

Research Article

3D Face Image Inpainting with Generative Adversarial Nets

Tongxin Wei ^{1,2}, Qingbao Li ², Jinjin Liu ², Ping Zhang ², and Zhifeng Chen ²

¹Software Institute, Zhengzhou University, Zhengzhou 450001, Henan, China

²State Key Laboratory of Mathematical Engineering and Advanced Computing, Zhengzhou 450001, Henan, China

Correspondence should be addressed to Tongxin Wei; tongxin_wei@foxmail.com

Received 4 September 2020; Revised 22 November 2020; Accepted 26 November 2020; Published 14 December 2020

Academic Editor: Gonglin Yuan

Copyright © 2020 Tongxin Wei et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In the process of face recognition, face acquisition data is seriously distorted. Many face images collected are blurred or even missing. Faced with so many problems, the traditional image inpainting was based on structure, while the current popular image inpainting method is based on deep convolutional neural network and generative adversarial nets. In this paper, we propose a 3D face image inpainting method based on generative adversarial nets. We identify two parallels of the vector to locate the planer positions. Compared with the previous, the edge information of the missing image is detected, and the edge fuzzy inpainting can achieve better visual match effect. We make the face recognition performance dramatically boost.

1. Introduction

To solve various problems encountered in the face collection, the following main traditional image restoration methods are used: structure-based face image inpainting [1] and face image restoration based on image structure and texture [2]. Most of these traditional image restoration methods are incomplete or even distorted for face restoration under different postures.

In deep learning, there are many different deep network structures, including towards real-time object detection with region proposal networks (R-CNN) [3], long-and short-term memory networks [4], and generative adversarial nets (GAN) [5]. In the field of computer vision, remarkable achievements have been made in the research and application of convolutional neural network (CNN) [6]. Inspired by the neural mechanism of animal vision, CNN network has excellent performance in machine vision and other fields [7, 8].

Image inpainting, based on convolutional neural networks [9–11], has an obvious advantage and is more logical, and inpainting effect is superior to the traditional inpainting method [12]. At the same time, generative adversarial net performs well in feature extraction and in image inpainting [13, 14] through the interaction between the generator and

discriminator, to repair the missing image. However, the high complexity of the network structure makes it difficult to modify the parameters. It wastes a lot of time in adjusting the weight parameters.

In this paper, we propose a stereo face image inpainting method based on facial feature structure. We use line segment to mark different planes to repair the structure and texture features of the image. Combined with the VGG network [15], we add dropouts and a complete connection layer. Unlike previous global feedback tweaks, we recommend adding local fix-part generation and counter to fix images more quickly.

In the feedback between generator and discriminator, parameters of the generator model are adjusted continuously, and the generated image is modified until both converge. The face image inpainting process is shown in Figure 1.

The contribution of the proposed method is threefold. First of all, combined with the inpainting method of 3D space structure, we make the structure of the image more clear. Secondly, the improved VGG network structure is adopted to remove the noise and make the image feature more obvious. Thirdly, we analyze the known structure to predict missing parts of the image and use generative adversarial nets to repeatedly repair from structure to

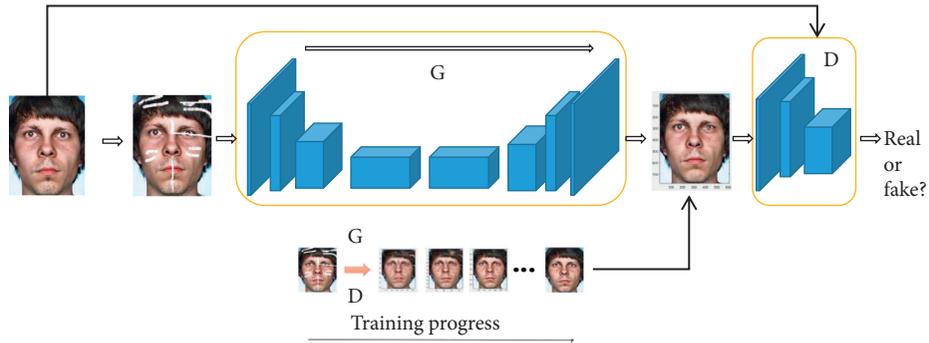


FIGURE 1: Our training starts with both the generator G and discriminator D having a low spatial resolution of $m \times n$ pixels. As the training advances, we incrementally add layers to G and D thus increasing the spatial resolution of the generated images. All existing layers remain trainable throughout the process. Here, $m \times n$ refers to convolutional layers operating on $M \times N$ spatial resolution [16]. This allows stable synthesis in high resolution and speeds up the training considerably.

texture. It has been proved that our method performs well in image restoration.

2. Related Work

Our recent work is to use the feature extraction function of convolutional neural network to extract face image features effectively. Generative adversarial net is used to complete the missing image. The missing image edge is fused to achieve the perfect restoration of face image.

