2020 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT) | 978-1-6654-1589-7/20/$31.00 ©2020 IEEE | DOI: 10.1109/ISSPIT51521.2020.9408901

Anti-Gan: Discriminating 3D reconstructed and real faces for robust facial Identity in Anti-spoofing Generator Adversarial Network

1st Miao Sun
State Key Laboratory of ASIC and System
School of Microelectronics, Fudan University
200433 Shanghai, China
18112020006@fudan.edu.cn

2nd Gurjeet Singh
Department of EECS
Oregon State University
Corvallis, USA
singhg@oregonstate.edu

3rd Patrick Yin Chiang
State Key Laboratory of ASIC and System
School of Microelectronics, Fudan University
200433 Shanghai, China
pchiang@fudan.edu.cn

***Abstract*—3D face reconstruction is an attractive topic in computer vision. We have seen dramatic rise in its development recently. Now the state-of-the-art method can reconstruct a face from a single 2D face image freely, which brings a threat to facial security society. Since they are very similar in feature distributions, an efficient work to discriminate reconstructed face and real face is vital. Since Generative Adversarial Nets (GAN) has been proposed by Ian J. Goodfellow in 2014, it is extensively trained to approximate data distributions of many applications. For its adversarial mechanism, GAN shows a powerful generative ability to get the state of art. Inspired by its adversarial mechanism, we propose a similar framework called Anti-GAN to discriminate an adversarial dataset from real 3D face datasets and reconstructed face datasets. Considering the computation of backpropagation, G and D all adopt convolutional neural network architecture. Additionally, experiments show that Anti-GAN is a powerful way to distinguish real faces and reconstructed faces. At the same time, it can also offer robust features for a facial identity task.**

***Index Terms*—3D reconstructed faces, Identity recognition, Anti-spfooing**

I. Introduction

The renaissance of machine learning provides a doable approach to take advantage of the complex neural networks to explore the connections hiding behind datasets. Many ingenious neural networks are created to solve the problems related to images, speech, digital numbers, sounds and videos. However, they are also vulnerable for some adversarial samples. Especially these days it is easy to make an adversarial attack on a network on the point of application. In [1], adversarial examples are strong perturbations to other samples in a machine learning model. Some solutions are produced to cope with these adversarial samples specially. In [2], the authors introduced the adversarial learning by adding some generated adversarial examples to help the learning model to be robust. [3,4] used Bayesian uncertainty estimates and density estimation in the subspace to distinguish the adversarial samples. Another way is redefining the aiming function considering the adversarial samples. In [5], an adversarial objective function based on the fast gradient sign method is designed to regularize the neural network. However, these neural networks benefit from the back-propagation to obtain an optimal point. Before GAN, many efforts are focusing on designing a powerful architecture to extract the features of different tasks. In[6], a spatial pyramid pooling pyramid in CNN is developed to handle scope invariant. ResNet [7] can extend the CNN to 152 layers without gradient vanishing problem by introducing the skip connection. By replacing the fully connected layer, [8] can realize an end-to-end object segmentation task. You Only Look Once(Yolo)[9] draw many attentions in these years, a multi-task CNN can reach a real-time object segmentation task. However, neural networks depend on finetuning the parameters are sensitive and hard to find a most stable moment to stop. So there is much workload on testing and regularizing the model, due to that the latent variables behind the enormous model are difficult to control and observe. Different from the normal training procedure, GAN introduces the adversarial mechanism to train neural network model. It offers a method to decide when the neural network gets appropriate parameters. In this case, estimation and execution are operated simultaneously generative model is improved dramatically compared with previous methods. Now that we know use the adversarial mechanism that can help us to push the generative model to approximate any data distributions, whether it can be transformed to differentiate two closer data distributions. From this point of view, we proposed a new methodology of GAN to show the usage in the discrimination issue. In Anti-GAN, the generative model is designed to map the space of real data distribution and adversarial data distribution to a high-dimensional space. Then the discriminative model is trained to judge whether the high-dimensional vector comes from a real dataset or an adversarial dataset. In the adversarial process, the generator works as a

