Discussion on the Development of Face-swap Methods

Yibo Wang

North China Electric Power University, Baoding 071003, China

ncepu.north@gmail.com

Abstract. Face-swap technology is a technology that can transfer source face identity information to the target face and maintain some other attributes of the target face, also known as deepfake. Among these methods, the first to appear were target-specific ones, which tend to have better performance but lower efficiency. As deep learning techniques continue to evolve, the pursuit of algorithmic generalization has begun, using pre-trained models to swap different faces. Although the generalized models are very efficient, the quality of certain face-swap images and videos generated is not satisfactory. Thus, people want to find an algorithm with both generality and performance. In the context of the current general pursuit of both generality and performance of deep forgery mehtods, this paper examines the development of deep forgery methods from the early days to the present and discusses the question of whether algorithm generality and performance can be achieved at the same time. After that, this paper proposes a deep forgery dataset, which contains the wild face videos we found from the Internet video community Bilibili, as well as the forged videos and images we generated using the four methods: Deepfakes, DeepFaceLab, FSGAN, and FaceDancer, and tests on this dataset the ability of the existing deep forgery methods for the generation capability of wild faces. Finally, we change the direction of identity vector usage for the effect of the target face on identity transfer during the generation process and propose a metric to measure the identity transfer error of the generated face caused by the target face.

Keywords: Deepfake ;Deepfake Datasets; Deepfake Evaluation Metric; Benchmark.

1. Introduction

Face exchange techniques aim to transfer identity information from a source face to a target face and maintain a portion of the target face's attributes such as expression, pose, background, and illumination. The earliest face exchange technology [[1],[2],[3],[4]] is the application of computer graphics, which is poorly generated and inefficient. With the development of artificial intelligence technology, methods for exchanging faces using deep learning [[5], [6], [7], [8], [9]] have grown by leaps and bounds. Face swapping using deep learning techniques to manipulate faces, also known as deep faking, has broad application prospects in the film and television, entertainment, and privacy protection industries.

The earliest face-swap methods that appeared were target-specific [[5], [6]], and it was very time-consuming to train a separate model for each pair of faces. With the continuous development of deep forgery technology, non-target-specific methods [[9]] appeared, which solved the disadvantage of time-consuming training models, utilized pre-trained models, and only needed to provide photos or videos of the source face and the target face to complete the exchange. However, these two methods have their advantages, target-specific methods due to the training of a specific target, the quality of the generated results better; non-target-specific methods, although the solution to the time-consuming problem, it performs poorly when dealing with wild faces. Thus, it became desirable to find a method that has both better performance and better generalization. This leads to several questions: 1. Can the method have both generality and performance? 2. Most of the existing methods are trained on some public datasets, if they face faces in wild face datasets, can they still show better generality, and what is their ability to fake? Besides, the evaluation metrics of the performance of existing methods seem to be not comprehensive enough. The contribution of this paper is:

- **Horizontal comparison of methods.** We select a few mainstream deepfake methods from different periods and with different methods and compare them side-by-side to determine how algorithmic performance has evolved. This kind of comparison is also often seen in some algorithmic literature, but they either select only the best methods at that time or methods with similar methods to theirs for comparison. We select methods from different periods and methods for horizontal comparison to be able to get more objective conclusions.

- **A more challenging wild face deepfake dataset.** We propose a new deep forgery fake face image video dataset GTW (Generation Test Datasets of Wild Face), which includes several wild face videos selected from the Internet video community, and use a variety of methods to generate deepfake videos and images on wild face to test the performance of face-swapping methods on wild faces..
- **A new evaluation metric for deepfake images**. We propose a new perspective on the application of identity vectors in the measurement of deep forgery evaluation metrics by obtaining a new evaluation metric, Tgt-ID Error. Tgt-ID Error is used to measure the error resulting from the mis-transmission of the target face's identity information to the generating face, which is a direction that cannot be evaluated by other metrics that currently exist.

2. Related Work

2.1 Deepfake Methods

3D-based methods. 3D-based depth forgery methods were a technique of wide interest to early researchers. Face2Face [[1]], proposed by Thies et al. is an early graphics-based 3D model reenactment algorithm. The algorithm maps one person's facial expression onto another person's facial model in real-time by capturing the 3D motion of the facial expression. This allows the user to manipulate the expression of the target person in real-time through the camera, achieving a highly realistic reenactment effect. Nirkin et al. [[3]] used a generic face model and used detected 2D landmarks to fit the pose and expression of the 3D model, overlaid the 3D model with the face obtained from FCN segmentation, and finally fused the estimated source face model to the target face to achieve face swapping. However, the 3D-based method cannot be popularized due to its high time cost and computational complexity.