2.1. Image Restoration. Image restoration methods can be roughly divided into traditional image restoration methods and image restoration algorithms based on neural networks. Representative methods include contextual encoders [17] and restoration methods based on image patching [9, 18–20]. However, if the edge pixel information is not processed smoothly, the image will have a cliff-like stitching effect. We should not only refer to the content of inpainting but also combine the image structure and texture information.

Another model [21] is based on neural network, which maps image feature extraction to high-dimensional space, and then integrates the extracted feature information by piecing together.

2.2. Face Inpainting. Most of the traditional face reinforcement algorithms [2, 22, 23] are based on the relative position structure of face, so it is difficult to complete the face under different rotation or tilt positions. Because of the limitations of planar face structure inpainting and the development of artificial intelligence, Xiong et al., Zhi and Sun, Nazeri et al., and Liao and Yan [24–27] proposed the face image inpainting method based on convolutional neural network and Liu et al., Li et al., Liao et al., and Portenier et al. [28–31] proposed the face image inpainting method based on generative adversarial nets. Meanwhile, the emergence of 3D face image inpainting [32] based on deep learning made face image inpainting achieve better effect. To some extent, this method is superior to the traditional face image

inpainting method, but the effect of the edge and line structure inpainting of the face image is unclear.

There are many sets of parallel lines according to a plane. We identify the support for each plane by locating the overlapping positions of two sets of line segments corresponding to two parallels of the vector. The structure of the face is generated according to different plane directions to generate the relevant line segment representation. Our method makes the image edge structure clearer and the inpainting effect better.

3. Model

VGG [15] has the following advantage compared with other neural networks: convolution operation makes full use of the spatial 3D structure of images. The common convolution kernel makes the operation easier and expands the receptive field. The increase of the number of convolutional layers makes the feature extraction of structure deepening more delicate. The network structure is deepened to make the detail processing of the image more accurate. Under the interaction of the generator and discriminator, the facial structural features are extended and adjusted.

3.1. Image Generation. The neural network extracts the features of the known face images and retains the structure information and texture information. Firstly, the whole contour structure of the image is extracted. Then, the texture structures in different planes are annotated with vectors. Finally, the texture structure information and structure characteristics of different planes are integrated.

3.1.1. Face Feature Extraction. When the surveillance video capture image is missing, it can be repaired in the following steps. Firstly, the missing part of the image is judged by the structure. The facial features are extracted according to the convolutional neural network. Finally, according to the local and overall structural features of the missing points, we fill the structure and texture of the missing image.

We first describe the large frame structure to roughly calculate the facial structural features of faces. Then, the

small texture information is filled in by a fine calculation. We use the median filter to remove the noise information, and the texture information of the missing image is repaired accurately. The face texture structure is shown in Figure 2.

3.1.2. 3D Representation of Image. Instead of the original square pixel repair, we propose a 3D face image inpainting method based on the facial feature structure. We identify the support for each plane by locating the overlapping positions of two sets of line segments corresponding to two parallels of the vector, according to different parallel lines to determine different planes so that the face stereo features more obvious. Fixing the structure and texture of different planes is more targeted. We distinguish the different plane structures by two sets of parallel lines on different planes. We use straight line segment fitting-edge curve to make face contour clearer.

Figure 3 shows our method about the facial three-dimensional image structure. Two groups of different parallel lines are used to distinguish different planar structures of faces. We use line segment fitting curve to smooth the structure of facial lines.

3.2. Discriminator. In the process of generative adversarial nets, it is necessary to import database data into the model for training. The model performs feature extraction on the structural information of the relevant face data. By training the model, the inpainting speed and accuracy of missing face image are further improved. The following is the image completion algorithm, which uses the cross-entropy loss function, the convergence conditions of the final generator, and discriminator parameters.

The missing image is used as the test data input model. We calculate the missing part of the image of the probability distribution. We initialize the processed reconstructed image. Each channel is set independently in the original image and the target image discrete Poisson equation. We combine with the known target area to constantly adjust to the new information in the known image. The unknown pixel position of hybrid Poisson for linear equations is calculated in advance. In the image recovery process, the missing part of unknown pixel position near the boundary is calculated in advance to construct the sparse matrix.

At the same time, the discriminator uses the cross-entropy loss to compare and judge the images. The cross-entropy loss function is as follows:

$$H(p, q) := \sum_i p_i \log q_i. \quad (1)$$

Tips: in the formula, p_i is real sample distribution and q_i is generator generation distribution.

In the case of the current model, the discriminator is a binary classification problem, so the basic cross-entropy can be expanded more specifically.

After extending to N samples, add N samples to obtain the corresponding formula as follows:

$$H((x_i, y_i)_{i=1}^N, D) = - \sum_{i=1}^N y_i \log D(x_i) - \sum_{i=1}^N (1 - y_i) \log(1 - D(x_i)). \quad (2)$$

For the sample points x_i in GAN, they correspond to two sources, which are either real samples or samples generated by the generator that are subject to the noise distribution in the generator.