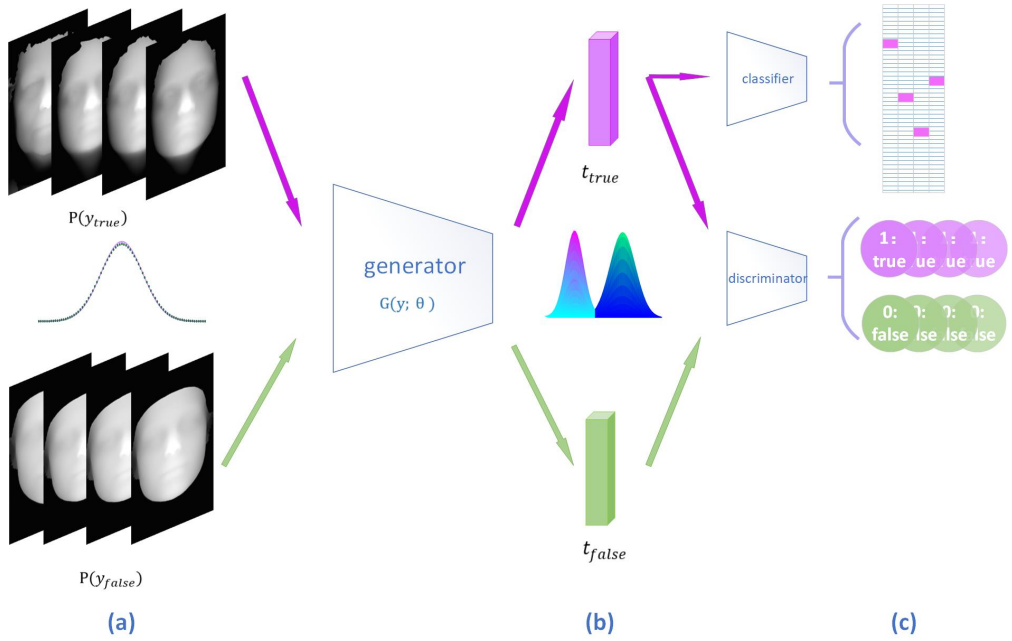

Fig. 1. Overview of the proposed Anti-GAN. (a)The input images are real data distribution $P(y_{true})$ and adversarial smaples$P(y_{false})$. They are close in two-dimension space. (b) Generator in Anti-GAN is used to map the input images into high-dimension space and try to make them be far away from each other. (c) Discriminator is used to realize a binary classification for 3D facial anti-spoofing. Classifier aims to realize the 3D facial authentication by using one-hot code labels.

helper to extract some important features which can be used to differentiate these two candidates and the discriminator works as a detective using these features to determine which dataset the input belongs to. Cooperation in this scheme pushes the generator to find the most effective features until the discriminator can give a correct verdict. The proposed Anti-GAN inherits the advantageous training algorithms from the original GAN. Feature candidates from the generator are produced by the individual forward propagation. Similarly, no approximate inference or Markov chains are necessary. In this framework, we train the model based on the foundation of DCGAN [10], which offers some practical tricks to guarantee the entire model can converge successfully. In this paper, we apply Anti-GAN to solve a 3D face anti-spoofing problem. 3D reconstruction face aims to reconstruct a 3D face from 2D facial images. It offers an efficient way to acquire 3D facial information without real measurement tools and benefits 2D face authentication with rich features. For its attractive values, it has been a popular issue in face analysis. However, it also presents a challenge for recognizing 3D face spoofing attacks and it is empty in this field for mining the subtle differences between 3D reconstructed faces and real 3D faces. To bridge the gap, we use the proposed Anti-GAN to generate the representation features for discriminating these two datasets. At the same time, the output features from generator are robust for identification cases. We also add a subnetwork to realize a 3D facial authentication task. The brief graph of the proposed architecture is shown in Fig.1. Some comparisons with previous classification works are listed to show the efficiency of Anti-GAN. To summarize, the contributions in this paper are as follows:

. 1. We proposed a new framework, Anti-GAN, to realize a dataset discriminating task. We apply Anti-GAN on 3D reconstructed face spoofing robustly. It can also generate useful features for other tasks.

. 2. A new case for our proposed framework which lies the key point on the 3D anti-spoofing scenario. Experiments show that Anti-GAN provides a method to safeguard the neural network from such attacks compared with other classification works.

