial Restoration Generative Network

In the facial restoration network, we use the GAN framework and reconstruct the internal generators and discriminators, as shown in Figure 3. In the generator part, we adopt Resnet as the main network and use dilated convolution instead of upsampling and downsampling operations. The size of the perceptual field of view affects the generation of image texture features during image generation; therefore, we use dilated convolution [37] to reduce excessive convolution operations while increasing the perceptual field of view. Resnet's use of multiple residual blocks avoids the appearance of image generation overfitting, generates more detailed image texture, and generates missing face part features and global facial features to maintain consistency.

In the discriminator part, a combination of global and local discriminators is used to discriminate the authenticity of the generated images. The global discriminator consists of multiple convolutional layers and a fully connected layer. All convolutional layers use a 2 × 2-pixel step to reduce the image resolution. A similar structure is used for the local discriminator, except that the size of the input image block is half that of the global discriminator. The authenticity of the generated face restoration images is finally confirmed by integrating the global and local discriminator outputs and the transformer module's discriminator results for the generated images in the sigmoid activation function calculation.
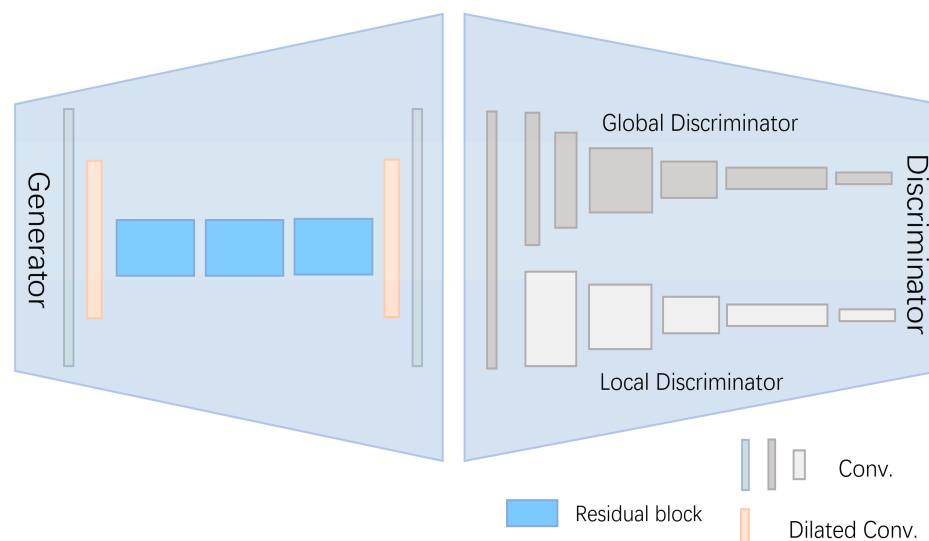
Figure 3. Facial restoration generative network architecture.

The face restoration generation network finally generates fake images after the continuous game and update between the generator and the discriminator. Its loss function is as follows:

$$L_{\text{gan}}(\mathrm{D},\mathrm{G}) = \mathbb{E}_{\boldsymbol{x} \sim p_{\text{data}}(\boldsymbol{x})}[\log D(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{z} \sim p_z(\boldsymbol{z})}[\log(1 - D(G(\boldsymbol{z})))] \tag{2}$$

In the equation, D represents the discriminator, G represents the generator, and z is the corrupted image.

4. Experiments

4.1. Datasets

We selected four datasets for evaluation in our experiments (Figure 4): VGG Face2 [38], Celeba-HQ [39], FFHQ [40], and EDFace-Celeb [41]. The VGG Face2 dataset is made of around 3.31 million images divided into 9131 classes, each representing a different person's identity and the images have an average resolution of 137×180 pixels. Images are downloaded from Google Image Search and have large variations in pose, age, illumination, ethnicity, and profession (e.g., actors, athletes, and politicians). The CelebA-HQ contains 30,000 face images with a 1024×1024 resolution. The FFHQ contains 70,000 high-quality PNG images at a resolution of 1024×1024 with considerable variation in age, ethnicity, and image background. The EDFace-Celeb is a publicly available ethnically diverse facial dataset that includes 1.7 million photos covering different countries with a balanced ethnic composition.

