Qingyu Liu [0009-0000-9774-808X]*,**, *Roben A. Juanatas*[0009-0000-0572-0785]*

MASK FACE INPAINTING BASED ON IMPROVED GENERATIVE ADVERSARIAL NETWORK

Abstract

Face recognition technology has been widely used in all aspects of people's lives. However, the accuracy of face recognition is greatly reduced due to the obscuring of objects, such as masks and sunglasses. Wearing masks in public has been a crucial approach to preventing illness, especially since the Covid-19 outbreak. This poses challenges to applications such as face recognition. Therefore, the removal of masks via image inpainting has become a hot topic in the field of computer vision. Deep learning-based image inpainting techniques have taken observable results, but the restored images still have problems such as blurring and inconsistency. To address such problems, this paper proposes an improved inpainting model based on generative adversarial network: the model adds attention mechanisms to the sampling module based on pix2pix network; the residual module is improved by adding convolutional branches. The improved inpainting model can not only effectively restore faces obscured by face masks, but also realize the inpainting of randomly obscured images of human faces. To further validate the generality of the inpainting model, tests are conducted on the datasets of CelebA, Paris Street and Place2, and the experimental results show that both SSIM and PSNR have improved significantly.

1. INTRODUCTION

Image inpainting is the filling of missing areas of an image with pixels to make it as close to the original image as possible. Image inpainting or image completion has been one of the research focuses in the field of computer vision. Face image inpainting differs from general image inpainting in that the human face has distinctive features and complex geometric structures, especially the mouth, nose, and eyes, which contain rich and high-level semantic information(Jiang et al., 2022). Face image inpainting pays more attention to the rationality of image structures and semantics, and preserves the unique identity attributes of individuals.

Traditional inpainting algorithms(Ding et al., 2019; Jia & Tang, 2003; Simakov et al., 2008) and deep learning-based inpainting algorithms(He et al., 2022; Shao et al., 2022; Zeng et al., 2019; Zheng et al., 2019) are the two primary categories of image inpainting. Traditional inpainting algorithms are mainly based on the diffusion principle, which

---

* National University, College of Computing and Information Technologies, Philippines,
liuq@students.national-u.edu.ph, rajuanatas@national-u.edu.ph
** Huainan Normal University, School of Computer Science, China, liuqy@hnnu.edu.cn

propagates neighbourhood pixels from intact areas to the missing areas and uses pixel similarity to complete image inpainting. Such methods can restore small missing areas, but they are less effective for large missing areas or images with meaningful texture structures. In recent years, due to the advent of convolutional neural networks and deep learning, many new inpainting models have emerged. Deep learning-based inpainting algorithms can effectively extract information from unobscured regions and generate pixels with rich semantics to fill the missing regions. The emergence of generative adversarial network technology has pushed image inpainting to a new level (Goodfellow, 2014). Context Encoder (CE) algorithm used adversarial training of the discriminator and generator to obtain the generator to complete the image inpainting for the first time (Pathak et al., 2016). The images restored by CE have obvious blurring. To enhance the effect of image inpainting, algorithms such as GLCIC (Globally and Locally Consistent Image Completion) (Iizuka et al., 2017), GMCNN (Generative Multi-column Convective Neural Networks) (Wang et al., 2018) and PEPSI (Parallel Extended-decoder Path for Semantic in Inpainting) (Sagong et al., 2019) appeared one after another on the basis of CE, but still did not eliminate the problems of blurring and artifacts.

In our paper, a new face inpainting model based on generative adversarial network is proposed by improving the original model of pix2pix (Isola et al., 2017). Spatial attention mechanism (Jaderberg et al., 2016) and channel attention mechanism (Zhang et al., 2018) are introduced in the Encoder and Decoder of the model, while multi-branch residual modules are added in the down-sampling to improve the model's capacity for feature extraction, so as to better accomplish image inpainting. The improved inpainting model is tested on the face mask dataset with better inpainting results. In order to confirm the model's efficacy, we also do experiments on Celeb A, Paris Street and Place2 datasets, and the results show that the inpainting of regular mask regions and random mask regions can be accomplished with a better visual performance. Figure 1 shows some examples of the inpainting results of our method.

| (a) Ground truth | (b) Masked imges | (c) pix2pix | (d) SRGan | (e) Ours |

Fig. 1. Inpainting results of our method

The following are the contributions of this paper:

(1). Construction of the face mask dataset based on public CelebA.

(2). Improvement of the UNet structure and proposition of a face image inpainting model that can effectively remove face masks.

(3). Improvement of the residual module of the single branch by adding multiple branches with different convolution kernel sizes, and the model feature extraction capability is enhanced. The number of branches can be added more if needed.

(4). One color picture contains both spatial and channel information. Spatial attention and channel attention are subtly blended into one model. This is crucial for face image inpainting.

2. RELATED WORK

2.1. Image Inpainting based on Gan

Goodfellow et al. introduced a novel framework for building generative models through adversarial training in their paper. The framework simultaneously trains the discriminator network D and the generator network G. The generator network generates new images with the aim of "fooling" the discriminator as much as possible, and we can use the generator to restore images.

Many methods of image inpainting have emerged on the basis of Gan. Yu et al.(Yu et al., 2018) proposed a depth generation model based on a contextual attention mechanism, which "borrows" features and pixels from relatively distant regions. During the training process, the surrounding image features are explicitly used as a reference, and the long-range correlation between the occluded and other regions is established by using an inflated convolution with a large perceptual field. The idea of partial convolution is further generalized by using gated convolution(Yu et al., 2019) to extend the feature selection mechanism of each layer in the network to learnable location information, and hand-drawn sketches are used to guide the inpainting process. Liu(Liu et al., 2018) proposed partial convolution to ensure that the pixel responses in the masked region are all synthesized from the unmasked region. And the occlusion mask is automatically updated as the effective region keeps increasing during network forwarding to improve the model's inpainting performance for non-regular occlusions.

