Hindawi
Mathematical Problems in Engineering
Volume 2020, Article ID 4589260, 8 pages
https://doi.org/10.1155/2020/4589260

Research Article

Facial Landmark Detection Using Generative Adversarial Network Combined with Autoencoder for Occlusion

Hongzhe Liu,[1] Weicheng Zheng,[1] Cheng Xu,[1] Teng Liu,[1] and Min Zuo[2]

[1]*Beijing Key Laboratory of Information Service Engineering, College of Robotics, Beijing Union University, Beijing, China*
[2]*National Engineering Laboratory for Agri-Product Quality Traceability, Beijing Technology and Business University, Beijing, China*

Correspondence should be addressed to Cheng Xu; xc-f4@163.com

Received 13 June 2020; Revised 24 October 2020; Accepted 9 November 2020; Published 26 November 2020

Academic Editor: Mariko Nakano-Miyatake

The performance of the facial landmark detection model will be in trouble when it is under occlusion condition. In this paper, we present an effective framework with the objective of addressing the occlusion problem for facial landmark detection, which includes a generative adversarial network with improved autoencoders (GAN-IAs) and deep regression networks. In this model, GAN-IA can restore the occluded face region by utilizing skip concatenation among feature maps to keep more details. Meanwhile, self-attention mechanism that is effective in modeling long-range dependencies is employed to recover harmonious images for occluded faces. Deep regression networks are used to learn a nonlinear mapping from facial appearance to facial shape. Benefited from the mutual cooperation of GAN-IA and deep regression networks, a robust facial landmark detection model is achieved for the occlusion problem and the performance of the model achieves obviously improvement on challenging datasets.

1. Introduction

Facial landmark detection is essential to many facial analysis applications, e.g., face recognition, facial expression analysis, and 3D face modeling. Unfortunately, this important task still suffers from many challenges in the real world, such as occlusion, extreme pose, and illumination. Facial occlusion is a main cause of the failure of the facial landmark detection algorithms and could be caused by objects or self-occlusion due to large head poses. For decades, many methods which are devoted to exploring robust facial landmark detection under control condition or even in the wild perform well for near frontal and clear face images, while their performances usually degenerate severely under partial occlusions. How to model occlusion regions is the essential core of dealing with occlusion issue for facial landmark detection; it is difficult to be tackled because parts of the face can be occluded by irregular, complex, and arbitrary objects.

Recently, some related works have been proposed to deal with this issue. Burgos-Artizzu et al. [1] proposed a robust cascaded pose regression (RCPR) method dividing the face into different blocks, and for each time, only one non-occluded block is used to predict facial landmarks. It achieves impressive results on the challenging dataset, e.g., 300 W and AFLW. However, the training of the RCPR model depends on occlusion annotations, and it is very expensive to achieve annotating occlusion for large-scale datasets. Xing et al. [2] combined an occlusion dictionary with the face appearance dictionary to restore face shape under partial occlusions and modeling different partial face occlusions.

In recent years, convolutional neural networks (CNNs) and generative adversarial network (GAN) [3] have been achieved significant performance improvements for face occlusion problem. It is due to the fact that CNN has significant powerful ability to learn feature representation and that GAN can train a well performance generator via an adversarial process to generate a filled image as output that is a high-quality and natural image for occlusion regions. Zheng et al. [4] designed a probabilistically principled framework with a reconstructive path and a reconstructive path to rebuild the original image from this distribution

prior distribution of missing parts, which are supported by GANs. Lee et al. [5] proposed an effective facial landmark detection network and an associated learning framework with the geometric prior-generative adversarial network. The method achieves fine performances.

Occlusion sensitivity is a challenging issue for CNN as well [6]; it probably guides the detection model to learn bad feature representation when facial landmarks are being detected. Benefited from methods mentioned above, we propose a generative adversarial network with improved autoencoders (denoted as GAN-IA) to explicitly tackle the specific occlusion problem via cascaded deep regression networks for facial landmark detection. Our work builds upon the recently proposed DRDA [7], which utilizes an autoencoder-like neural network that is trained with only by minimizing the reconstruction error; it probably leads to blurry output. We substitute GAN-IA for the decorrupt autoencoder of DRDA to output more natural and real face images for partial occlusion regions to improve facial landmark localization accuracy. The work of GAN-IA depends on the recently proposed model [8].