Where, for a sample that comes from a real sample, we want to check for the correct distribution y_i . We want to make sure it has an error distribution in the generated sample. The above formula is further written in the expected form of probability distribution, and the formula shown below can be obtained for GAN loss function:

$$H((x_i, y_i)_{i=1}^{\infty}, D) = -\frac{1}{2} E_{x \sim P_{\text{data}}} [\log D(x)] - \frac{1}{2} E_z [\log(1 - D(G(z)))]. \quad (3)$$

4. Experiments

We report on the qualitative and quantitative evaluation of the experiment in detail. It includes original image data, damaged image, inpainting image, effect comparison of various experimental methods, and experimental data analysis.

4.1. Datasets. In our experiment with the evaluation method, we use publicly available dataset called CelebA [33] and AFLW [34]. CelebA contains 10,177 images of 202,599 faces corresponding to celebrity status. Each image label identifies the label and features, including the Face Bbox annotation box, the coordinates of the human face feature points, and the face attribute label. CelebA is widely used for face-related recognition and image inpainting tasks. AFLW is a large-scale face database with multipose and multiview. Each face is marked with 21 feature points, which included images of different poses, expressions, lighting, and race. It has a high research value. We use Set14, BSD100, URBAN100 [35], and ManGA100 [36] for evaluating model inpainting images' effect.

4.2. Network Structure of the Model. The convolutional layer and ReLU activation function are added to the convolutional network to extract features through the convolutional layer. The ReLU function is activated and passed into the pooling layer. After feature classification, the full connection layer is used for feature stitching. Figure 4 shows the network structure frame of neural network to extract face image features. It consists of six convolutional layers and pooling layers to extract face image features. Finally, features are extracted and stored through the cross combination of three fully connection layers and two dropout layers. At the same time, when the incomplete face image is input, the incomplete part is analyzed globally and locally to complete the missing image. Figure 4 shows the network structure of face image feature extraction.

4.3. Experimental Details and Results. We create image structures from coarse to fine on both linear and logarithmic scales. Image structure patch: the image size of each layer of

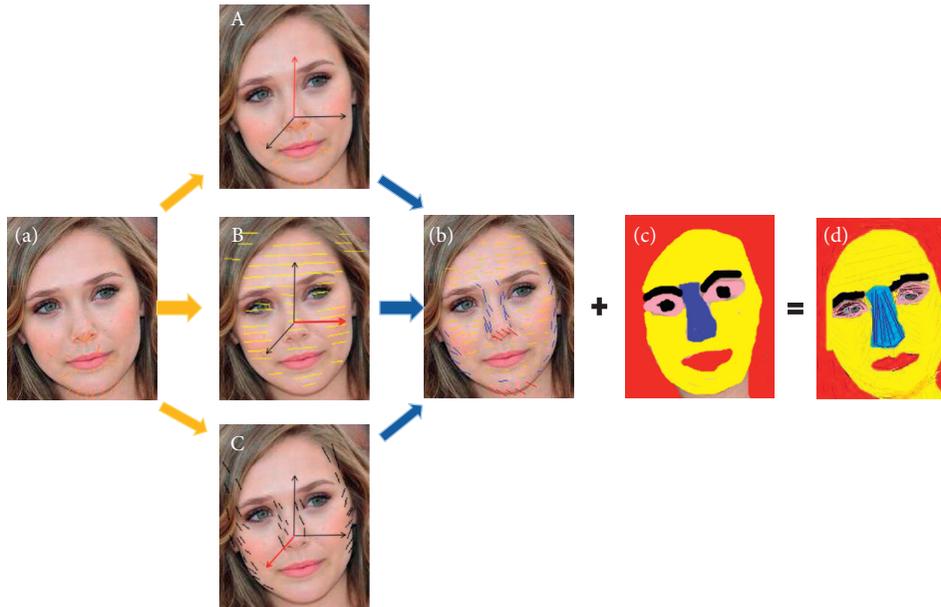


FIGURE 2: The spatial structure of different plane structures and textures. Picture A is the original image, while a, b, and c represent the three-dimensional hierarchical effects of different planes of the image with line segment texture. At the same time, image C and texture feature B interact with each other to form the representation result of image D through feature extraction result of face structure.