The UNet structure is useful in Gan models because of its symmetry, which can effectively reduce the model's parameters. Ronneberger(Ronneberger et al., 2015) first proposed the UNet network model, which was used to capture the sampling structure and symmetric connections, replacing sliding window convolution and fully connected layers with better restoration. Zhou(Zhou et al., 2018) proposed the UNET++ network framework based on the UNet network, and replaced the original skip-connection with a dense connection, which improved the segmentation effect of the model. Mou(Mou et al., 2022) proposed a new model DGUNET(Deep Generalized Unfolding Network) based on the UNet network structure model, which was not friendly to large corrupted images.

2.2. Attention

Attention mechanism is a machine learning and deep learning data processing technique that is widely utilized in numerous learning tasks, including image processing, and natural

language processing. In essence, the attention mechanism resembles how people observe their environment. In image processing, the most used attention mechanisms are spatial attention (SA) and channel attention (CA). These two attention mechanisms have different levels of resource allocation. SA locates the spatial region of interest for transformation to obtain the weights, while CA is to allocate the resources among the convolutional channels. Both of them focus on different parts, but both emphasize on focusing on the important information and ignoring the unimportant information. Yu(Wu et al., 2020) proposed a new depth-based generative model for large broken images and added contextual attention. They used convolution to calculate the matching scores of foreground and background images and then applied softmax for comparison to obtain the attention scores of each pixel. Xie C et al. in the literature(Xie et al., 2019) found the incoherence of existing methods for restoring the missing region of a hole, and therefore proposed a deep generative modelling method based on refinement. The method used a new semantic attention layer to not only preserve the contextual structural information but also to model the information around the hole.

3. AUTHORS' METHOD

In order to effectively recover the face obscured area by the mask, an improved generative adversarial network model is proposed in this paper based on the UNet structure. The whole model contains two sub-networks, the generator and the discriminator. The down-sampling modules of the generator incorporate the spatial attention mechanism, while the up-sampling modules use both the channel attention mechanism and the spatial attention mechanism. In addition, the authors improve the common residual module by adding convolutional branches using convolutional kernels of different sizes to increase the feature extraction capability in the down-sampling module. Figure 2 displays the model's overall framework.
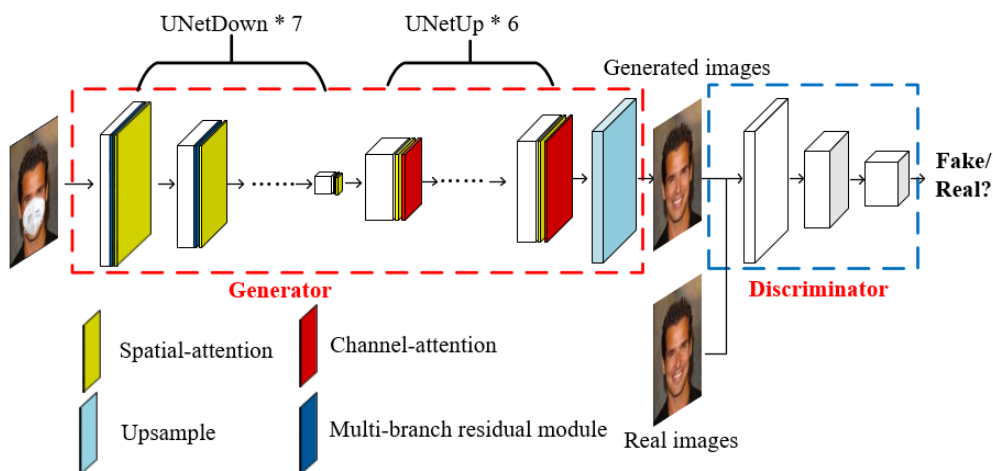

Fig. 2. Overall framework

28

The masked face images are input to the generator network, and the faces without masks of the same size are obtained by convolution operations. The output images of the generator network and the real face images are both fed into the discriminator network at the same time to complete the determination of true or false. In this iteration, the generator network and the discriminator network are trained against each other to gradually improve their respective performance. We use the stable generator for face image inpainting.

3.1. Generator network

Figure 3 illustrates the general structure of the authors' generator network, where all images are uniformly resized to 128*128. The generator network consists of 7 down-sampling modules, 6 up-sampling modules and 1 up-sampling convolutional layer. The down-sampling and up-sampling modules perform the 'Concatenate' operation at the same size level to better synthesize the extracted features from the bottom to top layers. The resolution of the feature images becomes smaller during down-sampling, and consequently, some of the effective features are lost. Therefore, a multi-branch residual module and a spatial attention mechanism to each down-sampling module to compensate for the lost information was added.
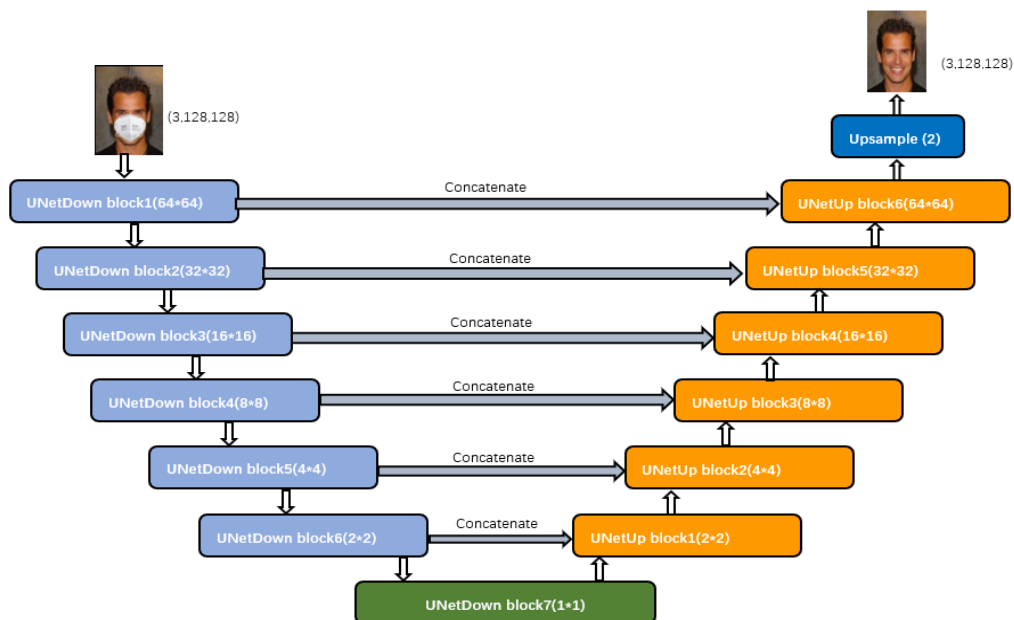