As illustrated in Figure 1, the model GAN-IA recovering face part occluded is not a simple autoencoder-like neural network in DRDA; it consists of a generator, a local discriminator, and a global discriminator. The generator is trained to recover the occlusion region, and the two discriminators are used to distinguish the recovered contents in the occlusion and whole generated images as real and fake. To restore more details in occlusion regions, we introduce skip concatenation following "U-Net" [9] between feature maps during generating the occlusion-free image. Moreover, self-attention mechanism is introduced so that the generator can draw images in which fine details at every location being carefully coordinated with fine details in distant portions of the image, and discriminators can enforce complicated geometric constraints on global image structure [10]. The recovered images are fed into the deep regression network for landmark detection task. By combining GAN-IA the with deep regression network, a robust facial landmark detection model to partial occlusions is attained and detection accuracy is obviously improved as we shall see.

The major contributions of our work are summarized as follows:

(1) We propose an improved method based on DRDA to deal with the occlusion problem for facial landmark detection.

(2) We design a new framework based the existing model to restore the partially occluded region automatically, and the restored regions are employed together with nonoccluded parts, which results in better performance.

(3) Extensive experiments conducted on challenging datasets show that our proposed approach has a significant performance improvement compared with the existing methods for facial landmark detection task under occlusions.

2. Related Work

Early representative works on landmarks estimation include active contours models (known as snakes), active shape models (ASMs) [11], constrained local models (CLMs) [12], active appearance models (AAMs) [13], and Gauss–Newton deformable part models [14]. This category of algorithms belongs to template methods and employs principal component analysis (PCA) [15] to model the variation in face shape or simultaneously establish shape and appearance models. However, these methods suffer from poor performance when the reconstruction error spreads over the whole face under occlusions and each of these approaches hardly reach state-of-the-art performance on "in the wild" datasets.

Apart from methods mentioned above, there are regression-based methods which directly learn the mapping from face images to landmark coordinate vectors with linear or deep architecture. Many early methods use handcrafted features to extract facial texture information and leverage machine learning algorithms, e.g., SVM and random forest, as the regressors. Kazemi and Sullivan [16] presented a general framework based on gradient boosting for learning an ensemble of regression trees that optimizes the sum of square error loss and naturally handles missing or partially labelled data. Ren et al. [17] proposed a set of local binary feature to jointly learn a linear regression for the final landmarks output. These methods like this usually adopt cascade architecture to estimate and refine the shape iteratively until convergence. Unfortunately, these early works are suboptimal because of not considering the underlying relationship between feature extraction and regression process. Instead, recent methods combine feature extraction with regression process to train the model using the end-to-end way in CNN. DSRN [18] is a direct shape regression network for end-to-end face alignment by jointly handling the highly nonlinear relationship between face images and associated facial shapes in a unified framework. DeCaFA [19] is an end-to-end deep convolutional cascade architecture for face alignment; it uses fully-convolutional stages to keep full spatial resolution throughout the cascade and significantly outperforms existing approaches on challenging databases. Wang et al. [20, 21] put forward the idea of combining the face GAN network with the cascaded network to improve the face alignment algorithm, realize the accurate positioning of key points of the face, and solve the problem of facial expression lighting and occlusion for face detection.

There are also some works based on heatmap regression and have achieved promising results in facial landmark detection. Kowalski et al. [22] proposed the deep alignment network (DAN) to take landmark heatmaps and the entire face as the input of the midstage in multiple stage architecture, and landmark heatmaps can provide visual information about landmark coordinates. Most recently, LAB [23] proposed is a boundary-aware face alignment algorithm by utilizing boundary lines as the geometric structure of a human face to help facial landmark localization.