. 3. Compared with other conventional neural networks, Anti-GAN outperforms a powerful spatial feature extracting ability. It means Anti-GAN learns some meaningful features when it is trained for discriminating adversarial samples, which is helpful for realizing other classification works.

II. Related Work

A. Generator Adversarial Network (GAN)

Generative Adversarial Network[5] is famous for its powerful ability to approximate intractable data distribution by a two-player game. GAN has been developed for many versions either of the applications or theoretical study. Interesting applications in GAN has gushed out with [11-14]. In another aspect, GAN

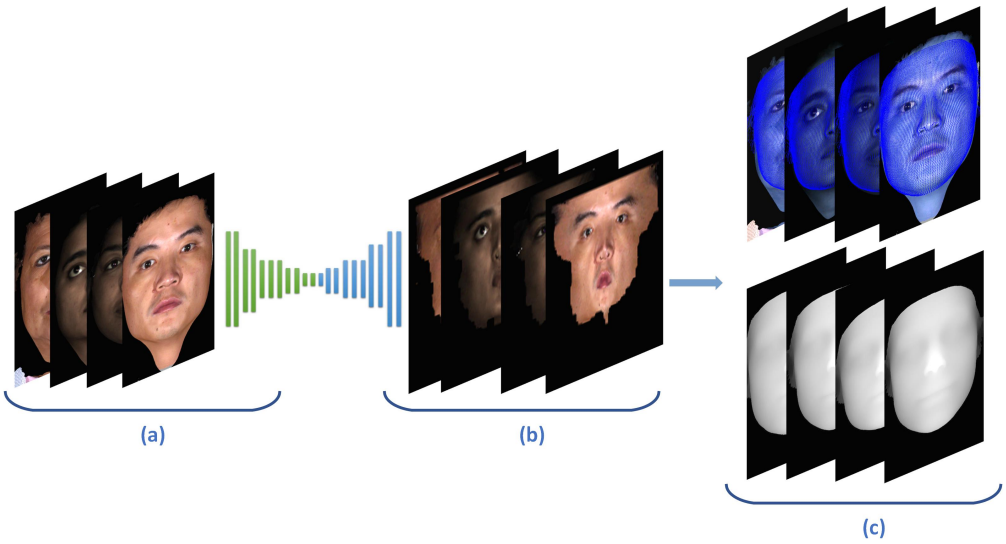

Fig. 2. The preparation for adversarial samples by PRNet.(a) The single image for every sample. (b)PRNet uses a simple encoder-decoder network to regress an UV position map. (c) UV position maps in (b) record both of the position information and dense correspondence. So face alignment and 3D face reconstruction can be obtained simultaneously

has been suffered from the uncontrolled training direction [15], gradient undisappearing [16], stable converging [17]. These works enrich the usage of GAN to get a practical level.

B. Classification Algorithms

The classification task is an important field in Machine learning. Support Vector Machine(SVM) has been the dominated method before CNN. By using the kernel method, SVM can be trained to separate a group of data distributions in high-dimensional space. For some complex scenarios, SVM is limited by its feature extractor ability. In [18], SVM is connected with a feature extractor CNN to complete object detection and classification task. Compared with the end-to-end model, a combination model is troublesome to train and difficult to speed up. Benefiting from the convolutional neural network, classification solutions perform a qualitative leap. From LeNet [19], CNN can get high accuracy in some simple classification tasks. Then AlexNet introduced the dropout, data augmentation and norm layer to CNN, which acquires a great success. After that AlexNet[20], VGG Net[21] and GoogLeNet[22] enrich the CNN family from the number of the architecture and the usage of convolutional layers. In the latest study on machine learning, CNN compression[23] and application fields extension[24] are new trends. Minor changes are made to explore the different forms of neural network layers. Transforming learning and robust learning are more valuable due to that the basic form of CNN has been fixed.

3D face reconstruction With the mature of machine learning, a broad security concern rises up. For a learned system, similar samples may make errors easily. In [25], the authors gave a situation that the attackers may produce a[26], attackers can use an adversarial check to fool the learned model to read a false number rather than the real one. In this paper, we pay attention on a potential attack of 3D facial spoof roused by the 3D face reconstruction technology in these days.