4.2. Implementation Details

Our facial restoration generative networks, T-GANs, contain two parts and three different network architectures, so they need to be trained separately for optimization to achieve the best results. In the transformer module, we set the initial learning rate to 0.0001, and the generator learning rate in the generative network is set to 0.001. If the discriminator learning rate is too high it will directly affect the generation quality, and its learning rate should not be higher than the generator learning rate to prevent the generators from failing to converge and the network from crashing. The overall network is optimized using the Adam optimizer [42], and the batch size is set to 20. Additionally, we added the beam search algorithm to optimize the learning rate to obtain the best performance of the network.
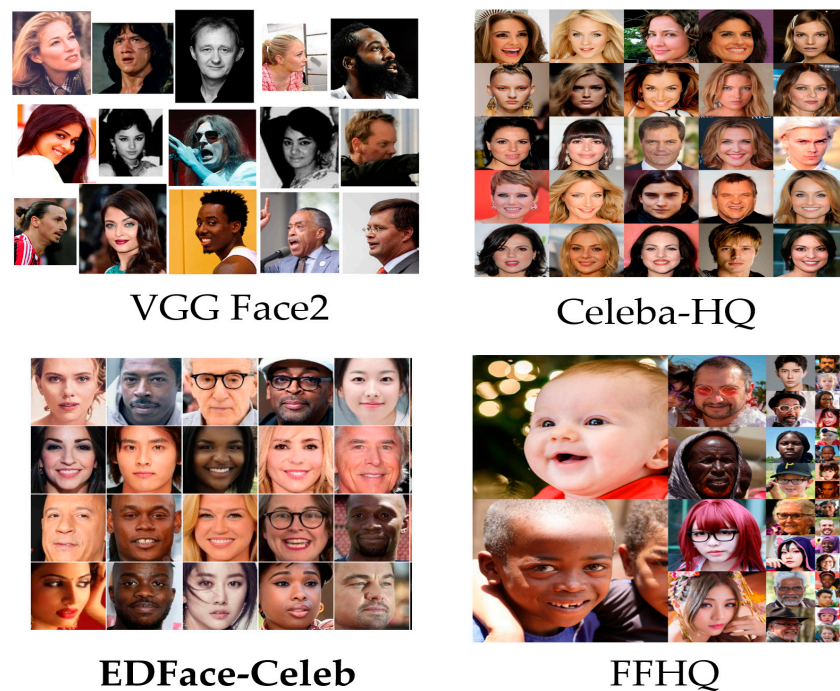
Figure 4. Open-source datasets.

In conducting the quality assessment of the generated images, we evaluated the network performance of our proposed network using the commonly used FID (Frechet inception distance) image generation quality assessment method; however, with a single assessment method, the image quality could not be verified from various aspects. We also used the PSNR (peak signal-to-noise ratio) and SSIM (structural similarity index map) to evaluate the quality of the images generated by the proposed networks.

4.3. Main Results

In the local feature repair generation task, compared to the global image generation task, the generated part often appears to be inconsistent with the surrounding area features, and the texture features appear distorted. At the same time, artifacts appear on the edges of the generated features, and the repaired areas are clearly distinguished from other normal areas, resulting in unnatural images. The occurrence of this situation is mainly due to the inability of most generative networks to accurately detect missing feature regions and specific missing features.

Our proposed network performs the restoration of the image by first using the transformer module to compute the global features of the image with multiple attention, detect the missing feature regions, and compute the correct features of the missing faces, as shown in Figure 5. It solves the problem of the repaired image anomalies that occur in most of the generative networks due to the inability to confirm the repaired content. For images missing different facial features, as shown in Figure 6, our proposed network can accurately detect the feature missing locations and missing contents, and repair them accurately to generate reasonable facial images, while maintaining the integrity and naturalness of the image features.