Fig. 3. Generator network

The original residual module has only one convolution branch, and the result of convolution and the input images are directly fused for processing. Considering that different sizes of convolution kernels can extract features at different levels, we transform the residual module into three branches with convolution kernels sizes 1*1, 3*3 and 5*5. Depending on the specific situation, we can also add more branches, as shown in figure 4. Each convolution branch can control the scaling of the image by the stride.
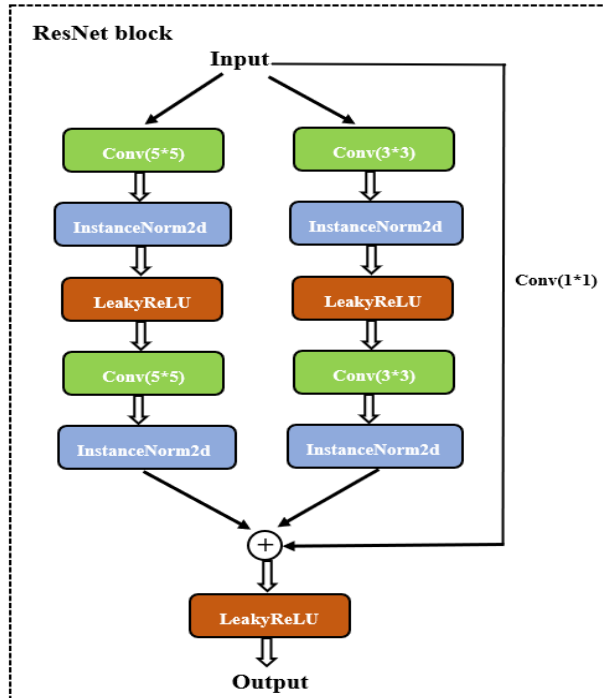
Fig. 4. Improved ResNet block

Two improved residual modules are used consecutively in the down-sampling module, one of which has a stride set to 2 to reduce the feature map resolution. The spatial attention module is connected behind the residual module to focus on the more important pixel information in space. Figure 5 displays the down-sampling's structure.
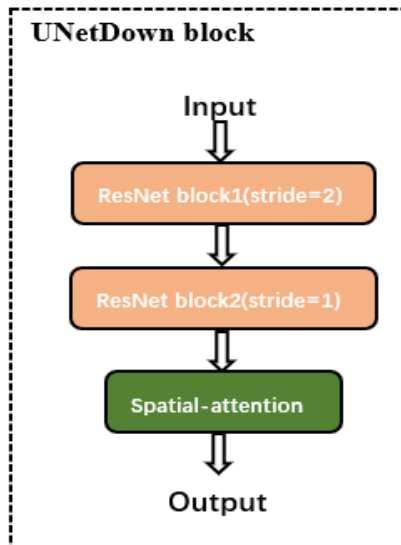

Fig. 5. UNetDown block

Figure 6 shows the up-sampling module, which reconstruct the feature image to the size of the original input by successive up-sampling. Each module contains residual module, spatial attention and channel attention in turn. The channel attention must be added after the 'Concatenate' operation in order to better extract useful information between channels.
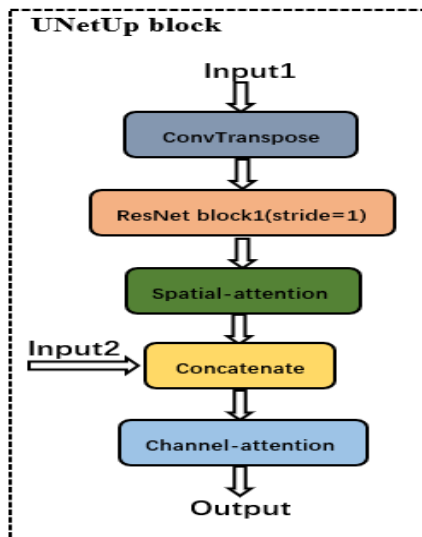
Fig. 6. UNetUp block

3.2. Discriminator network

The discriminator network's primary job is to judge whether the input images are fake or real, so its input size is the same as the generator's output images. Figure 7 demonstrates that the output of the discriminator network is a 1*16*16 matrix, where each pixel point represents the discriminant value of a small region of the input image. This structure makes the discriminator network do a whole discriminant for a small region with better results.
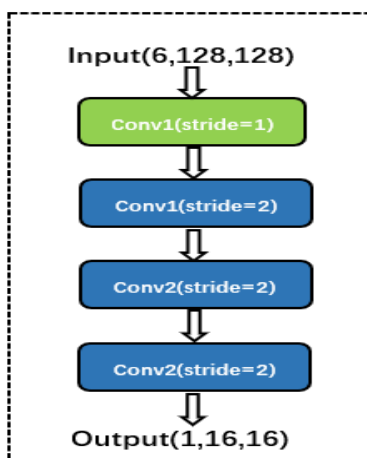
Fig. 7. Discriminator network