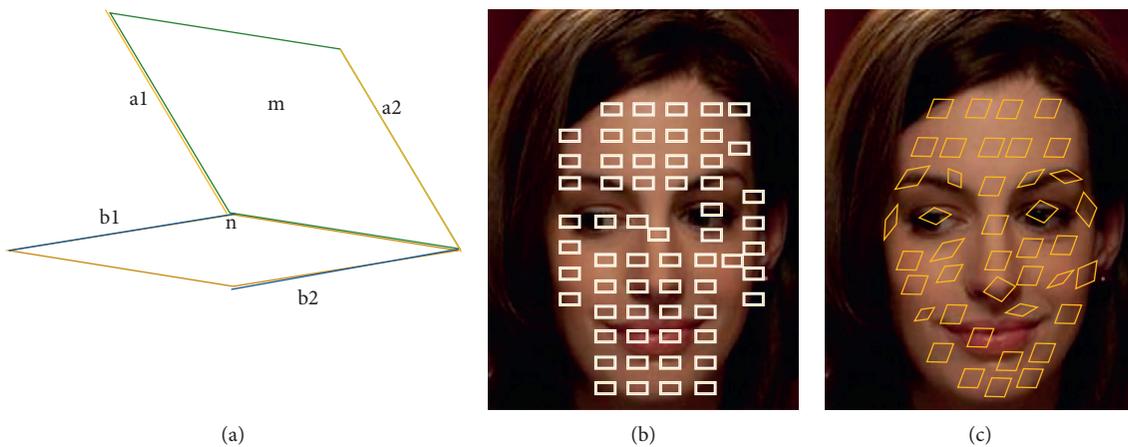


FIGURE 3: (a) Parallel to a_1 and a_2 , a_1 and a_2 determine plane m ; parallel to b_1 and b_2 , b_1 and b_2 determine plane n , m , and n different from two different planes to determine a three-dimensional space. (b) Traditional pixel structure arranged. (c) Our difference between different planes.

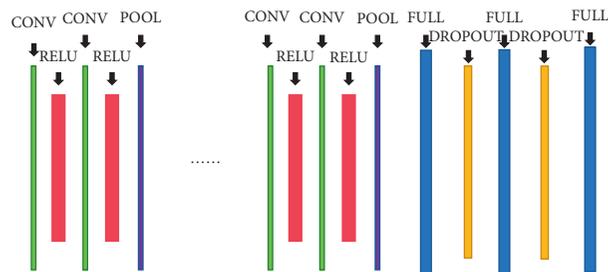


FIGURE 4: Feature extraction structure.



FIGURE 5: Inpainting results of AFLW.



FIGURE 6: The image generation results in Set14, ManGA100, BSD100, and URBAN100.

the neural network structure and the image size of the high-resolution image. To avoid the dark value near the image boundary, we transform the resampling mask into a logical type by fuzzy inversion resampling. We calculate the coarsest scale of the image.

VGG network extracts the dataset and tag data according to the characteristics of the dataset. At the same time, we use the image outside the database as the test set to test and verify the performance of the model.

We divide the images into training data and test data. Based on the lack of test data, the missing image generator is generated on our network to repair the missing image content. At the same time, we use the feedback effect of the discriminator and generator to repair the structure and texture of the image and adjust the local optimal effect.

The plane model and the regular model structure are calculated in advance according to the image. We scale the