3D face reconstruction has drawn much attention due to that 3D face information can offer abundant information in other tasks such as face authentication, 3D face printing, and virtual reality. From the results of [27,28], the 3D reconstructed face is lifelike even under large poses. This application also creates a new hidden danger for 3D face ID which could be used to attack the 3D camera or data processors to pass the authentication. Because the current 3D reconstructed faces are very similar to real 3D faces, distinguishing a fake 3D face is a challenging task.

III. Anti Generator and Adversarail Network

A. Adversarial nets in Anti-GAN

We propose the Anti-GAN to distinguish two lower-dimension approximated distributions over data (y_{true}, y_{false}) and generate a feature vector containing the distinguishing information. Different from the original architecture in [3], we sample the paired samples from a real dataset and an adversarial dataset. Then map the two similiar domains to a high-dimensional space as $G(y;\theta)$, where y symbols (y_{true}, y_{false}) and G adopts the a deep convolutional structured generator as [10]. We train the generator to learn a different distribution $t_{true} = G(x_{true}, \theta)$ far away from $t_{false} = G(x_{false}, \theta)$. For discriminator, we input vector $t_{true} = G(x_{true}, \theta)$ or $t_{false} = G(x_{false}, \theta)$ for discriminator to do the binary classification. The discriminator is trained to maximize the probability of marking a correct label to

Algorithm Iterative stochastic descent gradicent training of Anti-GAN.

while i < iters **do**

while j < step **do**

- Sampling a batch of N adeversarial samples $(y^1_{false}, ..., y^N_{false})$ from the adeversarial dataset
- Sampling a batch of N real samples $(y^1_{true}, ..., y^N_{true})$ from the real dataset.
- Caculating the gradient of adversarial loss, update the D by stochastic ascend gradicent algorithm.

$$\nabla_{\theta_D} \frac{1}{N} \sum_{i=1}^{N} [log(D(G(y^i_{true})))] + \sum_{i=1}^{N} [log(1 - D(G(y^i_{false})))]$$

end

- Caculating the gradient of G loss, update the G by stochastic descent gradicent algorithm.

$$\nabla_{\theta_G} \frac{1}{N} \sum_{i=1}^{N} (\lambda_1 (1/dist(t_{true}, t_{false})) + \lambda_2 D_{loss})$$

end

the vector from a real dataset or an adversarial dataset. Meanwhile we train the generator to minimize the loss of discriminator $log(D(G(y_{true})) + D(G(y_{false})))$. So we design the adversarial nets in Anti-GAN as a two-player 'minmax' game with the loss fuction $L(G, D)$:

$$\begin{aligned} \min_G \max_D L(D, G) =& \mathbb{E}_{y_{true}}[log(D(G(y_{true})))] \\ &+ \mathbb{E}_{y_{false}}[log(1 - D(G(y_{false})))] \end{aligned} \tag{1}$$

B. Generalized Anti-GAN

a) Discriminator: Discriminator D is trained to judge which dataset the feature vector belongs to. The adversarial samples try to fool the discriminator to ensure that they are close to the real dataset by its similarity. The real dataset makes efforts to obtain a distant vector from the adversarial samples. Therefore, for separate loss function of real dataset and adversarial dataset, we adopt cross entropy to evaluate the estimation mistake:

$$y_i^{'} = \frac{exp(y_i)}{\sum_i^N (exp(y_i))} \tag{2}$$

$$H_{l_i}(y) = -\frac{1}{N} \sum_{i}^{N} \left(p(l_i) log(y_i^{'}) \right) \tag{3}$$

$$\begin{aligned} D_{loss} =& -\frac{1}{N} (\sum_{i}^{N} (p(l_{true,i}) log(y^{'}_{true,i})) \\ &+ \sum_{i}^{N} (p(l_{false,i}) log(y^{'}_{false,i}))) \end{aligned} \tag{4}$$

where y_i is the estimation output from D, $y_i^{'}$ is a softmax output according to the outputlayer, l_i is the label and N is the batch number in training process.

b) Generator: To regularize the generator G to learn an obvious feature vector to differentiate the two datasets. A meaningful loss function is designed to illustrate this iteration direction. According to the goal of G, it is convenient to evaluate the distance between t_{true} and t_{false} in high-dimentional space by calculating euclidean distance. Then the reciprocal of euclidean distance is counted as a final cost function of G as G_{loss}:

$$\begin{aligned} & 1/dist(t_{true}, t_{false}) \\ =& \left(\sqrt{(t_{true} - t_{false})(t_{true} - t_{false})^T} \right)^{-1} \end{aligned} \tag{5}$$

$$G_{loss} = \lambda_1 (1/dist(t_{true}, t_{false})) + \lambda_2 D_{loss} \tag{6}$$