3.3. Loss function

During the generative adversarial network training process, the loss functions are essential. The total loss of the generator network includes L1 loss, adversarial loss and TVLoss (Rudin et al., 1992), while the discriminator network's loss uses only adversarial loss. L1 loss enables pixel-level evaluation of the masked images and the original images, which plays a key role in the outcome of image inpainting. The adversarial loss is mainly used to help the alternate, adversarial training of the generator network and the discriminator network, thus improving the respective performance of the 2 sub-networks. During image inpainting, any noise will have a relatively large impact on the original image result, especially the boundary of the occluded area will become incoherent. TVLoss, as a regular term, can keep the image boundary smooth after inpainting.

$$Loss_G = \gamma_{L1} * Loss_{L1} + \gamma_{adv} * Loss_{adv} + \gamma_{TV} * Loss_{TV} \qquad （1）$$

$$Loss_D = Loss_{adv}(D) \qquad （2）$$

In Equation (1), $\gamma_{L1}, \gamma_{adv}, \gamma_{TV}$ denote the weights of losses respectively.

4. EXPERIMENTAL RESULTS AND ANALYSIS

Based on the CelebA face dataset, we select the images of the front faces and manually add the three most common masks to form the mask face dataset. Meanwhile, to confirm the generality of the inpainting model, we also conduct tests on Celeb A, Place2 and Paris Street datasets respectively. The test results are compared with two classical methods, pix2pix and SRGan (Ledig et al., 2017). The optimizer used for model training is Adam, and b1, and b2 are 0.5, 0.999 respectively. The learning rate of the discriminator network and generator network is both set to $2*10^{-4}$. The weights of the generator loss function are γ_{L1}=100, γ_{adv}=100, γ_{TV}=1.

4.1. Datasets

The Chinese University of Hong Kong created the CelebFaces Attributes Collection (CelebA), a sizable face attribute dataset that includes 202599 photos of 10177 famous identities. The CelebA dataset is one of the most commonly used datasets for model training or testing related to face image processing. This dataset is very comprehensive and includes Europeans, Asians, men, women, older, middle-aged, laughing, serious, stationary, athletic, etc. Considering the need of our face inpainting experiment, 140000 images with frontal faces are randomly selected from the whole dataset, among which, the ratio of men to women is about 1:1. In addition, this paper takes the 3 most common masks that people use in daily life for construction. During the process, Photoshop software is used for mask masking and this work produces paired images of masked and unmasked faces as shown in figure 8. The left part is the original face and the right is the masked face. These two parts are put together for comparison.

Fig. 8. Mask face dataset

In the experiment, in addition to the CelebA dataset, we also use the Place2 and Paris Street datasets, and the distribution of each dataset is shown in table 1.

Tab. 1. Distribution of datasets

| Datasets	Training set	Test set
Mask face dataset	29261	500
CelebA	30000	500
Paris Street	25000	500
Place2	30000	500

4.2. Qualitative Evaluation

Both pix2pix and SRGan algorithms have some blurring in the inpainting results, and cannot generate semantically coherent pixels in the mask-obscured region. However, there are 'Concatenate' connections in the pix2pix inpainting model, which can generate complex organs such as the nose and mouth. Therefore, the overall effect of pix2pix inpainting is better than that of SRGan. The improved residual module can extract features of different dimensions, while spatial attention and channel attention can selectively focus on some of the key information in space and channels which are very important for face image inpainting. Figure 9 shows that the proposed face image inpainting model in this paper generates semantically reasonable content in addition to the best local texture details.
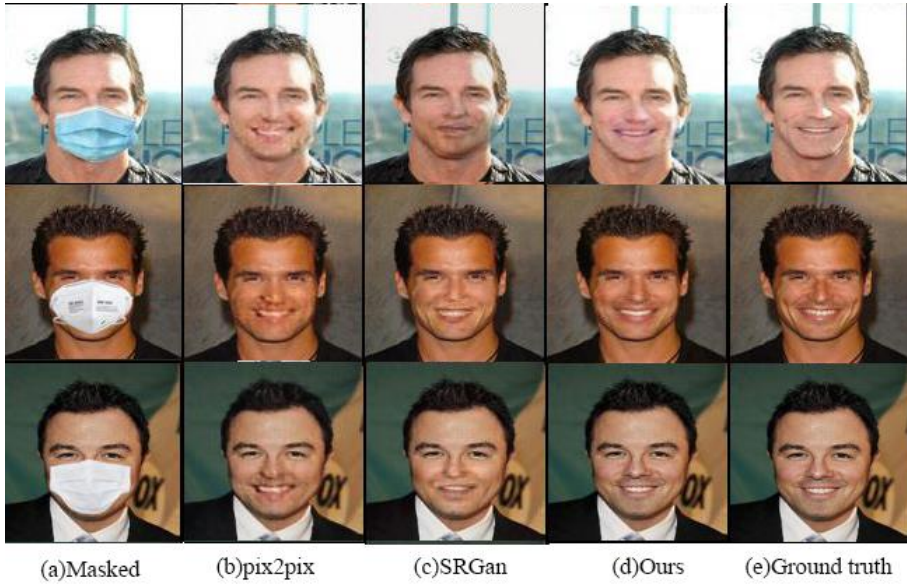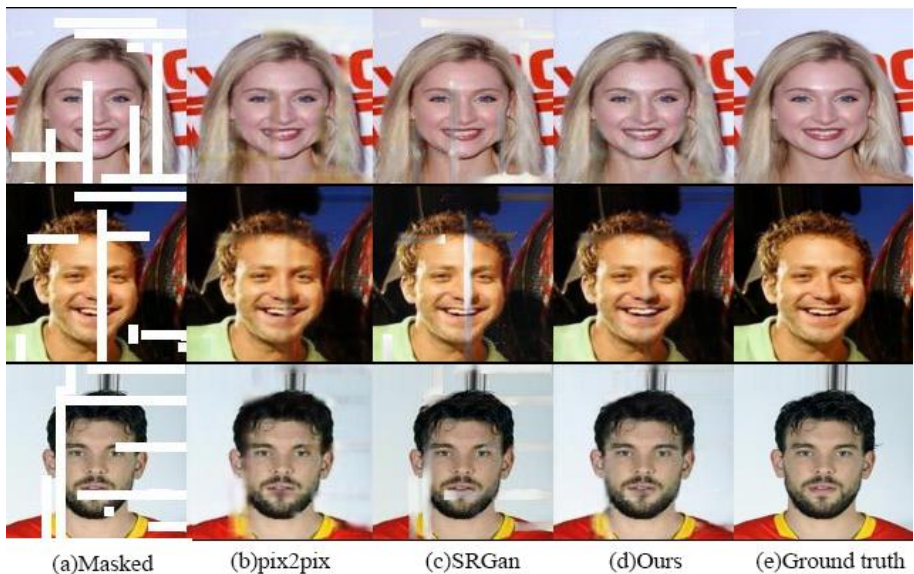
Fig. 9. Comparison results on mask occlusion faces