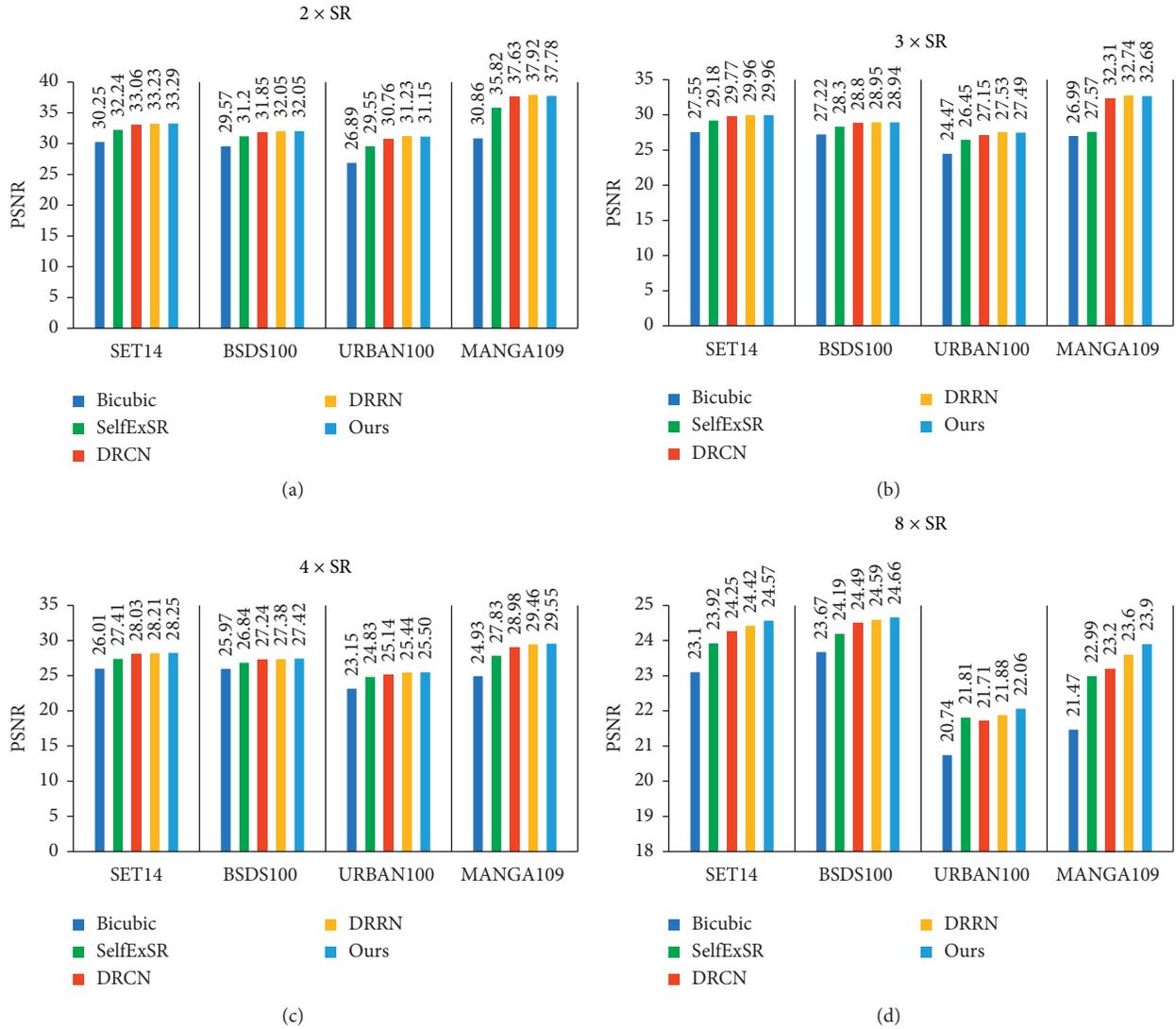


FIGURE 7: We report the average PSNR for [37] (a) 2x SR, (b) 3x SR, (c) 4x SR, and (d) 8x SR.

parameters according to the image size. After updating the plane model, we construct the plane probabilistic updating correction matrix and rotation parameters.

We train the data in AFLW database and then do image inpainting after irregular occlusion processing. The facial inpainting effect is shown in Figure 5.

4.4. Image Regional Ability Evaluation. In face image restoration, we find that the resolution of newly generated images is significantly improved. We do the following tests to verify the model's effect in the super-resolution direction. We repair some damaged or unclear images, and we need to measure the result through image quality evaluation index in addition to direct visual perception. Common indicators are the peak signal-to-noise ratio (PSNR) and the structural similarity index (SSIM). In this paper, we use PSNR and

SSIM as performance metrics to evaluate our model. Figure 6 shows the newly generated image in Set14, BSD100, URBAN100 [35], and ManGA100 [36].

We report the average PSNR/SSIM for 2x, 3x, 4x, and 8x SR. As shown in Figures 7 and 8, the experimental results of this paper are used to compare data [37]. The quantitative results are shown in Figure 8. Our method is superior to [38, 39], especially in the case of 8x SR. We show visual comparisons of SET, BSDS100, URBAN100, and MANGA109 datasets to generate new images. Our method makes the original image structure and texture clearer. Instead, our method reconstructs the order gradually from coarse-grained to fine-grained calculations. In contrast, our model shows a faster rate of image reconstruction under the conditions of 2x SR and 3x SR. However, compared with DRRN [40], in the 2x SR model, the resolution