According to different scenes, the loss of G should be changed into other reasonable forms.

c) Overall procedure: Overall trainging process is depicted in Algorithm 1. In Algorithm 1, step is a hyper-parameter for the update of D and we chose it as 2 in case that D converges too fast in our experiments and $\lambda_1 = 1, \lambda_2 = 1$ For parameter iters, it depends on the scale of your datasets. A reasonable set of iters is observing the loss of D. When D_{loss} is close to 1 over the entire training dataset, it is time to stop training.

IV. Anti Generator and Adversarial Network for 3D facial anti-spoofing application

In this section, we introduce a new scene of the adversarial attack. Two main 3D reconstruction algorithms will be shown for training datasets preparation. A brief description of a 3D facial dataset in this case is contained. In the end, the detail information on Anti-GAN in this paper will be discussed.

A. Texas 3D Face Recognition Database

To evaluate our algorithm, Texas 3D Face Recognition Database [29,30] is used as the main test bench. It contains 1149 2D and 3D facial images in pairs of 105 adult human

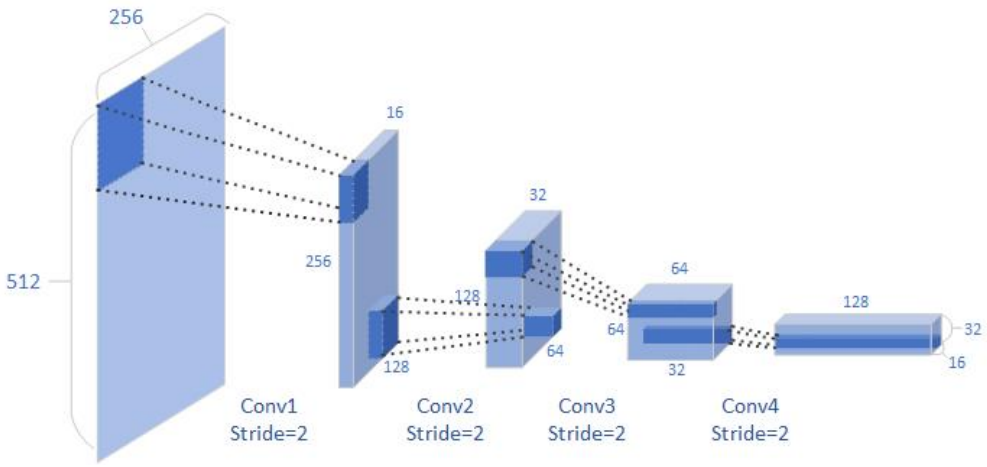

Fig. 3. Anti-GAN generator architecture for 3D facial anti-spoofing. In generator, we use a 4-layer convolutional neural network to produce a 32x16x128 vector for discriminator and other expanded tasks.

subjects. Because it is collected by a stereo imaging system manufactured by 3Q Technologies (Atlanta, GA), it can obtain a very high spatial resolution of 0.32 mm along the x, y, and z dimensions. It can satisfy our requirements for 2D and 3D images. In the next step, we use the 2D images to generate an adversarial samples and use the measured 3D images as the real samples.

B. PRNet

3D reconstruction algorithms can be used as a prerequisite for printing 3D facial masks without a depth sensor. Since 3D Morphable Model(3DMM) is developed in 1999 [31], it is a standard template for other researchers to study[32-34]. This method depends on the accuracy of facial landmarks heavily. [35,36] introduced the CNN into 3DMM to solve the nonlinear optimization problem. However, how to obtain the accurate 3DMM coefficients is remaining to be thought. The latest study on this field in [37,38], they use a cascaded CNN architecture and a volumetric CNN structure separately to estimate the 3DMM shape parameters at the cost of time consuming. In [39], the authors introduced the UV position map into 2D space to construct an end-to-end CNN to realize to reconstruct a 3D facial graph from a single image.Because they do not use any template to estimate the face model, they can take use of the semantic meaning well. Besides, considering the speed performance, their network is light-weighted with 9.8ms for processing one image. In this paper, we adopt PRNet to construct our adversarial dataset. In Fig.2, the preprocessing of the adversarial is illustrated:

C. Anti-GAN structure for 3D facial anti-spoofing task

a) Foundation of Deep convolutional Generative and Adversarial Network(DCGAN): Among variant versions of GAN, DCGAN bridges the gap between CNN and unsupervised learning. Because it tries to learn the hierarchy of representations of object parts, its architecture is constrained to a certain combination of convolutional layers. Due to that factor, DCGAN is a powerful feature extractor. Compared with maximum likelihood techniques, GAN suffers from unstable to train, sometimes leading to generators producing nonsensical outputs. In DCGAN, the authors gave some useful tricks on constructing a stable GAN. Based on the above reasons, we adopt some efficient structures of DCGAN.

b) Anti-GAN: Anti-GAN is proposed to solve the discriminating closer datasets. For 3D facial anti-spoofing application, the entire framework is shown in Fig.3. In the entire network, all the spatial pooling functions are forbidden, which benefits the generator and discriminator can learn the spatial downsampling by themselves. Behind every convolutional layer, we do batchnorm except the output layer of generator and the first layer of discriminator in order to erase the effect of poor initialization and sample oscillation. At last, we use ReLU activation in the generator aiming to cover the depth space quickly. Leaky rectified activation is used in the discriminator because it works well in [10].

c) Indentity classifier: For showing the advantages from our training mechinism, we add a simple 4-layer convolutional architecture to be a classifier. If Anti-GAN can generate a distant vector pair in high-dimentional space, they should be helpful to classifier task. In this paper, we take sigmoid cross entropy to be the loss of classifier:

$$z_j^{'} = \frac{exp(z_j)}{\sum_j^N exp(z_j)} \tag{7}$$

$$C_{loss} = -\frac{1}{N}\sum_j^N \left(p(I_j)log(z_j^{'})\right) \tag{8}$$

TABLE I
COMPARISION RESULTS WITH CONVENTIONAL CNN

	Texas Database	
Methods	3D facial anti-spoofing Top-1 accuracy	Face ID Top-1 accuracy
VGG net(phase 1)	0.729	0.667
VGG net(phase 2)	0.840	—
VGG net(phase 3)	0.476	0.000
GoogLeNet(phase 1)	0.337	0.047
GoogLeNet(phase 2)	0.767	—
GoogLeNet(phase 3)	0.476	0.000
Anti-GAN	0.913	0.892

where $z = C(t_{true})$, I is the label for all the people in training dataset. When we train Anti-GAN. we do not add C_{loss} into G_{loss}. So the classifier is trained separately. This method is a fair way to verify the feature extractor characteristic of generator in Anti-GAN.

V. METHODOLOGY COMPARISON

Other works such as [41,42,43], they focus on investigating facial feature points for single classification task in specific datasets. In [41], they used a two-stream CNN to improve the facial recognition accuracy on public datasets. [42] added RGB, IR and depth information to learn strong facial features and also proposed a new multi-modal architecture for facial recognition. [43] noticed the discriminative details of 3D moving faces and transfer the single -frame task to multi-frame task, which achieve better results on current existing datasets. Compared with these works, we choose 3D reconstructed faces as the main attack objects and demonstrate the different features among them in next experiments.

VI. EXPERIMENTS

a) Experimental setup: Anti-GAN is evaluated on Texas 3D Face Recognition Database. To construct the adversarial samples, we use PRNet to generate a 3D depth image from every RGB facial image. A same architecture of discriminator and classifier is used for different tasks, but they do not share parameters. For discriminator, we use it to construct the adversarial net and classifier is used to verify that the output of generator can be used to other tasks. For comparison, we choose two classical CNN to train for the same tasks: VGG Net and GoogLenet. From the point of function of Anti-GAN, it gives a different train mechainism to find a useful feature vector in high-dimension space. So we use simple convolutional structure for stressing on the 'minmax' problem. VGG Net and GooLenet stand for the conventional neural network which contain complex architectures without adversarial training. When we train these two CNNs on our experiments, we finetune the weight based on the original code provided by the authors.