In the following figure, 64*64 size masks are used to replace the general face masks. The experimental results show that the algorithm in this paper is more reasonable and coherent in inpainting details, and highly restores the masked area. The other two inpainting algorithms, have poorer inpainting results, with significant blurring and distortion.
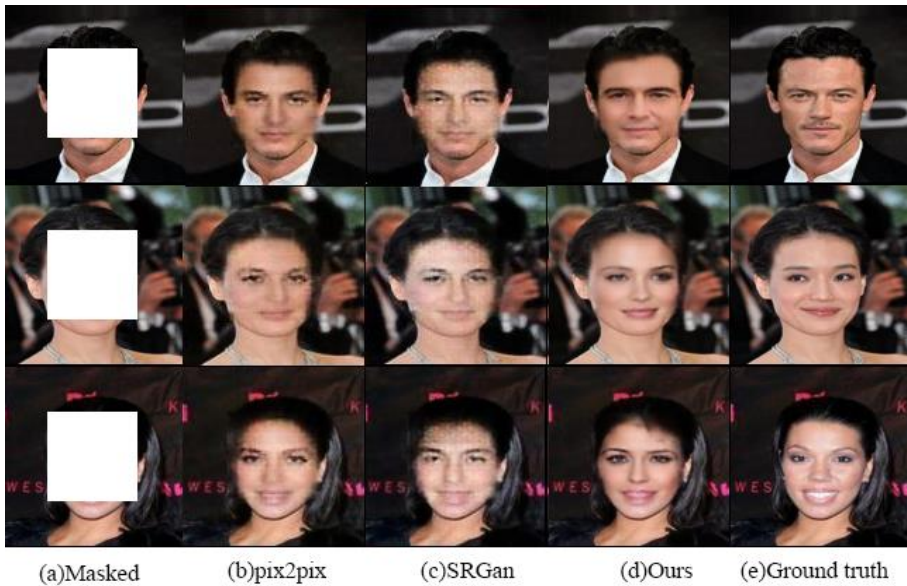
Fig. 10. Comparison results on rectangle occlusion faces

To further verify the inpainting effect of the model on face images, we perform random masks on face images. Figure 11 shows the comparison of random masks around 10%, and figure 12 shows the comparison of random masks around 30%. Pix2pix and SRGan have poor inpainting effects, and there are clear blurs and visual blindness. The comparison results show the obvious superiority of the face inpainting algorithm in this paper.
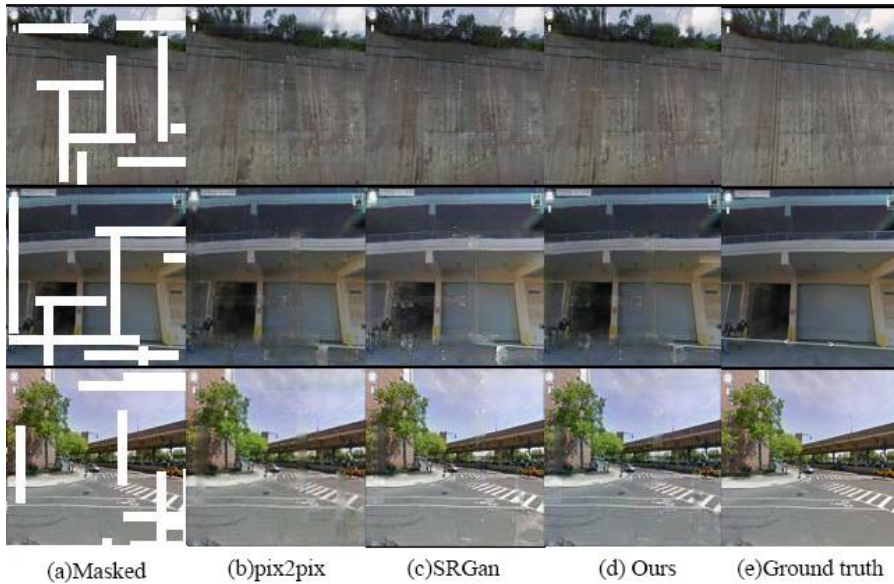