For large areas of missing facial feature image repair, such as the emergence of pandemic viruses nowadays when it is necessary to wear masks that obscure large areas of facial features, the face image cannot be accurately recognized. As shown in Figure 7, our network can still accurately perform reasonable face feature generation for the masked area.
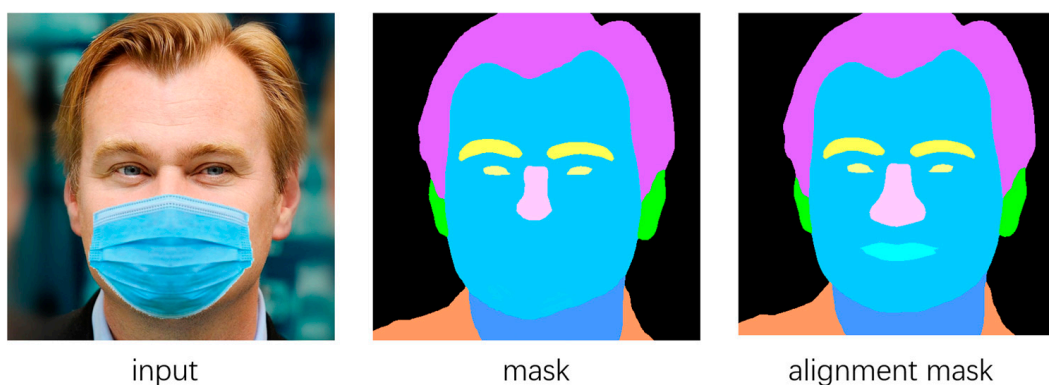
Figure 5. Facial missing feature mask.

Figure 6. The missing features of different parts of the face generate contrast.

In the evaluation of generated image quality, we use the FID image quality evaluation method to test the quality of images generated by the restoration of missing blocks (mask) of different sizes. The results of our proposed generative network and other excellent generative networks are shown in Table 1. Our proposed generation network performed better. The missing content of different sizes can be accurately repaired to achieve the effect of real ones.

Table 1. Different sizes of missing regions; image quality comparison generated by different models under the FID.

	Mask				
	10–20%	**20–30%**	**30–40%**	**40–50%**	**Center**
GAFC [43]	7.29	15.76	26.41	38.85	7.50
PIC [44]	6.57	12.93	20.12	33.71	4.89
Region Wise [45]	7.05	15.53	24.58	31.47	8.75
Edge Connect [46]	5.37	9.24	17.35	27.41	8.22
Ours	**4.35**	**7.23**	**12.41**	**17.29**	**4.91**

Figure 7. Large-area face feature loss (wearing a mask) generated result comparison.

We also validated the performance of the proposed network using PSNR and SSIM image quality evaluation methods, as shown in Table 2; our model is more robust in the restoration generation of missing parts (mask) of different areas compared to other good models.

Table 2. Performance comparison of PSNR and SSIM for different models in the face of different missing content restoration.
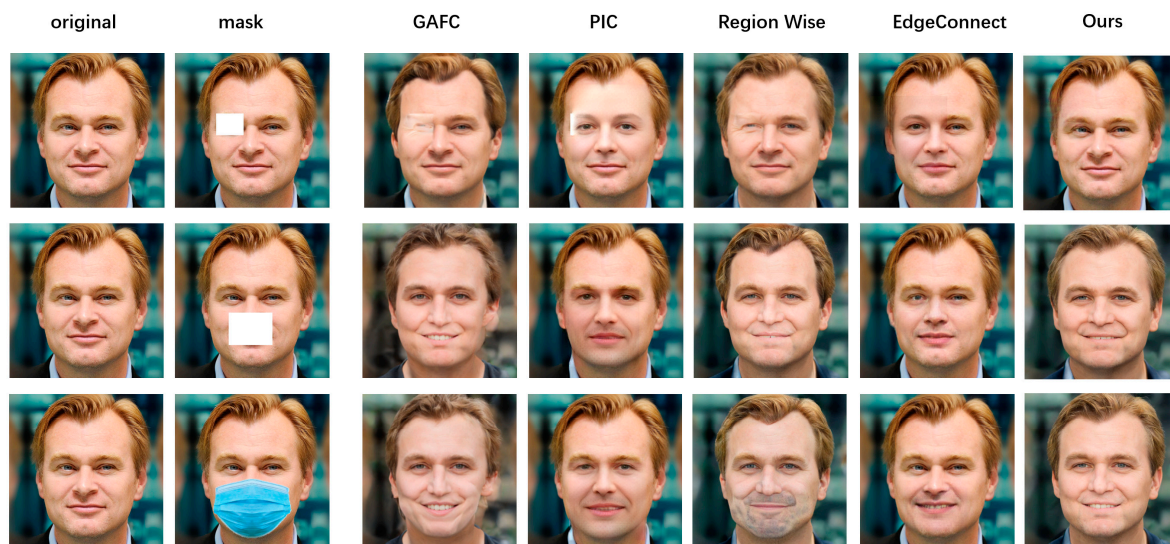
	Mask	GAFC	PIC	Region Wise	Edge Connect	Ours
	10–20%	27.51	30.33	30.58	30.73	**35.17**
	20–30%	24.42	27.05	26.83	27.55	**29.53**
PSNR	30–40%	22.15	24.71	24.75	25.21	**27.01**
	40–50%	20.30	22.45	22.38	23.50	**25.27**
	center	24.21	24.27	24.05	24.79	**29.74**
	10–20%	0.925	0.962	0.963	0.971	**0.983**
	20–30%	0.891	0.92.7	0.930	0.941	**0.974**
SSIM	30–40%	0.832	0.886	0.889	0.905	**0.939**
	40–50%	0.760	0.829	0.855	0.859	**0.907**
	center	0.865	0.869	0.871	0.874	**0.925**