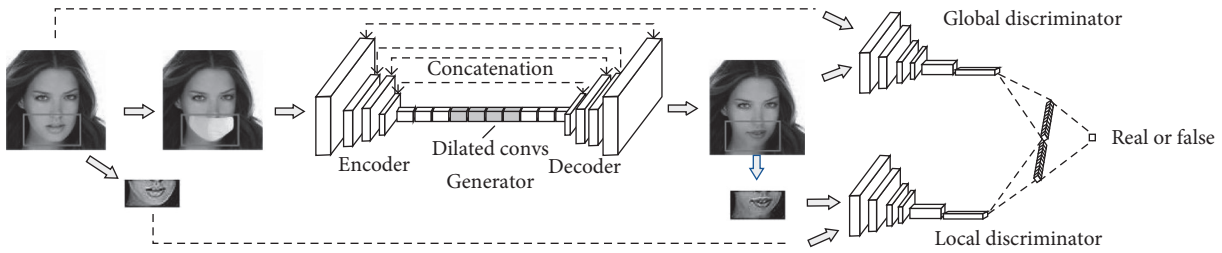

Figure 1: The architecture of generative adversarial network combined with improved autoencoders. Our proposed framework mainly consists of three modules: generator (autoencoder), global discriminator, and local discriminator. The generator takes the occluded images as input and outputs the generated images. The two discriminators are used only for training the generator cooperatively while they are not needed in the testing; they are learned to distinguish the generated contents in the occlusion and whole generated images as real and fake.

3. Our Approach

Our approach is based upon an existing method which employs a deep regression network coupled with autoencoders to tackle partial occlusion problem under a cascade structure for facial landmark detection. We replace its autoencoder with GAN-IA. A more robust and accuracy method, consisting of a deep regression network and GAN-IA, is proposed to explicitly deal with partial occlusion problem. In this section, the overview of the proposed method will be illustrated firstly. Then, the more details about the deep regression network for facial landmark localization and GAN-IA for restoring the occluded face region will be demonstrated separately.

3.1. A Method Overview. In this paper, we propose a robust and accuracy method including a generative adversarial network with improved autoencoders for facial landmark detection, which aims to effectively handle the occlusion problem. This work builds upon the proposed DRDA that is a cascade structure, and we modify its two major modules to yield more precise results: deep regression network which only has neural network layers lacking enough data fit ability and decorrupt autoencoders which is modeled as an autoencoder-like neural network having a weak ability of feature representation learning. To be specific, the deep regression network is replaced with a practical facial landmark detector, denoted as PFLD [24], which is an accurate, efficient, and compact facial landmark detector but is not an exclusive model for occlusion problem. In addition, we elaborately design a new network framework GAN-IA based on the convolution network as a substitute for original autoencoders to recover the occluded face region realistically. In our experiments, we also use the original regression network as a comparison to verify effectiveness of our method.

As illustrated in Figure 1, it diagrams the new framework of restoring the jaw and mouth part occluded by an occlusion. The specific regression process details can be seen in DRDA, and simple process can be seen in Figure 2.

Cascade regression iteration process can be characterized by the following equation:

$$S_{t+1} = S_t + \Phi_t(\phi(I, S_t)), \quad (1)$$

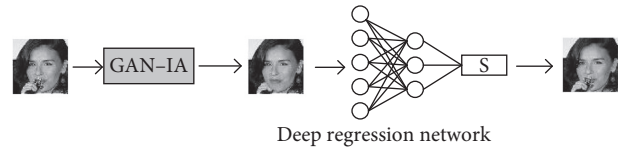

Figure 2: Simple process of landmark detection. First, GAN-IA restores the occlusion region for reducing the impact of occlusion, and then landmarks are detected by the deep regression network.

where S_t is the landmark locations estimated at iteration t and Φ_t is the regression network which returns the update to S_t given a feature ϕ from the image I at the landmark locations.

3.2. Deep Regression Network for Shape Prediction. The objective of the deep regression network is to characterize the nonlinear mapping from facial appearance to facial shape. For the face image, the convolutional neural network has better performance than the traditional deep neural network for facial landmark detection due to the fact that feature extraction process and regression process are trained simultaneously in CNN that can directly infer the underlying relationship between facial appearance and facial shape. As a result, we substitute the original deep neural network relying on hand-crafted features in DRDA with PFLD based on CNN recently proposed which has superior performance for facial landmark detection. Although PFLD is an efficient detector, we make a comparison between the result of models with or without it for demonstrating our progress.

3.3. GAN-IA Restoring Occlusion Region. The generator is designed as an improved autoencoder to learn complex distributions in an unsupervised setting and to reconstruct new contents of input images with occluded regions. It is based on a fully convolutional network. Unlike the traditional autoencoder model, the improved autoencoder in the generator introduces concatenation strategy in channel dimension to make up missing feature information during the encoder stage from low-level feature to high-level feature while it is difficult to restore in the decoder stage inversely. This work is built upon an efficient structure, the so-called "U-Net." We modify this architecture so that it works with very few training images and leads to better results. Without

only using the traditional standard convolutional layers, a variant called the dilated convolution layer [25] also is employed, which allows extending receptive field each pixel point of feature maps and not increasing the number of learnable weights, which is important for the recovering occlusion task. In the encoder stage, we employ standard convolution layers to reduce the size of the feature map while increasing feature map depth to get better performance. In the decoder stage, feature maps in the midlayers are restored to the original input resolution by utilizing deconvolution layers [26] including convolution with fractional strides, which is crucial to output nonblurry texture in the occlusion regions.