Auto-encoder based methods. Deepfakes [[5]], which appeared in 2017, is an early deep learning-based face-swap project on the web, popular for its ease of use and high-quality synthesis results. Deepfakes trains two pairs of encoders and decoders for two inputs, where the encoders share weights.DeepFaceLab [[6]] is another open source face-swap tool that subsequently emerged with a face-swap process similar to Deepfakes' face-swap process. It provides a clear exchange process and a flexible and variable framework for face exchange to meet different needs. In addition, it allows users to manually adjust the parameters, which greatly improves the quality of face swapping.

GAN-based methods. The GAN technique introduces two neural networks: a generator and a discriminator, and lets the two neural networks compete with each other during training to achieve better training results. Korshunova et al. [[7]] consider face-swap as a type of style transfer and use a network structure with multiscale inputs to capture the features of the target face.IPGAN [[8]] proposes a framework for decomposing the identity and attributes of faces based on the image inputs Decompose the identity and attribute vectors and synthesize faces by reorganizing the identity and attribute vectors from different faces.FSGAN [[9]] is a goal agnostic face-swap algorithm, that reenacts the information of the target face's pose and expression on the source face, segments, repairs according to the segmentation mask of the target face, and finally fuses it with the target face.

Latent code based methods. Latent codes are codes that are used in the generative model to represent potential features or concepts of the input data and contain key features of the data. By manipulating these latent codes, finer control over specific features can be achieved. MegaFS [[10]] is an algorithm for high resolution face swapping that uses a multiscale encoder to predict the latent codes and proposes a latent code manipulation module to synchronize the management of multiple attributes and finally generates the results through styleGAN2. In addition, MegaFS utilizes pre-existing face masks in a post-processing manner to enhance the face features. RAFSwap [[11]] enhances identity consistency in two directions, local and global, respectively, where the local branch introduces a transformer to enhance local features. Unlike Megafs, the algorithm proposes a soft mask prediction module and incorporates it into the algorithm to enhance identity features. However, both methods are more dependent on the mask provided in CelebA-HQ. After Megafs, another high-resolution face-swap method, FSLSD [[12]], was proposed by Xu et al. Unlike Megafs, they explicitly decouple face attributes and divide them into structural attributes (identity, expression, pose) and appearance attributes (illumination, background), and use a landmark-driven structure to transfer the expression and pose features to achieve the separation of these two attributes from identity information. Facedancer [[13]] introduces the Adaptive Feature Fusion Attention module in the encoder Adaptive Feature Fusion Attention adaptively learns to fuse the attribute features with the identity features and introduces another identity encoder to enhance the identity features.

2.2 Deepfake Datasets

FaceForensics++ [[15]] contains 1000 videos from YouTube, and a face detector is used to ensure that there are faces in consecutive frames during the screening of the footage to improve the video quality. Five deep forgery methods such as Deepfakes, Face2Face, FaceShifter, FaceSwap, and NeuralTextures are then used to generate the forgery videos, totaling 5000 videos, which are widely used in the task of deep forgery detection. Because it contains deep forgery videos of five methods, it is also often used as an algorithm evaluation benchmark.

CelebA-HQ [[16]] contains 30k 1024×1024 high-resolution celebrity images.CelebA-HQ is generated by increasing the resolution of low-resolution images by super-resolution techniques on images from CelebA [[17]], a large-scale dataset containing 20w celebrity images. CelebA-HQ has been applied to the training of face editing techniques because of its high resolution.

3. Performance Evaluation of Deepfake Methods on Public datasets

3.1 Experimental Settings

In this paper, some mainstream deep forgery methods in each period, including Deepfakes, Deepfacelab, FSGAN, Megafs, RAFaceSwap, and Facedancer, are selected, involving three methods: autoencoders, GANs, and latent codes. Two datasets FaceForensics++ and CelebA-HQ are used as benchmark datasets for this evaluation. The 1000 videos in FaceForensics++ are mainly used directly as input data, whereas for mehtods that use images for exchanges, we take ten frames from each of the 1000 videos in FaceForensics++, totaling 10k frames, as input data. For these 1000 videos and the frames they contain, in this paper, we use the first 500 as source faces and the last 500 as target faces to generate 5000 fake faces. It is divided into two groups according to the different benchmark datasets, Deepfakes, Deepfacelab, FSGAN, and Facedancer are tested on FaceForensics++, Megafs, and RAFSwap rely on the mask on CelebA-HQ and can only be tested on this dataset, in addition, Facedancer is also tested on this dataset.