In addition, we compared the proposed model with other good models in terms of computational efficiency, testing the average running time and the FLOPs during each image, as shown in Table 3.

Table 3. Efficiency comparison of different models.

	FLOPs	Time
GAFC	103.1 G	1.74 s
PIC	109.0 G	1.62 s
Region Wise	114.5 G	1.82 s
Edge Connect	122.6 G	2.05 s
Ours	95.5 G	1.03 s

The task of facial local feature repair generation is difficult due to the complexity of facial features, and it is easy to have problems such as distorted features, poor quality, and unnaturalness in the images generated by the repair. As shown in Figure 8, our proposed generative network is compared with other advanced models in the task of face image restoration for different areas and locations with missing features, and the generated face images suppress the appearance of artifacts well and generates features with higher accuracy while maintaining face feature consistency and naturalness.

Figure 8. Comparison of the restoration effect of different generative networks for different missing facial features.

5. Discussion

Our proposed face restoration generation network can perform effective restoration of the face of multiple scenarios of face loss and generate high-quality, natural-looking face images. At the same time, we test the performance of our network on face images with more than 50% missing facial features or even those missing the whole facial features with only hair features, as shown in Figure 9. Our proposed network still performs well and can repair their image to generate reasonable facial features and maintain the global feature consistency and naturalness of the face images. In the following research, we have improved the performance of our proposed network model in a near step to enhance the image quality, generating reasonably natural face images while preserving the background features and the accurate restoration generation of the effect of lighting on skin tones.
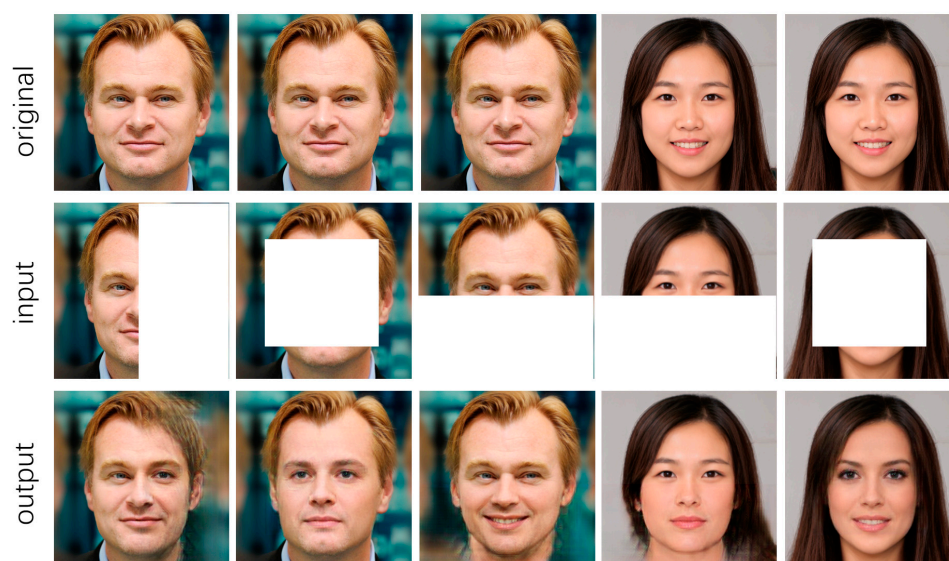
Figure 9. Image restoration generation for large facial feature loss.