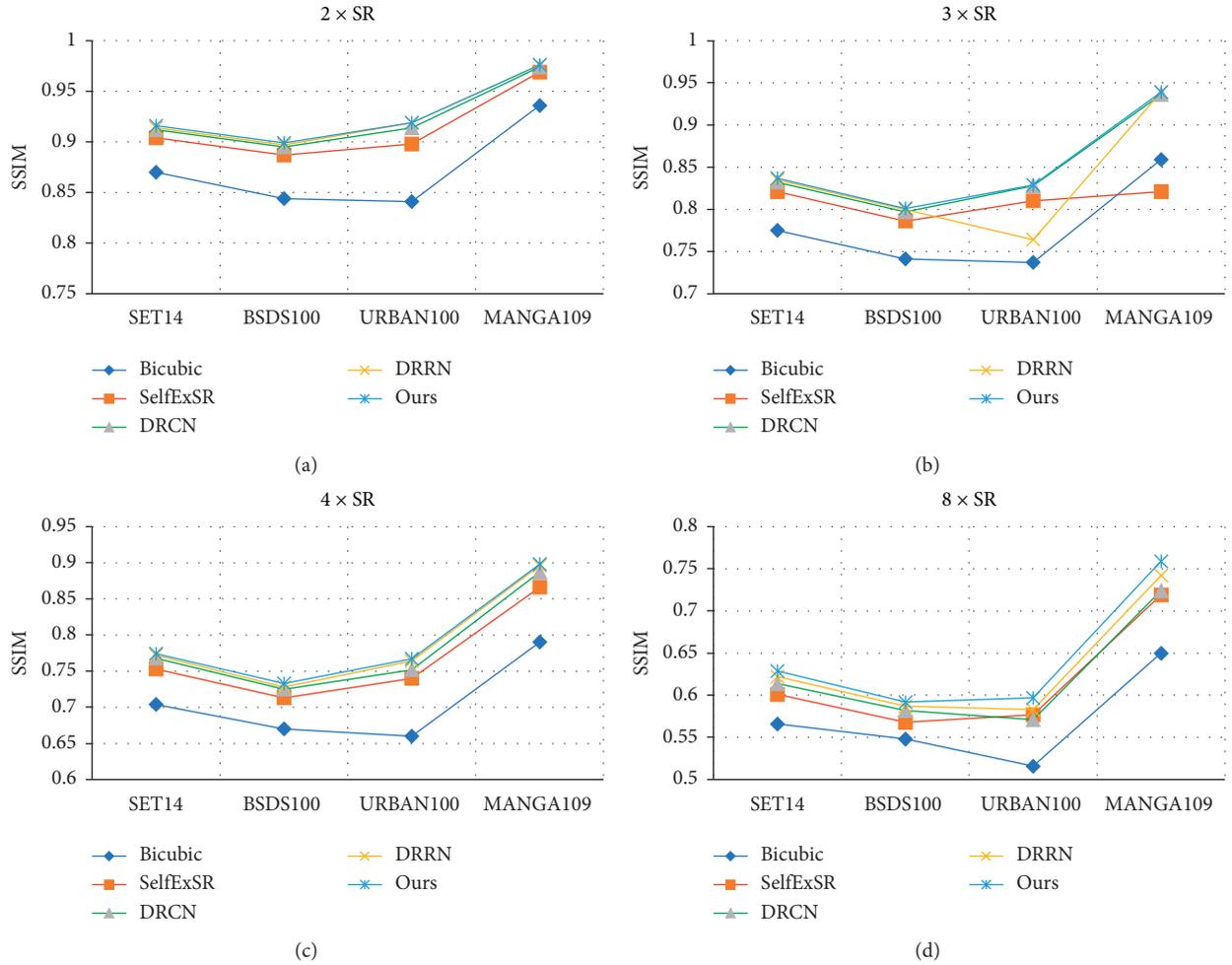


FIGURE 8: We report the average SSIM for [37] (a) $2 \times$ SR, (b) $3 \times$ SR, (c) $4 \times$ SR, and (d) $8 \times$ SR.

improvement effect is not as good as DRRN due to its complexity of network structure. This does not affect the repair effect of our model on missing face image.

4.5. Performance Improvement on Face Recognition. We adopt the inpainting method to repair the damaged image. We compare the inpainting image with the original identity information to judge the performance of the feature domain.

In face recognition, Euclidean distance and cosine distance are usually used to measure the similarity of face features. In this paper, we use cosine distance to measure the difference between two individuals.

We verify the improvement of the masking face recognition performance. We take some of the character images from CelebA as training data. We partially occlude the face data. We calculate the similarity between the missing data

and the original data and the similarity between the repaired data and the original data. The face inpainting can be verified by face identification. First, we use the original dataset as the training set. Then, we tested the occlusion dataset and the recovery dataset, respectively. The experimental results show that occlusion restoration is helpful to improve the recognition performance.

Figure 9 shows the original image selected from the CelebA dataset, the damaged image, and the face image repaired in our method. First, we detect the similarity between the missing image and the original image to obtain the value of similarity. Then, we detect the similarity between the restored image and the original image to obtain the similarity value of the restored image. Through comparison, it is found that the repaired face image has obvious advantages over the result of face recognition. This proves that our method has great significance for face image inpainting and face recognition.