Although the generator can be trained to restore the occluded face region, it cannot ensure that whether the restored image is visually realistic or has been restored. To encourage more natural face, we adopt a local discriminator and a global discriminator to distinguish whether an image is real or has been restored. The local discriminator focuses on the restored contents in the occlusion region while the global discriminator determines the faithfulness of an entire image. Like the generator, this work also is based on an existing architecture, and we improve it by introducing self-attention mechanism which is complementary to convolutions and help with modeling long-range and multilevel dependencies cross image regions. By doing so, the proposed model can more accurately enforce complicated geometric constraints on the global image structure, and fine details of output at every location can be coordinated with ones in distant portions of output. Figure 3 shows the comparison of facial occluded image recovery with or without self-attention mechanism.

The self-attention structure is shown in Figure 4, where x is the original layer input, y is the new output, and the output is as (3). In equation (3), β is the calculated attention map, defined as follows: $W_h \in R^{c\times c}$, where $s_{i,j}$is the feature map.

$$y_i = \gamma O_i + x_i. \tag{2}$$

Note that pooling layers are not used in the generator and two discriminators because spatial information within a receptive field will be lost during pooling which may be critical for precise image restoring.

To ensure the generator can recover the occluded image realistically, we introduce two loss functions. One is the reconstruction loss L_r that is the L_2 distance between the network output and the original image. Another loss is the adversarial loss which is the crucial part of training in our work and involves turning the standard optimization of a neural network into a min-max optimization problem in which at each iteration the discriminator networks are together updated with the generator. It is defined as follows:

$$\begin{aligned} L_{d_i} = \min_G \max_D \varepsilon_{x\sim p_{\text{data}}(x)}[\log D(x)] \\ + \varepsilon_{z\sim p_z(z)}[\log(1 - D(G(z)))], \end{aligned} \tag{3}$$

where $p_{\text{data}}(x)$ and $p_z(z)$ represent the distributions of noise variables z and real data x. The two discriminators $\{d_1, d_2\}$ share the same definition of the loss function. The unique difference is that the local discriminator returns loss gradients for the occluded region while the global discriminator back-propagates loss gradients through the whole face image. The overall loss function is defined by

$$L = L_r + \lambda_1 L_{d_1} + \lambda_2 L_{d_2}, \tag{4}$$

where λ_1 and λ_2 are weighting hyper parameters to balance different losses.

3.4. Cascade Deep Regression Networks Combined with GAN-IA. The central idea of our approach is the design of coarse-to-fine cascade. The deep regression network and GAN-IA complement each other under a cascade structure. On the one hand, with more accurate face shape, the appearance variations within each face component become more consistent, leading to more compact GAN-IA for better generated images. On the other hand, the deep regression networks that are robust to occlusions can be attained by leveraging better generated faces.

To learn a GAN-IA network, we need a training dataset including the genuine face images and the face images occluded by various occlusions on genuine images. We depend on the dataset CelebAMask-HQ to do this work and follow DRDA to collect kinds of occlusions to randomly place them in given positions in a face, as seen in Figure 4.

4. Experiments

To the best of our knowledge, it is not difficult to predict the landmarks of normal faces for most of existing methods. However, the localization accuracy of these methods would drop significantly if face components are partially occluded. To evaluate the effectiveness of our proposed method for occlusion problem of facial landmark detection, we carry out our experiments on two challenging datasets, including OCFW and COFW. For restoring the occluded face part, we employ the CelebAMask-HQ dataset for training GAN-IA.

4.1. Datasets

4.1.1. OCFW. OCFW is a collection of 3, 837 face images from existing datasets: LFPW, HELEN, AFW, and IBUG. Its 68 landmarks for each face image are manually annotated and published in website. Following the setting reported in [2], we use 2591 images as training samples and 1246 images as testing samples. All training images are occlusion-free while all testing images are occluded partially, which poses a challenge scenario for facial landmark detection under occlusion condition.