b) Results and Discussions: For Anti-GAN, we train the classifier loss and adversarial net loss simultaneously, because they have no shared layers during training period. For VGGNet and GoogLeNet, we train the binary classification first on training set. Then we modified the output to give us vector of true and false for dataset as well as vectors for person identification. Both these vectors have separate cross entropy loss. According to our dataset, some person contains only one 3D image. So it is not easy to learn every identity feature for individual person.

c) Training details: For binary classification and multi-class classification, we adopt the Top-1 accuracy as metric in this paper. The results are listed in Tabel 1. In phase 1,we use the original weight pretrained on ImageNet[40].In phase 2, we start to set the layers to trainable with D_{loss} with only 3D anti spoofing loss. In phase 3, we train VGG Net and GoogLenet for 3D face ID and 3D facial anti-spoofing together.

d) Results analysis: As shown in Table1, in phase 1, VGG Net and GoogLenet can give reasonable accuracies, even without learning about dataset.. These can be because of number of reasons such as extensive training of network on large dataset such as imagenet. It indicates that both of their architectures are useful feature extractors. In phase 2, they can get 100% accuracy for 3D facial anti-spoofing on Texas Database. However, when the 3D face ID is added into the total loss function, VGG Net and GoogLenet fell into chaos eventually. In Anti-GAN, it always learns a stable direction to decrease the loss function. In test set, it can converge quickly to satisfy the classification tasks. In a sense, the adversarial mechanism in Anti-GAN shows a good feature extractor characteristic. At the same time, for 3D reconstruction algorithm , we find there is still a gap from true depth image. For these CNNs and Anti-GAN in this paper, they all can recognize the 3D adversarial samples. It is reasonable because when we observe the two datasets, 3D reconstruct faces are

not so accurate as a true 3D face at some details. They are much like to a generalized model. During the training process of G and D, we find that the loss of generator converges slowly due to that the D_{loss} is close to 1 after 36 epoches(100 epoches in total). So there is still room for generator to cover more complex tasks. With the output feature vectors from G, the classifier is can learn classification margins easily. In our experiments, Anti-GAN outperforms the conventional CNN.

VII. Conclusion

In this paper, we review some popular 3D facial reconstruction algorithms in recent years. Based on these methods, we analysis the potential threat and uncertainty for face ID security issue. For solving this problem, we construct some reconstruction face datasets by using existing real 3D face dataset as attack datasets. Aiming at differentiating these fake faces, we proposed a novel Anti-GAN which is used for discriminating two closer data distributions and extracting robust features for another classification task as a branch in this paper. We also add a distant loss on Gloss to achieve our goals. Anti-GAN in 3D facial anti-spoofing experiments indicates that it outperforms conventional CNN on extracting high-dimension meaningful features.

References

[1] Goodfellow, I. J., Shlens, J. and Szegedy, C., Explaining and harnessing adversarial examples' ICLR. arXiv preprint arXiv:1412.6572, 2014.

[2] Szegedy, Christian, Zaremba, Wojciech, Sutskever, Ilya, Bruna, Joan, Erhan, Dumitru, Goodfellow, Ian J and Fergus, Rob, Intriguing properties of neural networks, ICLR. abs/1312.6199, 2014.

[3] Goodfellow, I. J., Warde-Farley, D., Mirza, M., Courville, A. and Bengio, Y., Maxout networks. arXiv preprint arXiv:1302.4389, 2013.

[4] Jarrett, K, Kavukcuoglu, K, and LeCun, Y, What is the best multi-stage architecture for object recognition?, in 2009 IEEE 12th international conference on computer vision, 2009, pp.2146–2153.

[5] Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S.and Bengio, Y. , Generative adversarial nets, in Advances in neural information processing systems. 2014, pp.2672–2680.

[6] He, K., Zhang, X., Ren, S., and Sun, J., Spatial pyramid pooling in deep convolutional networks for visual recognition, in IEEE transactions on pattern analysis and machine intelligence, 2015, vol. 37, pp.1904–1916.

[7] He, K., Zhang, X., Ren, S., and Sun, J., Deep residual learning for image recognition, in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp.770–778.

[8] Long, J., Shelhamer, E., and Darrell, T., Fully convolutional networks for semantic segmentation, in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp.3431–3440.

[9] Redmon, J. and Farhadi, A., Yolov3: An incremental improvement, 2018 arXiv preprint arXiv:1804.02767.