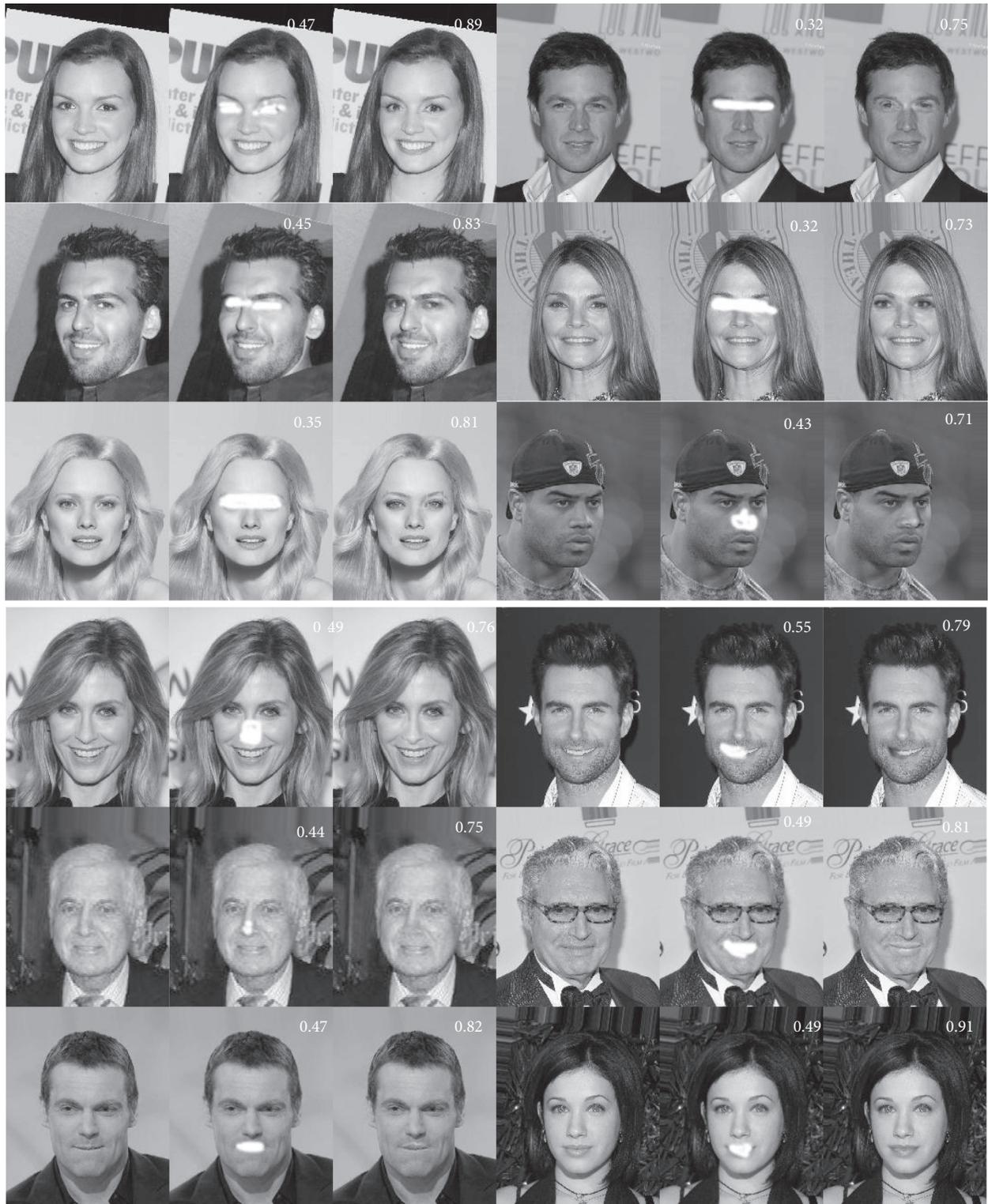


FIGURE 9: The results of CelebA datasets.

5. Conclusions

In this paper, we propose a 3D face image restoration method based on generative adversarial nets. The results of the experiment indicate the algorithm is effective and competitive for face image inpainting. The repair effect for a small part of the missing image is satisfactory. However, when there are too many missing parts in the image, the repair effect will be different from the original image. In the future, we are committed to simplifying the model structure and improving the model generalization ability.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

There are no potential conflicts of interest.