4.1.2. COFW. COFW [1] is an occluded face dataset in the wild and consists of 1,345 images for training and 507 images for testing. Specifically, its training set is composed of two parts: (1) 845 faces from LFPW training set and (2) extra 500 faces with severe occlusions. All samples of test set are partially occluded and all face images of the entire dataset are

(a) (b) (c) (d)

Figure 3: The comparison of facial occluded image recovery with or without self-attention mechanism. (a) Original image, (b) the occluded image, (c) the left eye and right eye are asymmetric of the third image recovered face without self-attention mechanism, and (d) recovered face without self-attention mechanism.

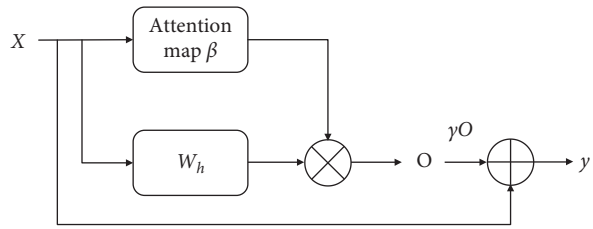

Figure 4: Self-attention structure diagram.

hand annotated using 29 landmarks. There is another manually annotated 68 landmarks dataset for test set for the objective to compare with previous methods easily. In our experiments, we only use 68 landmarks to evaluate the effectiveness of addressing occlusion problem of our proposed model.

4.1.3. CelebAMask-HQ. CelebAMask-HQ introduced in [27] is a large-scale face image dataset that has 30,000 high-resolution face images selected from the CelebA dataset. The images are originally stored as HDF5 format (.h5), which are not suitable for common data loaders. Therefore, we write a script code to generate the 256×256 images and save CelebA-HQ images in JPEG format (.jpg) for the needs of GAN-IA.

4.2. Implementation Details. The deep regression networks of our improved model follow DRDA for fair comparison with the original model. To verify effectiveness of our work in improving the performance, an ablation study is conducted by utilizing other deep regression networks, e.g., PFLD. In our experiments, we reimplement the PFLD model by employing the tensorflow framework as the code is not released by its authors.

We also divide the face shape of 68 landmarks into seven components, and for each component, the GAN-IA model pretrained on the CelebA dataset is used to recover the facial occluded parts. An overview of the general training procedure is as suggested in [7]. The training is split into three phases: first, the generator network is trained with the MSE loss for T_G iterations; second, the generator network is frozen, and the discriminators are trained from scratch for T_D iterations; finally, both the generator network and two discriminators are trained jointly until the end of training. The pretraining of the generator and the discriminator networks has proved critical for successful training. The input for the global context discriminator is the full 256×256-pixel image, and for the local context discriminator, the input is a 128×128-pixel patch centered around the generated region (or a random area for real samples). To avoid the generated area having subtle color inconsistencies with the surrounding regions, we perform simple post-processing by Poisson image blending. Figure 5 shows the recovered results of samples in CelebAMask-HQ, and the proposed method can well recover the genuine appearance from the occlusions.

All models are trained with the tensorflow framework on a NVIDIA GTX 1080Ti GPU.

4.3. Evaluation Metrics. To verify the performance of our model on occluded faces, we use two evaluation metrics: the normalized root mean squared error (NRMSE) calculated with respect to the interocular distance and the cumulative error distribution (CED) curve. NRMSE can be formulated as follows:

$$\text{NRMSE} = \frac{1}{N} \sum_{i=1}^{N} \frac{\left\| S_i - \widehat{S}_i \right\|^2}{L\Omega_i}, \tag{5}$$

where L denotes the number of landmarks on one face and Ω represents the interocular distance.

4.4. Evaluations on OCFW Datasets. Firstly, we conduct the experiments on the OCFW dataset to evaluate our method and the existing approaches. All methods employed to predict 68 facial landmarks are trained on OCFW training set and are evaluated on OCFW testing set. As illustrated in Figure 6, we compare the improved method with other representative methods via two kinds of evaluation criteria. As seen, RCPR performs better than SDM, and it is possible that RCPR designs the interpolated and applies smart restart strategy to relieve the sensitivity of shape initialization [6].Then, DRDA outperforms RCPR when the NRMSE is above 0.04, which owes to the powerful ability of the modeling facial occlusion region and the favorable ability of modeling nonlinear mapping from facial appearance to facial shape. Furthermore, our methods (GAN-IA coupled with the deep regression network (DRN) or PFLD) achieve superior performance in comparison with other methods, with an improvement up to 10% compared with DRDA when NRMSE is 0.06. This obvious improvement can be

Figure 5: Visualized results of reconstruction on CelebAMask-HQ. The first row shows the original face images. The images of the second row are occluded by various occlusions. The corresponding recovered faces are shown in the third row.