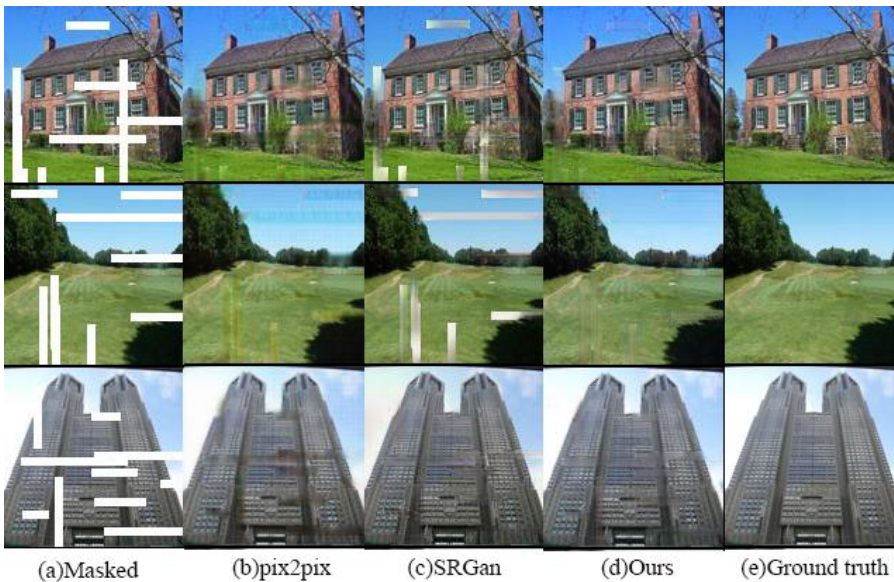**Fig. 11. Comparison results on random occlusion faces(10%)**

Fig. 12. Comparison results on random occlusion faces(30%)

The proposed inpainting model in this paper not only accomplishes the inpainting task well on face images, but also works well on other datasets. So we also conduct comparison experiments on Place2 and Paris Street datasets. The experimental results demonstrate that the inpainting results of pix2pix and SRGan algorithms have artifacts and blind areas, while the inpainting model in this paper highly restores the occluded areas as shown in figure 13 and figure 14.

(a)Masked   (b)pix2pix   (c)SRGan   (d) Ours   (e)Ground truth

Fig. 13. Comparison results on random occlusion Paris Street(10%)

(a)Masked   (b)pix2pix   (c)SRGan   (d)Ours   (e)Ground truth

Fig. 14. Comparison results on random occlusion Place2(10%)

4.3. Quantitative Evaluation

In order to better demonstrate the superiority of the inpainting method in this paper, we also perform a quantitative analysis of the inpainting results. PSNR and SSIM(Hore & Ziou, 2010) are used as evaluation metrics for the inpainting images, respectively. PSNR is generally used to measure the image quality value between the maximum signal and the background noise. A larger value of PSNR indicates less distortion in the inpainting image. SSIM is a metric that measures the similarity of two input images, and a larger value of SSIM indicates better image inpainting quality. Table 2-table 5 show the results of the quantitative analysis. It is very clear that the inpainting results of our proposed algorithm are better than pix2pix and SRGan.

Tab. 2. Quantitative comparison of mask faces

	pix2pix	SRGan	Ours
PSNR	27.72	26.56	29.61
SSIM	89.11%	89.24%	92.71%

Tab. 3. Quantitative comparison on rectangular occlusion with CelebA

	pix2pix	SRGan	Ours
PSNR	23.79	19.28	24.84
SSIM	84.02%	74.13%	85.59%

Tab. 4. Quantitative comparison on random occlusion with CelebA

Datasets		pix2pix	SRGan	Ours
CelebA (10% masked)	PSNR	26.47	24.48	28.03
	SSIM	86.93%	83.58%	89.28%
CelebA (30% masked)	PSNR	23.79	19.28	24.84
	SSIM	84.02%	74.13%	85.59%

Tab. 5. Quantitative comparison on random occlusion with Place2 and Paris Street

Datasets		pix2pix	SRGan	Ours
Place2 (10% masked)	PSNR	24.08	23.41	25.08
	SSIM	81.16%	81.62%	83.48%
Paris Street (10% masked)	PSNR	23.95	22.35	25.43
	SSIM	80.59%	79.41%	84.13

4.4. Loss function analysis

For generative adversarial network, the generator and discriminator are trained against each other, so adversarial loss is necessary. In addition to the adversarial loss, the total loss

function of the generator includes the L1 loss and the TVLoss loss. As we can see in figure 15, the final face inpainting results are very poor if L1 loss is not used, which enables pixel-level processing. And for TVLoss, it can make the restored image smoother, especially at the boundary of the occluded area.
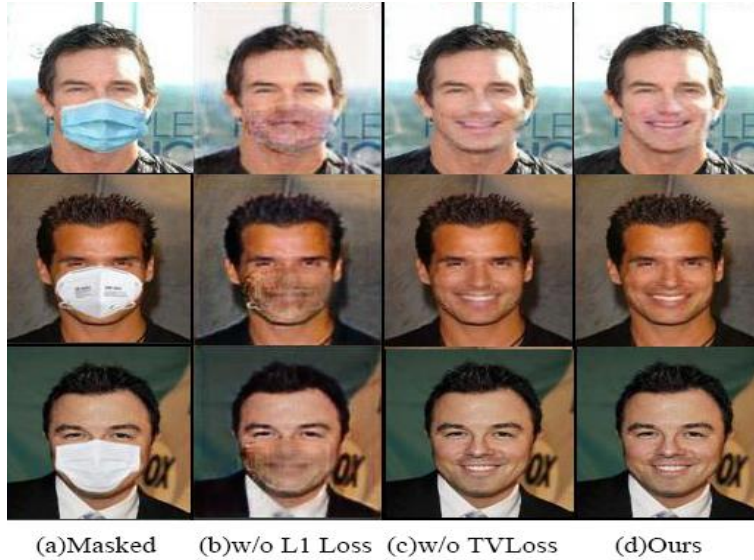

(a)Masked   (b)w/o L1 Loss (c)w/o TVLoss   (d)Ours

Fig. 15. Qualitative comparison of loss function

The quantitative analysis is performed in table 6, and the use of both L1 Loss and TVLoss loss leads to the improvement of both PSNR and SSIM2. These two losses play a good role in the whole image processing.