We choose the widely used ID similarity, pose error, expression error, and The Frechet Inception Distance (FID) as the evaluation metrics to evaluate the performance of each algorithm in various aspects. We use an identity vector extractor [[18]] to extract identity vectors, an open source pose estimation model [[19]], and a 3D facial model [[20]] to extract pose and expression vectors to compute the above metrics.

3.2 Experiments on public datasets

In this section, we perform qualitative and quantitative comparisons of the two sets of exchange results, respectively. For quantitative comparison, we first manually check the generation results to remove the images that failed to be generated, followed by face region cropping of the generated images using the face detector provided by Dlib, and finally testing.

3.2.1 Qualitative comparison

Figure 1 shows the exchange results on CelebA-HQ. It can be seen that FaceDancer generates a poor quality image, which may be caused by the fact that the pre-trained model used in this paper has not been trained on CelebA-HQ. In contrast, MegaFS and RAFSwap, which rely on CelebA-HQ, have better results, with RAFSwap's swapping results possessing a texture closer to that of real faces.RAFSwap's handling of the background and illumination is better than that of MegaFS, and Megafs' swapping results generally have hair color inconsistencies with that of the target face. In addition to this, MegaFS also fails to generate facial regions in the experiments, which may be related to the different masks used by the two mehtods, Megafs uses the hard mask from CelebA-HQ, while RAFSwap uses its own predicted soft mask. In the fourth line, we can also see that Megafs incorrectly retains the eyebrow information of the target face.

Figure 1 Exchange results on CelebA-HQ

Figure 2 shows the exchange results on FaceForensics++, and it can be seen that among the four mehtods, FaceDancer has the best exchange results.Deepfakes has obvious artifacts and slightly blurred faces. DeepFaceLab(DFL) is a target-specific algorithm, has some improvement over the former, but the illumination processing is still problematic.FSGAN, on the other hand, has both artifacts and illumination problems, which may be related to the instability of GANS. In addition to this, as can be seen from the first and third rows, FSGAN's expression information delivery also produces some errors in the visualization.

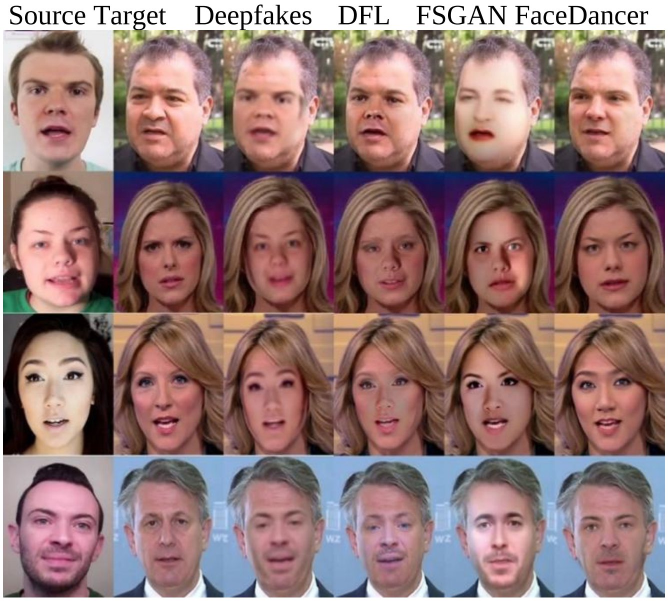

Figure 2 Exchange results on FaceForensics++

Overall, the performance of face-swap mehtods shows a trend of improvement in terms of visual effects: FaceDancer swaps face with a texture closer to the real face than previous methods, and this is largely due to its excellent illumination processing; RAFSwap performs better in terms of hair and skin texture compared to MegaFS. All of these advances make the swapping results more realistic. However, there is also the phenomenon that the new mehtods are not as good as the earlier proposed mehtods, e.g., the visualization of faces generated by FSGAN is worse compared to DeepFaceLab. This is mainly due to the fact that FSGAN does not handle light well.