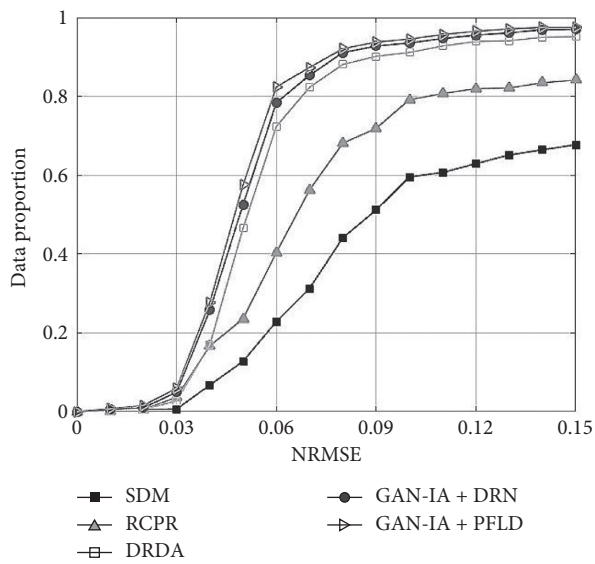

Figure 6: Comparisons of the CED curve on the OCFW dataset.

attributed to the schema of effectively coupling deep regression networks with GAN-IA and demonstrates that better recovering the regions can achieve better landmark detection performance.

Figure 7 shows landmark detection results of DRDA and our improved model on some challenging samples. It can be obviously observed that there is a clear boost in the robustness of facial landmark detection to partial occlusions under different conditions.

From Table 1, it can be found that each proposed improved module plays an essential part in improving the performance.

4.5. Evaluations on COFW Datasets. We further evaluate the detection accuracy of all methods on COFW. COFW is another challenging dataset with real-world faces occluded to different degrees. In addition, it also contains large shape and appearance variations due to pose and expression. All methods are evaluated on COFW test set in terms of 68 facial

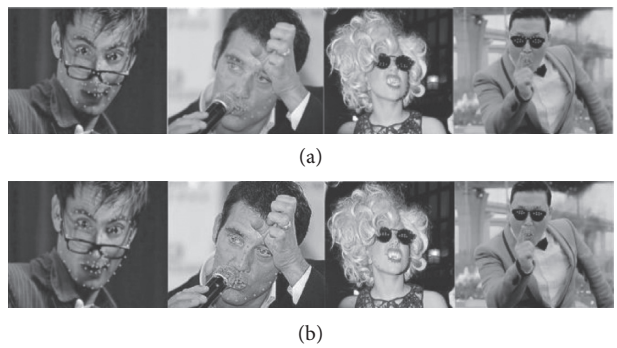

Figure 7: Visualized results of (a) DRDA and (b) our proposed model.

Table 1: NRMSE($\times 10^{-2}$) comparisons of our proposed model with different modules on challenging set.

Model	NRMSE
Autoencoder + DRN	7.23
GAN-IA + DRN	6.96
GAN-IA + PFLD	6.71
GAN-IA(without SA) + DRN	7.15

landmarks according to another version dataset. For RCPR, we use the model released by authors for evaluations, which is trained with occlusion annotations for robust face alignment under occlusions and achieves promising results on COFW test set [7].

The performance of all methods is illustrated in Figure 8. Similar conclusions can be achieved. As seen, RCPR always performs better than SDM during the entire NRMSE section. RCPR explicitly predicts occlusions and utilizes the occlusion information to help shape prediction so as to achieve robustness to partial occlusions. Benefited from the GAN-IA networks, our method performs better than RCPR and DRDA, with an improvement up to 8% compared with DRDA when NRMSE is 0.08.