References

- [1] Z. Niu, H. Li, Y. Li, Y. Mei, and J. Yang, "An adaptive face image inpainting algorithm based on feature symmetry," *Symmetry*, vol. 12, no. 2, p. 190, 2020.
- [2] J. Sulam and M. Elad, "Large inpainting of face images with trainlets," *IEEE Signal Processing Letters*, vol. 23, no. 12, pp. 1839–1843, 2016.
- [3] S. Ren, K. He, and B. G. Ross, "Faster R-CNN: towards real-time object detection with region proposal networks," *Advances in Neural Information Processing Systems*, vol. 39, no. 6, pp. 91–99, 2015.
- [4] S.-S. Ho, P. Dai, and F. Rudzicz, "Manifold learning for multivariate variable-length sequences with an application to similarity search," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 6, pp. 1333–1344, 2016.
- [5] M. Nakao, K. Imanishi, N. Ueda, Y. Imai, T. Kirita, and T. Matsuda, "Regularized three-dimensional generative adversarial nets for unsupervised metal artifact reduction in head and neck CT images," *IEEE Access*, vol. 8, pp. 109453–109465, 2020.
- [6] A. Krizhevsky, I. Sutskever, and E. H. Geoffrey, "ImageNet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 60, no. 6, pp. 1097–1105, 2012.
- [7] J. Liu, Q. Li, P. Zhang, G. Zhang, and M. Liu, "Unpaired domain transfer for data augment in face recognition," *IEEE Access*, vol. 8, pp. 39349–39360, 2020.
- [8] M. Liu, J. Liu, P. Zhang, and Q. Li, "PA-GAN: a patch-attention based aggregation network for face recognition in surveillance," *IEEE Access*, vol. 8, pp. 152780–152789, 2020.
- [9] M. Ghorai, S. Samanta, S. Mandal, and B. Chanda, "Multiple pyramids based image inpainting using local patch statistics and steering kernel feature," *IEEE Transactions on Image Processing*, vol. 28, no. 11, pp. 5495–5509, 2019.
- [10] G. Liu, F. A. Reda, and K. J. Shih, "Image inpainting for irregular holes using partial convolutions," *Proceedings of the European Conference on Computer Vision*, vol. 11, pp. 85–100, 2018.
- [11] C. Yu, F. Li, and C.-I. Chang, "Deep 2D convolutional neural network with deconvolution layer for hyperspectral image classification," *International Conference in Communications, Signal Processing, and Systems*, vol. 516, no. 2, pp. 149–156, 2018.
- [12] W. Zhang, "A precise-mask-based method for enhanced image inpainting," *Mathematical Problems in Engineering*, vol. 2016, Article ID 6104196, 2016.
- [13] U. Demir and B. Ü. Gözde, "Patch-based image inpainting with generative adversarial networks," 2018, <https://arxiv.org/abs/1803.07422v1>.
- [14] D. Hazra and Y.-C. Byun, "OEER-GAN: object extraction and background recovery generative adversarial networks," *IEEE Access*, vol. 8, pp. 135730–135741, 2020.
- [15] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2015, <https://arxiv.org/abs/1409.1556v6>.
- [16] T. Karras, T. Aila, and S. Laine, "Progressive growing of GANs for improved quality, stability, and variation," 2018, <https://arxiv.org/abs/1710.10196v3>.
- [17] A. Marafioti, N. Perraudin, N. Holighaus, and P. Majdak, "A context encoder for audio inpainting," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 12, pp. 2362–2372, 2019.
- [18] Z. Akhtar and F. Gian Luca, "Face spoof attack recognition using discriminative image patches," *Journal of Electrical and Computer Engineering*, vol. 2016, Article ID 4721849, 2016.
- [19] J. Zbontar and Y. LeCun, "Stereo matching by training a convolutional neural network to compare image patches," *The Journal of Machine Learning Research*, vol. 17, no. 1, pp. 2287–2318, 2016.
- [20] L. He, "An adaptive image inpainting method based on continued fractions interpolation," *Discrete Dynamics in Nature and Society*, vol. 2018, Article ID 9801361, 2018.
- [21] G. Zhao, J. Liu, J. Jiang, and W. Wang, "A deep cascade of neural networks for image inpainting, deblurring and denoising," *Multimedia Tools and Applications*, vol. 77, no. 22, pp. 29589–29604, 2018.
- [22] M. Jampour, C. Li, L.-F. Yu, K. Zhou, S. Lin, and H. Bischof, "Face inpainting based on high-level facial attributes," *Computer Vision and Image Understanding*, vol. 161, pp. 29–41, 2017.
- [23] H. Qiang, C. Dong, and F. Zhang, "A novel approach to face verification based on second-order face-pair representation," *Complexity*, vol. 2018, Article ID 2861695, 2018.
- [24] H. Xiong, C. Wang, X. Wang, and D. Tao, "Deep representation calibrated bayesian neural network for semantically explainable face inpainting and editing," *IEEE Access*, vol. 8, pp. 13457–13466, 2020.
- [25] Z. Zhi and Yi Sun, "An image inpainting algorithm based on energy minimization," *International Journal of Image and Graphics*, vol. 12, no. 3, 2012.
- [26] K. Nazeri, E. Ng, and T. Joseph, "EdgeConnect: generative image inpainting with adversarial edge learning," 2019, <https://arxiv.org/abs/1901.00212v3>.
- [27] H. Liao and Y. Yan, "Age estimation of face images based on CNN and divide-and-rule strategy," *Mathematical Problems in Engineering*, vol. 2018, Article ID 1712686, 2018.
- [28] X. Liu, Z. Yao, and C. Xie, "Bidirectional face aging synthesis based on improved deep convolutional generative adversarial networks," *Information*, vol. 10, no. 2, 2019.
- [29] Z. Li, H. Zhu, L. Cao, L. Jiao, Y. Zhong, and A. Ma, "Face inpainting via nested generative adversarial networks," *IEEE Access*, vol. 7, pp. 155462–155471, 2019.
- [30] H. Liao, G. Funka-Lea, and Y. Zheng, "Face completion with semantic knowledge and collaborative adversarial learning,"

- Asian Conference on Computer Vision*, vol. 1, pp. 382–397, 2018.
- [31] T. Portenier, Q. Hu, and A. Szabó, “Faceshop: deep sketch-based face image editing,” 2018, <https://arxiv.org/abs/1804.08972v2>.
 - [32] J. Roth, Y. Tong, and X. Liu, “Adaptive 3D face reconstruction from unconstrained photo collections,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, vol. 39, no. 11, pp. 4197–4206, 2017.
 - [33] <http://mmlab.ie.cuhk.edu.hk/projects/CelebA.html>.
 - [34] <https://www.tugraz.at/institute/icg/research/team-bischof/lrs/downloads/aflw/>.
 - [35] <https://github.com/jbhuang0604/SelfExSR/tree/master/data>.
 - [36] <https://dl.acm.org/doi/10.1145/3011549.3011551>.
 - [37] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, “Fast and accurate image super-resolution with deep laplacian pyramid networks,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 11, pp. 2599–2613, 2019.
 - [38] J.-B. Huang, A. Singh, and N. Ahuja, “Single image super-resolution from transformed self-exemplars,” in *Proceedings of the IEEE Conference on Computer Vision And Pattern Recognition*, pp. 5197–5206, IEEE, Boston, MA, USA, June 2015.
 - [39] J. Kim, L. Jung Kwon, and K. Mu Lee, “Deeply-recursive convolutional network for image super-resolution,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, pp. 1637–1645, June 2016.
 - [40] Y. Tai, J. Yang, and X. Liu, “Image super-resolution via deep recursive residual network,” *Proceedings of the IEEE Conference on Computer Vision And Pattern Recognition*, IEEE, pp. 2790–2798, July 2017.