Tab. 6. Quantitative comparison of loss function

	w/o L1 Loss	w/o TVLoss	Ours
PSNR	23.69	27.85	29.61
SSIM	85.85%	90.25%	92.71%

4.5. Ablation Experiment

In addition to performing quantitative and qualitative analyses, we also design ablation experiments to demonstrate the role of each module in the overall model. The experimental protocols are set up without channel attention, without spatial attention and without the improved residual module, respectively. The ablation experiments are still tested on the constructed mask faces dataset shown in figure 16.

(a)Masked   (b)w/o channel attention   (c)w/o spatial attention   (d)w/o Resnet module   (e)Ours
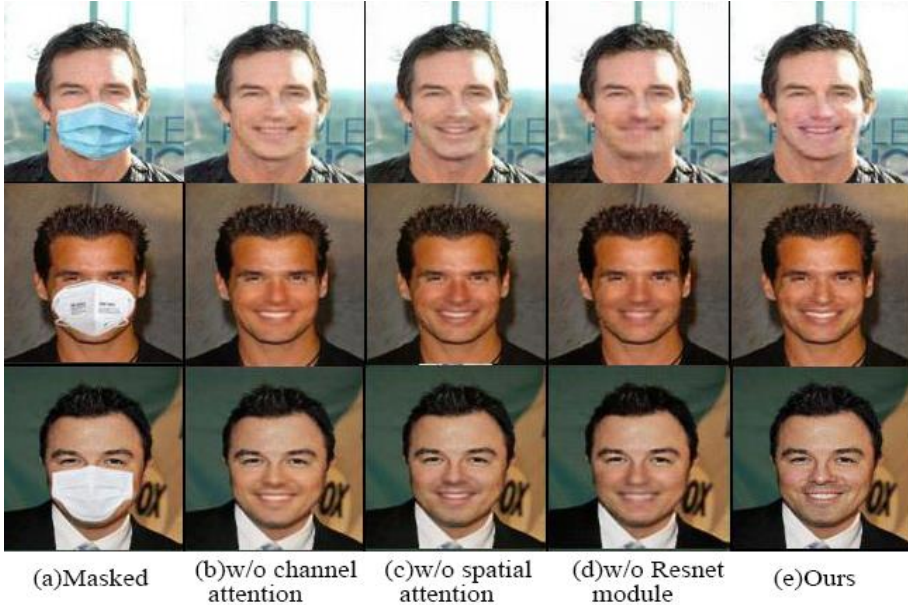
Fig. 16. Qualitative comparison of ablation experiment

Table 7 clearly shows the roles of the channel attention module, the spatial attention module and the improved residual module in the overall model. Among them, the improved residual module is particularly effective in extracting features and reconstructing pixels in the occluded region.

Tab. 7. Quantitative comparison of ablation experiment

	w/o Channel Attention	w/o Spatial Attention	w/o ResNet	Ours
PSNR	28.02	28.24	27.48	29.61
SSIM	91.51%	91.22%	88.94%	92.71%

5. DISCUSSION

In this paper, the authors propose an improved face inpainting model based on generative adversarial network, which can remove face masks. Face images are different from other general images in that face images have complex structures and faces contain rich semantic information. Therefore, face image inpainting is more difficult than general image inpainting. Face image inpainting has gained a lot of attention and some results have been achieved. However, the restored face images often have problems such as blurring and inconsistent boundaries. The uathors add improved residual modules to the generator network based on the UNet structure, and skillfully employ two attention mechanisms to improve the feature extraction of images for efficient mask removal. The experimental results show that the proposed inpainting method can not only achieve face image inpainting, but also be effective for general image inpainting. Therefore, if one person wears sunglasses

or a girl's flowing sea covers her forehead, it should also be possible to remove it. This provides convenience for the police to catch criminals, etc.

Considering that human faces have distinct geometric contours, especially eyes, ears, and nose have symmetric properties, we plan to add such a priori knowledge to the face inpainting model to further enhance the inpainting of face images. In addition, since different faces have similarities, we can do more scientific planning for the training dataset. For example, the dataset should contain images of various skin tones, various age groups, and various expressions. A more complete dataset may result in a better-trained inpainting model.

6. CONCLUSIONS

To effectively restore the blocked area by face mask and solve the problems of blurring and artifacts in the restored face images, this paper proposes an improved face image inpainting model based on the pix2pix generative adversarial network. Compared with the original method, authors' model introduces the attention mechanisms, improves the residual module, and uses skip-connections to connect the corresponding down-sampling module and up-sampling module. The experimental results show that the algorithm can realize face mask occlusion inpainting, rectangle occlusion inpainting and random occlusion inpainting. Compared with pix2pix and SRGan algorithms, the inpainting results of the improved model are significantly improved. In addition, good inpainting results have been obtained on Paris Street and Place2 datasets, which proves that the inpainting model has good portability.