Some localization results of our method are shown in Figure 9. The first row shows the alignment results under occlusions simultaneously with varying poses. The secondary row exhibits exemplars under simultaneous occlusions and expressions. Samples with a variety of occlusions (e.g., sunglasses and respirator) are illustrated in the last row. As seen, our method is robust to different types of occlusions under various poses and expressions.

4.6. Ablation Study. Our improved network for facial landmark detection task consists of two pivotal modules: GAN-IA and deep regression networks. In this subsection, we carry out the ablation study to validate their effectiveness on challenging set. Based on the introduced methods for increasing performance, we analyse the necessity of existence for each proposed module. Table 1 reports the comparison results of NRMSE.

From Table 1, it can be found that each proposed improved module plays an essential part in improving the performance. Furthermore, it can be obviously observed that the best performance comes from DRDA (the first row) equipped with all modules simultaneously. Moreover, in our

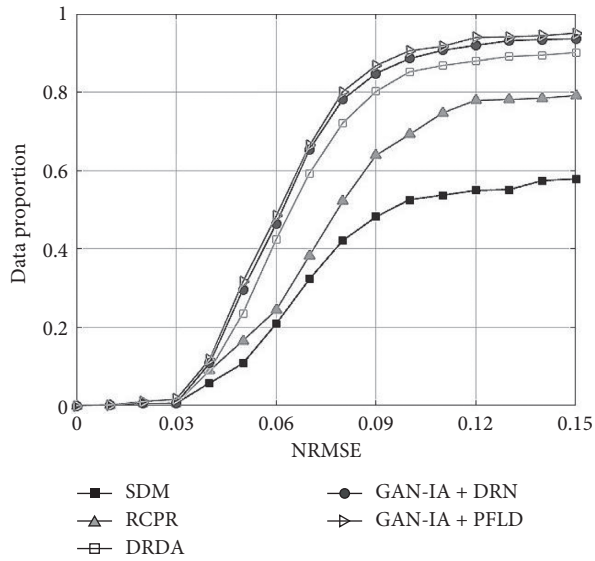

Figure 8: Comparison results of different methods on the COFW dataset.

Figure 9: Exemplar results on COFW. (a) The first row: occluded images under various poses; (b) the secondary row: images under occlusions together with different expressions; and (c) the last row: images occluded by a variety of objects.

Table 2: AUC0. 08 and failure rate of the model key points with different modules on challenging set.

Model	AUC0.08	FR (%)
GAN-IA (without SA) + DRN	49.97	5.08
GAN-IA+PFLD	52.12	4.21
GAN-IA+DRN	52.68	3.86
Autoencoder + DRN	53.76	2.94

proposed framework, self-attention mechanism is imposed to networks and makes it work better as well. In addition, it can be seen from Table 2 that the AUC of the model equipped with all modules reached 53.76, and the failure rate reached the lowest 2.94%

5. Conclusion

In this work, we introduce a novel generative adversarial network with improved autoencoders to restore the partially occluded face region via deep regression networks to solve the occlusion problem for facial landmark detection task, which consists of three main modules: generator module, local discriminator module, and global module. The generator employs skip concatenation to restore more details in occlusion regions. The local discriminator and global discriminator are considered as the auxiliary network to help produce realistic images. Meanwhile, self-attention mechanism that is effective in modeling long-range dependencies is introduced to recover harmonious images for occluded faces. We conduct the experiments on challenging datasets to evaluate the effectiveness of our work under occlusion condition. The experimental results show that our proposed method has a significant performance improvement and achieves robustness against occlusions.

Data Availability

In our research, we made experiments to verify our theory. However, the experimental data in our paper used to support the findings of this study have not been made available because the data will be used commercially.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (grant nos. 61871039, 61932012, 61802019, and 61906017), the Beijing Municipal Commission of Education Project (Nos. KM202111417001 and KM201911417001), National Engineering Laboratory for Agri-Product Quality Traceability Project (No. AQT-2020-YB2), the Supporting Plan for Cultivating High Level Teachers in Colleges and Universities in Beijing (grant no. IDHT20170511), and the Academic Research Projects of Beijing Union University (grant nos. ZK80202001, XP202015, BPHR2020EZ01, BPHR2019AZ01, 202011417004, 202011417005, 202011417SJ025, and KYDE40201702).

References

[1] X. P. Burgos-Artizzu, P. Perona, and P. Dollár, "Robust face landmark estimation under occlusion," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1513–1520, Sydney, Australia, December 2013.