6. Conclusions

We propose a novel T-GANs image generation network for the restoration generation task of missing feature parts of faces. In the network, the transformer module is added to effectively obtain global information on the image by using its self-attention mechanism to focus on the missing features of the face while considering all facial features. In the face restoration generation network, a ResNet-based network is used as the base network, and a dilated convolutional computation is added as the generator for the generation of missing faces. A dual discriminator is used in the discriminator to improve the ability to perceive both local and global features of the image. Compared to other conventional GANs-based image generation networks, we address the problem faced in maintaining the consistency and naturalness of the global features of images while generating local images in the image generation task, especially in the face feature missing restoration task. Our model is validated using different datasets and various image quality evaluation methods, where the FID is 17.29, PSNR and SSIM reach 25.27, and 0.907 is achieved in the image quality validation generated by large-area feature missing (40–50%) restoration. Additionally, our network is demonstrated to be more robust compared to other state-of-the-art networks; its generated face images perform better in terms of image quality, image feature uniformity, and naturalness.

References

1. Boutros, F.; Damer, N.; Kirchbuchner, F.; Kuijper, A. Unmasking Face Embeddings by Self-restrained Triplet Loss for Accurate Masked Face Recognition. *arXiv* **2021**, *preprint*. arXiv:2103.01716.
2. Damer, N.; Grebe, J.H.; Chen, C.; Boutros, F.; Kirchbuchner, F.; Kuijper, A. The effect of wearing a mask on face recognition performance: An exploratory study. In Proceedings of the 19th International Conference of the Biometrics Special Interest Group, BIOSIG 2020, Online, 16–18 September 2020.
3. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016; IEEE Computer Society, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
4. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. Ima-geNet large scale visual recognition challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [CrossRef]
5. Efros, A.A.; Freeman, W.T. Image quilting for texture synthesis and transfer. In Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques, Los Angeles, CA, USA, 12–17 August 2001; pp. 341–346.
6. Efros, A.A.; Leung, T.K. Texture synthesis by non-parametric sampling. In Proceedings of the Seventh IEEE International Conference on Computer Vision, Corfu, Greece, 20–27 September 1999; Volume 2, pp. 1033–1038.
7. Darabi, S.; Shechtman, E.; Barnes, C.; Goldman, D.B.; Sen, P. Image melding: Combining inconsistent images using patch-based synthesis. *ACM Trans. Graph.* **2012**, *31*, 1–10. [CrossRef]
8. ACriminisi, P.P.; Toyama, K. Region filling and object removal by exemplar-based image inpainting. *IEEE Trans. Image Process.* **2004**, *13*, 1200–1212. [CrossRef] [PubMed]
9. Satoshi, I.; Edgar, S.S.; Hiroshi, I. Globally and locally consistent image completion. *ACM Trans. Graph.* **2017**, *36*, 1–14.
10. Wan, Z.; Zhang, J.; Chen, D.; Liao, J. High-Fidelity Pluralistic Image Completion with Transformers. *arXiv* **2021**, *preprint*. arXiv:2103.14031.
11. 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 (CVPP), Las Vegas, NV, USA, 27–30 June 2016; pp. 2536–2544.
12. Zeng, Y.; Fu, J.; Chao, H.; Guo, B. Learning pyramid-context encoder network for high-quality image inpainting. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 1486–1494.
13. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In Proceedings of the Advances in Neural Information Processing Systems (NeurIPS), Montreal, QC, Canada, 8–13 December 2014; pp. 2672–2680.
14. Denton, E.; Gross, S.; Fergus, R. Semi-Supervised Learning with Context-Conditional Generative Adversarial Networks. *arXiv* **2016**, *preprint*. arXiv:1611.06430.
15. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, *preprint*. arXiv:1409.1556.
16. 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, Las Vegas, NV, USA, 17–19 June 1997; pp. 5485–5493.
17. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 5998–6008.
18. Ballester, C.; Bertalmio, M.; Caselles, V.; Sapiro, G.; Verdera, J. Filling-in by joint interpolation of vector fields and gray levels. *IEEE Trans. Image Process.* **2001**, *10*, 1200–1211. [CrossRef]
19. Bertalmio, M.; Sapiro, G.; Caselles, V.; Ballester, C. Image inpainting. In Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques, New Orleans, LA, USA, 23–28 July 2000; pp. 417–424.
20. Xu, Z.; Sun, J. Image inpainting by patch propagation using patch sparsity. *IEEE Trans. Image Process.* **2010**, *19*, 1153–1165. [PubMed]
21. 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]
22. Russell, B.C.; Torralba, A.; Murphy, K.P.; Freeman, W.T. LabelMe: A database andWeb-based tool for image annotation. *Int. J. Comput. Vis.* **2008**, *77*, 157–173. [CrossRef]
23. 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, 18–22 June 2018; pp. 5505–5514.
24. Liu, H.; Jiang, B.; Xiao, Y.; Yang, C. Coherent semantic attention for image inpainting. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 4170–4179.
25. 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; pp. 85–100.
26. Yan, Z.; Li, X.; Li, M.; Zuo, W.; Shan, S. Shift-net: Image inpainting via deep feature rearrangement. In Proceedings of the European conference on computer vision (ECCV), Salt Lake City, UT, USA, 18–23 June 2018; pp. 1–17.
27. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Singapore, 18–22 September 2022; Springer: Cham, Switzerland, 2015; pp. 234–241.