REFERENCES

Ding, D., Ram, S., & Rodriguez, J. J. (2019). Image Inpainting Using Nonlocal Texture Matching and Nonlinear Filtering. *IEEE Transactions on Image Processing*, *28*(4), 1705–1719. https://doi.org/10.1109/TIP.2018.2880681

Goodfellow, I. (2014). *NIPS 2014 Tutorial: Generative Adversarial Networks*. arXiv. https://doi.org/10.48550/arXiv.1701.00160

He, L., Qiang, Z., Shao, X., Lin, H., Wang, M., & Dai, F. (2022). Research on High-Resolution Face Image Inpainting Method Based on StyleGAN. *Electronics*, *11*(10), 1620. https://doi.org/10.3390/electronics11101620

Hore, A., & Ziou, D. (2010). Image Quality Metrics: PSNR vs. SSIM. *2010 20th International Conference on Pattern Recognition* (pp. 2366–2369). IEEE. https://doi.org/10.1109/ICPR.2010.579

Iizuka, S., Simo-Serra, E., & Ishikawa, H. (2017). Globally and locally consistent image completion. *ACM Transactions on Graphics*, *36*(4), 1–14. https://doi.org/10.1145/3072959.3073659

Isola, P., Zhu, J.-Y., Zhou, T., & Efros, A. A. (2017). Imavolge-to-Image Translation with Conditional Adversarial Networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 5967–5976). IEEE. https://doi.org/10.1109/CVPR.2017.632

Jaderberg, M., Simonyan, K., & Zisserman, A. (2016). *Spatial Transformer Networks*. arXiv. https://doi.org/10.48550/arXiv.1506.02025

Jiang, Y., Yang, F., Bian, Z., Lu, C., & Xia, S. (2022). Mask removal: Face inpainting via attributes. *Multimedia Tools and Applications*, *81*(21), 29785–29797. https://doi.org/10.1007/s11042-022-12912-1

Jia, J., & Tang, Ch.-K. (2003). Image repairing: Robust image synthesis by adaptive ND tensor voting. *2003 IEEE Computer Society Conference on Computer Vision springeand Pattern Recognition, 2003. Proceedings.* (pp. I-I). IEEE. https://doi.org/10.1109/CVPR.2003.1211414

Ledig, C., Theis, L., Huszar, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., & Shi, W. (2017). Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 105–114). IEEE. https://doi.org/10.1109/CVPR.2017.19

Liu, G., Reda, F. A., Shih, K. J., Wang, T.-C., Tao, A., & Catanzaro, B. (2018). Image Inpainting for Irregular Holes Using Partial Convolutions. In V. Ferrari, M. Hebert, C. Sminchisescu, & Y. Weiss (Eds.), *Computer Vision – ECCV 2018* (vol. 11215, pp. 89–105). Springer. https://doi.org/10.1007/978-3-030-01252-6_6

Mou, C., Wang, Q., & Zhang, J. (2022). Deep Generalized Unfolding Networks for Image Restoration. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 17378–17389). IEEE. https://doi.org/10.1109/CVPR52688.2022.01688

Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T., & Efros, A. A. (2016). Context Encoders: Feature Learning by Inpainting. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 2536–2544). IEEE. https://doi.org/10.1109/CVPR.2016.278

Ronneberger, O., Fischer, P., & Brox, T. (2015). *U-Net: Convolutional Networks for Biomedical Image Segmentation*. arXiv. https://doi.org/10.48550/arXiv.1505.04597

Rudin, L. I., Osher, S., & Fatemi, E. (1992). Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, *60*(1–4), 259–268. https://doi.org/10.1016/0167-2789(92)90242-F.

Sagong, M., Shin, Y., Kim, S., Park, S., & Ko, S. (2019). PEPSI: Fast Image Inpainting With Parallel Decoding Network. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 11352–11360). IEEE. https://doi.org/10.1109/CVPR.2019.01162

Shao, X., Qiang, Z., Dai, F., He, L., & Lin, H. (2022). Face Image Completion Based on GAN Prior. *Electronics*, *11*(13), 1997. https://doi.org/10.3390/electronics11131997

Simakov, D., Caspi, Y., Shechtman, E., & Irani, M. (2008). Summarizing visual data using bidirectional similarity. *2008 IEEE Conference on Computer Vision and Pattern Recognition(CVPR)* (pp. 1–8). IEEE. https://doi.org/10.1109/CVPR.2008.4587842

Wang, Y., Tao, X., Qi, X., Shen, X., & Jia, J. (2018). *Image Inpainting via Generative Multi-column Convolutional Neural Networks*. arXiv. https://doi.org/10.48550/arXiv.1810.08771

Wu, H., Zhou, J., & Li, Y. (2020). *Deep Generative Model for Image Inpainting with Local Binary Pattern Learning and Spatial Attention*. arXiv. https://doi.org/10.48550/arXiv.2009.01031

Xie, C., Liu, S., Li, C., Cheng, M.-M., Zuo, W., Liu, X., Wen, S., & Ding, E. (2019). Image Inpainting With Learnable Bidirectional Attention Maps. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)* (pp. 8857–8866). IEEE. https://doi.org/10.1109/ICCV.2019.00895

Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X., & Huang, T. S. (2019). *Free-Form Image Inpainting With Gated Convolution*. arXiv. https://doi.org/10.48550/arXiv.1806.03589

Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X., & Huang, T. S. (2018). *Generative Image Inpainting With Contextual Attention*. arXiv. https://doi.org/10.48550/arXiv.1801.07892

Zeng, Y., Fu, J., Chao, H.. & Guo, B. (2019). Learning Pyramid-Context Encoder Network for High-Quality Image Inpainting. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp.1486–1494). IEEE. https://doi.org/10.1109/CVPR.2019.00158

Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., & Fu, Y. (2018). Image Super-Resolution Using Very Deep Residual Channel Attention Networks. In V. Ferrari, M. Hebert, C. Sminchisescu, & Y. Weiss (Eds.), *Computer Vision – ECCV 2018* (vol. 11211, pp. 294–310). Springer. https://doi.org/10.1007/978-3-030-01234-2_18

Zheng, C., Cham, T.-J., & Cai, J. (2019). Pluralistic Image Completion. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 1438–1447). IEEE. https://doi.org/10.1109/CVPR.2019.00153

Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., & Liang, J. (2018). *UNet++: A Nested U-Net Architecture for Medical Image Segmentation*. arXiv. http://arxiv.org/abs/1807.10165