[2] J. Xing, Z. Niu, J. Huang et al., "Towards robust and accurate multi-view and partially-occluded face alignment," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 987–1001, 2017.

[3] I. Goodfellow, "NIPS 2016 tutorial: generative adversarial networks," 2016, http://arxiv.org/abs/1701.00160.

[4] C. Zheng, T. J. Cham, and J. Cai, "Pluralistic image completion," in *Proceedings of the IEEE Conference on Computer*

Vision and Pattern Recognition, pp. 1438–1447, Long Beach, CA, USA, November 2019.

[5] H. J. Lee, S. T. Kim, H. Lee et al., "Lightweight and effective facial landmark detection using adversarial learning with face geometric map generative network," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 3, pp. 771–780, 2019.

[6] M. Zhu, D. Shi, M. Zheng et al., "Robust facial landmark detection via occlusion-adaptive deep networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3486–3496, Long Beach, CA, USA, November 2019.

[7] J. Zhang, M. Kan, S. Shan et al., "Occlusion-free face alignment: deep regression networks coupled with de-corrupt autoencoders," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3428–3437, Las Vegas, NV, USA, June 2016.

[8] S. Iizuka, E. Simo-Serra, and H. Ishikawa, "Globally and locally consistent image completion," *ACM Transactions on Graphics*, vol. 36, no. 4, p. 1, 2017.

[9] O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," in *Proceedings of the International Conference on Medical Image Computing and Computer-assisted Intervention*, pp. 234–241, Springer, Cham, Switzerland, October 2015.

[10] H. Zhang, I. Goodfellow, D. Metaxas et al., "Self-attention generative adversarial networks," 2018, http://arxiv.org/abs/1805.08318.

[11] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models-their training and application," *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38–59, 1995.

[12] D. Cristinacce and T. F. Cootes, "Feature detection and tracking with constrained local models," *BMVC*, vol. 1, no. 2, p. 3, 2006.

[13] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 681–685, 2001.

[14] G. Tzimiropoulos and M. Pantic, "Gauss-Newton deformable part models for face alignment in-the-wild," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1851–1858, Columbus, OH, USA, June 2014.

[15] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 2, no. 4, pp. 433–459, 2010.

[16] V. Kazemi and J. Sullivan, "One millisecond face alignment with an ensemble of regression trees," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1867–1874, Columbus, OH, USA, June 2014.

[17] S. Ren, X. Cao, Y. Wei, and J. Sun, "Face alignment via regressing local binary features," *IEEE Transactions on Image Processing*, vol. 25, no. 3, pp. 1233–1245, 2016.

[18] X. Miao, X. Zhen, X. Liu et al., "Direct shape regression networks for end-to-end face alignment," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5040–5049, Munich, Germany, September 2018.

[19] A. Dapogny, K. Bailly, and M. Cord, "DeCaFA: deep convolutional cascade for face alignment in the wild," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 6893–6901, Seoul, Korea, November 2019.

[20] J. Wan, J. Li, Z. Lai et al., "Robust face alignment by cascaded regression and de-occlusion," *Neural Networks*, vol. 123, 2019.

[21] J. Wan, J. Li, J. Chang, Y. Wu, X. Li, and H. Zheng, "Face alignment by component adaptive mechanism," *Neurocomputing*, vol. 329, pp. 227–236, 2019.

[22] M. Kowalski, J. Naruniec, and T. Trzcinski, "Deep alignment network: a convolutional neural network for robust face alignment," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 88–97, Honolulu, Hawaii, July 2017.

[23] W. Wu, C. Qian, S. Yang et al., "Look at boundary: a boundary-aware face alignment algorithm," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2129–2138, Munich, Germany, September 2018.

[24] X. Guo, S. Li, J. Zhang et al., "PFLD: a practical facial landmark detector," 2019, http://arxiv.org/abs/1902.10859.

[25] L. C. Chen, G. Papandreou, I. Kokkinos et al., "Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834–848, 2017.

[26] H. Noh, S. Hong, and B. Han, "Learning deconvolution network for semantic segmentation," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1520–1528, Santiago, Chile, December 2015.

[27] C. H. Lee, Z. Liu, L. Wu et al., "MaskGAN: towards diverse and interactive facial image manipulation," 2019, http://arxiv.org/abs/1907.11922.