3.2.2 Quantitative comparison

RAFSwap achieves better results in both ID Similarity and Pose Error metrics, and there is a big improvement in ID Similarity, while MegaFS performs a little better in FID and Exp Error metrics, and there is a big improvement in FID, as shown in Table 1. This means that RAFSwap performs better on pose and identity information processing, while MegaFS performs better on expression processing.

FaceDancer achieves the best results in two of the four metrics, and it can be seen that the two indicators, Pose Error and FID, are gradually decreasing with the development of the algorithm, and Exp Error also shows a decreasing trend in general, as shown in Table 2. It indicates that the performance of the deep forgery method is improving under these three evaluation criteria, in which the gradual decrease of the first two indicators indicates that the algorithm is more successful in the reenactment of poses, and the algorithm is becoming more and more realistic in its generation. However, for ID Similarity, the result is floating.

Table 1 Quantitative comparison on CelebA-HQ.

Method	**ID Simi.↑**	**Pose Err.↓**	**Exp Err.↓**	**FID↓**
MegaFS	0.4523	4.2815	**2.5813**	**24.1407**
RAFSwap	**0.5176**	**3.6915**	2.7267	46.6952

Table 2 Quantitative comparison on FaceForensics++.（sorted by newness）

Method	**ID Simi.↑**	**Pose Err.↓**	**Exp Err.↓**	**FID↓**
Deepfakes	0.4884	4.2629	3.0748	61.2459
DeepFaceLab	**0.5377**	3.2452	2.9396	59.9866
FSGAN	0.4359	2.6773	**1.9443**	11.6990
FaceDancer	0.5176	**2.4405**	2.3731	**8.9494**

From these evaluation criteria, the overall development of algorithm performance is getting better, but not all aspects have improved. We find some problems with the existing methods: 1. Some of the face-swap methods suffer from artifacts and illumination processing problems. 2. Some of the methods rely too much on fixed datasets leading to their reduced utility.3. The cross-dataset generation ability of the mehtods is not good enough, MegaFS, RAFSwap are completely dependent on the CelebA-HQ dataset; on FaceForensics++,FaceDancer, the best performer, can barely generate faces correctly on CelebA-HQ. Through qualitative and quantitative comparisons, we also find that method performance cannot evolve to perform better and better in all respects. The target-specific mehtods generate even better quality than the new generalized face-swap mehtods. This is because target-specific mehtods are trained on a specific pair of faces, whereas the non-target-specific mehtods that are not trained on specific faces is limited by the across datasets.

4. Wild Face Dataset

In this section, we introduce our newly proposed dataset GTW (Generation Test Datasets of Wild Face) and test the generation ability of existing mehtods on this dataset for wild faces.We collect 16 high-quality 1080P videos from the Internet video community Bilibili and use Deepfakes, DeepFaceLab, FSGAN, and FaceDancer to generate fake videos to build the dataset. We downloaded their original videos through the video links, To ensure the high quality of the data, we first recorded the videos in the community to get the 1080P videos, and then edited and corrected these videos to ensure that there are human faces in each frame, and some of the wild faces of GTW are shown in Fig. 3. Table 3 shows the various data of GTW.

Table 3 Data of GTW

Dataset	Deepfake Videos	Deepfake Pictures	Methods	resolution	Vid Total Duration	Vid Avg Duration
GTW	498	1980	4	1920×1080	2h 9m 15s	32s

Figure 3 Sample video frames of wild faces on GTW

5. Experiments on GTW

Since some of the method selected above rely on masks, in this section we only use less restrictive mehtods for testing. The experimental setup is still swapped using ten frames taken from each video as above.

5.1 Quantitative Comparison

As shown in Table 4, the trends of these four mehtods on the GTW dataset are the same as the test results on FaceForensics++, but the specific metrics have different degrees of change and are getting worse in general. For ID Similarity, FSGAN and FaceDancer produced substantial decreases, but Deepfakes and DeepFaceLab showed a relatively little decrease, reflecting the advantage of target-specific mehtods. For FaceDancer's Pose Error and DeepFaceLab's FID still have some improvement, which indicates that they are more adaptable in these two metrics.