28. Li, Y.; Liu, S.; Yang, J.; Yang, M.-H. Generative face completion. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 3911–3919.

29. Guo, X.; Yang, H.; Huang, D. Image inpainting via conditional texture and structure dual generation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 14134–14143.

30. Ledig, C.; Theis, L.; Huszar, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; et al. Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4681–4690.

31. Zhang, K.; Luo, W.; Zhong, Y.; Ma, L.; Liu, W.; Li, H. Adversarial spatio-temporal learning for video deblurring. *IEEE Trans. Image Process.* **2018**, *28*, 291–301. [CrossRef] [PubMed]

32. Yu, Y.; Zhang, P.; Zhang, K.; Luo, W.; Li, C.; Yuan, Y.; Wang, G. Multi-Prior Learning via Neural Architecture Search for Blind Face Restoration. *arXiv* **2022**, *preprint*. arXiv:2206.13962.

33. Zhang, P.; Zhang, K.; Luo, W.; Li, C.; Wang, G. Blind Face Restoration: Benchmark Datasets and a Baseline Model. *arXiv* **2022**, *preprint*. arXiv:2206.03697.

34. Wang, T.; Zhang, K.; Chen, X.; Luo, W.; Deng, J.; Lu, T.; Cao, X.; Liu, W.; Li, H.; Zafeiriou, S. A Survey of Deep Face Restoration: Denoise, Super-Resolution, Deblur, Artifact Removal. *arXiv* **2022**, *preprint*. arXiv:2211.02831.

35. Ge, S.; Li, C.; Zhao, S.; Zeng, D. Occluded face recognition in the wild by identity-diversity inpainting. *IEEE Trans. Circuits Syst. Video Technol.* **2020**, *30*, 3387–3397. [CrossRef]

36. Xu, Z.; Yu, X.; Hong, Z.; Zhu, Z.; Han, J.; Liu, J.; Ding, E.; Bai, X. Facecontroller: Controllable attribute editing for face in the wild. In Proceedings of the AAAI Conference on Artificial Intelligence, Vancouver, BC, Canada, 2–9 February 2021; Volume 35, pp. 3083–3091.

37. Yu, F.; Koltun, V. Multi-scale context aggregation by dilated convolutions. *arXiv* **2015**, *preprint*. arXiv:1511.07122.

38. Parkhi, O.M.; Vedaldi, A.; Zisserman, A. Deep Face Recognition. In *BMVC 2015—Proceedings of the British Machine Vision Conference 2015*; British Machine Vision Association: Durham, UK, 2015; pp. 1–12.

39. Lee, C.H.; Liu, Z.; Wu, L.; Luo, P. Maskgan: Towards diverse and interactive facial image manipulation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 5549–5558.

40. 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; pp. 4401–4410.

41. Zhang, K.; Li, D.; Luo, W.; Liu, J.; Deng, J.; Liu, W.; Zafeiriou, S. EDFace-Celeb-1 M: Benchmarking Face Hallucination with a Million-scale Dataset. In Proceedings of the IEEE Transactions on Pattern Analysis and Machine Intelligence, Santa Clara, CA, USA, 1 March 2023; Volume 45, pp. 3968–3978.

42. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, *preprint*. arXiv:1412.6980.

43. Song, L.; Cao, J.; Song, L.; Hu, Y.; He, R. Geometry-aware face completion and editing. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019; Volume 33, pp. 2506–2513.

44. Zheng, C.; Cham, T.J.; Cai, J. Pluralistic image completion. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–17 June 2019; pp. 1438–1447.

45. Ma, Y.; Liu, X.; Bai, S.; Wang, L.; Liu, A.; Tao, D.; Hancock, E.R. Regionwise generative adversarial image inpainting for large missing areas. *IEEE Trans. Cybern.* **2022**. [CrossRef] [PubMed]

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