Table 4 Quantitative comparison on GTW

Method	**ID Simi.↑**	**Pose Err.↓**	**Exp Err.↓**	**FID↓**
Deepfakes	0.4751	5.7575	2.4483	63.0325
DeepFaceLab	**0.5320**	4.3179	**2.3498**	55.0617
FSGAN	0.3739	3.0180	2.8013	41.4517
FaceDancer	0.3993	**2.2441**	2.3812	**26.4837**

5.2 Qualitative Comparison

As shown in Figure 4, the quality of the generated results on GTW decreased compared to the public dataset. The two mehtods for specific targets, Deepfakes and DeepFaceLab(DFL), were able to transfer more source face identity information, but Deepfakes was inaccurate in transferring expressions for certain faces, in addition to which Deepfakes was limited by its algorithm itself, and the generation results were blurred; DeepFaceLab showed abnormal illumination; FSGAN still clearly has its inherent illumination and artifacts; FaceDancer, although generating better results, has a generation failure.

Source Target Deepfakes DFL FSGAN FaceDancer

Figure 4 **Qualitative comparison** on GTW

The qualitative and quantitative comparison of the experimental results show that the existing methods still have limitations in dealing with wildface. On the other hand, the small change in ID Similarity of the two target-specific mehtods again reflects that the cross-dataset generation ability of the target-specific mehtods is better than that of the non-target-specific mehtods, which is a problem that needs to be solved by the existing generalizability models in the pursuit of performance. On the other hand, we can also see that on the GTW dataset, the generation quality of the target-specific face-swap algorithm is not very different from that of the general-purpose algorithm, and even better than that of the general-purpose algorithm in some cases, so it is the future direction of face-swap mehtods still to adhere to the direction of improving the general-purpose nature of the model for the development of face-swap mehtods? Here we give a feasible suggestion for the future development of face-swap mehtods: optimize the structure of the target-specific algorithm model, lighten the model, improve the training speed, and achieve high-quality face-swap under small samples, so that each face-swap can be trained for a short time and generate the corresponding lightened model, and achieve the " generalizability".

6. A New Evaluation Metric

In the past, regarding the evaluation of identity information, people only focused on whether the identity information transferred from the source face to the generated face was accurate or not but ignored the possible influence of the target face on the identity information of the generated face. In some face-swap methods, it is necessary to distinguish the potential codes of different features, and there is a deep coupling between the identity information and other attribute information such as expressions and poses, which may lead to the transfer of erroneous identity information if it is not well decoupled when transferring information such as expressions or dealing with the illumination and background. Therefore, we compare the identity vectors of the generated face with those of the target face and calculate the cosine similarity as a measure of the error in the generated face caused by the wrong transfer of the identity information of the target face, which we call the Tgt-ID Error. The larger the value of this metric is, the more information is propagated by the error.

As shown in Fig. 4, although FaceDancer performs well in the metric, upon close inspection we can see that the lip color of the exchanged faces in the second and third rows, as well as the lower eyelids of the exchanged faces in the second row, are obviously attributes from the target face, which may be erroneously propagated by its processing of illumination. In contrast, DeepFaceLab's identity transfer does not suffer from such problems. We therefore introduced the Tgt-ID Error metric to measure the error generated by the mis-propagation of identity information of the target face. We measured Tgt-ID Error for the generated results on each previous dataset, and the results are shown in Table 5. The results show that DeepFaceLab indeed performs better on this metric, while other existing evaluation metrics do not reflect such a problem.

Table 5 Tgt-ID Error test results for mehtods on different datasets

Method / Dataset	**Deepfakes**	**DeepFaceLab**	**FSGAN**	**FaceDancer**	**MegaFS**	**RAFSwap**
FaceForensics++	**0.6328**	0.6480	0.7052	0.7622	\	\
CelebA-HQ	\	\	\	\	0.7790	**0.7227**
GTW	0.7711	**0.6676**	0.7527	0.7481	\	\

7. Limitations of Our Study

We have done a lot of work, such as comparing the performance of mainstream face-swap mehtods in various periods, testing the generation ability of each algorithm on wild human faces using the GTW dataset, and proposing a new evaluation metric. However, due to the limitation of time or other conditions, the research in this paper still has some shortcomings.1. The mehtods we selected do not cover all the mainstream mehtods because some mainstream mehtods do not have open source code, such as FaceShifter [[14]]. We will add the research about other mainstream face-swap mehtods later.2. There are still some public datasets that we have not tested, and we will continue to test them later.

8. Conclusion and Outlook

This paper discusses the issues of performance and generalization in the process of algorithm development and finds that in the pursuit of generalization, a certain amount of performance is lost, for example, although advanced generalized mehtods generate more realistic images, the ability to transfer identity information is often inferior to that of mehtods for specific targets. Secondly, this paper discusses the generation ability of existing mehtods on wild face data and finds that the performance of each algorithm decreases to different degrees when facing wild faces, indicating that the existing mehtods are not ready to face wild faces. In addition to this, we address the effect of the target face on identity information transfer and utilize the identity vector in another perspective, proposing Tgt-ID Error to measure the error of identity information mis-transfer from the target face to the generated face. For target-specific mehtods, model light weighting and shortening of training time can be achieved by optimizing the model structure. Achieving approximate generalization while maintaining the performance of target-specific mehtods can be considered as one of the future research directions.

References

[1] Thies, J., Zollhofer, M., Stamminger, M., Theobalt, C., & Nießner, M. (2016). Face2face: Real-time face capture and reenactment of rgb videos. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2387-2395).

[2] Suwajanakorn, S., Seitz, S. M., & Kemelmacher-Shlizerman, I. (2017). Synthesizing obama: learning lip sync from audio. ACM Transactions on Graphics (ToG), 36(4), 1-13.

[3] Nirkin, Y., Masi, I., Tuan, A. T., Hassner, T., & Medioni, G. (2018, May). On face segmentation, face swapping, and face perception. In 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018) (pp. 98-105). IEEE.

[4] FaceSwap. 2019. https://github.com/MarekKowalski/FaceSwap/

[5] Deepfakes. 2019. https://github.com/Deepfakes/faceswap

[6] DeepFaceLab. 2018. https://github.com/iperov/DeepFaceLab

[7] Korshunova, I., Shi, W., Dambre, J., & Theis, L. (2017). Fast face-swap using convolutional neural networks. In Proceedings of the IEEE international conference on computer vision (pp. 3677-3685).

[8] Bao, J., Chen, D., Wen, F., Li, H., & Hua, G. (2018). Towards open-set identity preserving face synthesis. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 6713-6722).

[9] Nirkin, Y., Keller, Y., & Hassner, T. (2019). Fsgan: Subject agnostic face swapping and reenactment. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 7184-7193).

[10] Zhu, Y., Li, Q., Wang, J., Xu, C. Z., & Sun, Z. (2021). One shot face swapping on megapixels. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 4834-4844).

[11] Xu, C., Zhang, J., Hua, M., He, Q., Yi, Z., & Liu, Y. (2022). Region-aware face swapping. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 7632-7641).

[12] Xu, Y., Deng, B., Wang, J., Jing, Y., Pan, J., & He, S. (2022). High-resolution face swapping via latent semantics disentanglement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 7642-7651).

[13] Rosberg, F., Aksoy, E. E., Alonso-Fernandez, F., & Englund, C. (2023). FaceDancer: pose-and occlusion-aware high fidelity face swapping. In Proceedings of the IEEE/CVF winter conference on applications of computer vision (pp. 3454-3463).

[14] Li, L., Bao, J., Yang, H., Chen, D., & Wen, F. (2019). Faceshifter: Towards high fidelity and occlusion aware face swapping. arXiv preprint arXiv:1912.13457.

[15] Rossler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., & Nießner, M. (2019). Faceforensics++: Learning to detect manipulated facial images. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 1-11).

[16] Karras, T., Aila, T., Laine, S., & Lehtinen, J. (2017). Progressive growing of gans for improved quality, stability, and variation. arXiv preprint arXiv:1710.10196.

[17] Liu, Z., Luo, P., Wang, X., & Tang, X. (2018). Large-scale celebfaces attributes (celeba) dataset. Retrieved August, 15(2018), 11.

[18] Deng, J., Guo, J., Xue, N., & Zafeiriou, S. (2019). Arcface: Additive angular margin loss for deep face recognition. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 4690-4699).

[19] Ruiz, N., Chong, E., & Rehg, J. M. (2018). Fine-grained head pose estimation without keypoints. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 2074-2083).

[20] Deng, Y., Yang, J., Xu, S., Chen, D., Jia, Y., & Tong, X. (2019). Accurate 3d face reconstruction with weakly-supervised learning: From single image to image set. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops (pp. 0-0).