[10] Radford, A., Metz, L., and Chintala, S., Unsupervised representation learning with deep convolutional generative adversarial networks, 2015, arXiv preprint arXiv:1511.06434.

[11] Dong, H., Neekhara, P., Wu, C., and Guo, Y., Unsupervised image-to-image translation with generative adversarial networks, 2017, arXiv preprint arXiv:1701.02676.

[12] Turkoglu, M. O., Thong, W., Spreeuwers, L., and Kicanaoglu, B., A Layer-Based Sequential Framework for Scene Generation with GANs, 2019, arXiv preprint arXiv:1902.00671.

[13] Mathieu, M., Couprie, C. and LeCun, Y., Deep multi-scale video prediction beyond mean square error, 2015, arXiv preprint arXiv:1511.05440.

[14] Zhu, J. Y., Zhang, R., Pathak, D., Darrell, T., Efros, A. A., Wang, O., and Shechtman, E., Toward multimodal image-to-image translation, in Advances in Neural Information Processing Systems, 2017, pp.465–476.

[15] Mirza, M., and Osindero, S., Conditional generative adversarial nets, 2014, arXiv preprint arXiv:1411.1784.

[16] Arjovsky, M., Chintala, S., and Bottou, L, Wasserstein gan, 2017, arXiv preprint arXiv:1701.07875.

[17] Cortes, C., and Vapnik, V., Support-vector networks, Machine learning, 1995, vol.20, pp.273–297.

[18] Girshick, R., Donahue, J., Darrell, T. and Malik, J., Rich feature hierarchies for accurate object detection and semantic segmentation, in Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp.580–587.

[19] LeCun, Y., Bottou, L., Bengio, Y. and Haffner, P., Gradient-based learning applied to document recognition, in Proceedings of the IEEE, 1998, vol.86, pp.2278–2324.

[20] Sutskever, I., Hinton, G. E., and Krizhevsky, A., Imagenet classification with deep convolutional neural networks, Advances in neural information processing systems, 2012, pp.1097–1105.

[21] Simonyan, K. and Zisserman, A., Very deep convolutional networks for large-scale image recognition, 2014, arXiv preprint arXiv:1409.1556.

[22] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D. and et al., Going deeper with convolutions, in Proceedings of the IEEE conference on computer vision and pattern recognition, pp.1–9.

[23] Han, S., Mao, H. and Dally, W. J., Deep compression: Compressing deep neural networks with pruning, trained quantization and huffman coding, 2015, arXiv preprint arXiv:1510.00149.

[24] Cohen, T. S., Geiger, M., Köhler, J. and Welling, M., 2018, Spherical cnns. arXiv preprint arXiv:1801.10130.

[25] Emil Mikhailov and Roman Trusov., How Adversarial Attacks Work, https://blog.ycombinator.com/how-adversarial-attacks-work, 2017.

[26] Papernot, N., McDaniel, P., Wu, X., Jha, S. and Swami, A., Distillation as a defense to adversarial perturbations against deep neural networks, in 2016 IEEE Symposium on Security and Privacy, 2016, pp.582–597.

[27] Yang, C., Chen, J., Su, N. and Su, G., Improving 3D face details based on normal map of hetero-source images, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops,2014, pp.9–14.

[28] Jourabloo, A., and Liu, X., Pose-invariant 3D face alignment, in Proceedings of the IEEE International Conference on Computer Vision, 2015, pp.3694–3702.

[29] Gupta, S., Markey, M. K. and Bovik, A. C., Anthropometric 3D face recognition, in International journal of computer vision, vol.90, pp.331–349.

[30] Gupta, S., Castleman, K. R., Markey, M. K., and Bovik, A. C., Texas 3D face recognition database, in 2010 IEEE Southwest Symposium on Image Analysis and Interpretation, 2010, pp.97-100.

[31] Blanz, V. and Vetter, T., A morphable model for the synthesis of 3D faces. in Siggraph, 1999, vol.99, pp.187–194.

[32] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[33] Huber, P., Feng, Z. H., Christmas, W., Kittler, J. and Rätsch, M., Fitting 3d morphable face models using local features, in 2015 IEEE international conference on image processing, 2015, pp. 1195–1199.

[34] Booth, J., Roussos, A., Zafeiriou, S., Ponniah, A. and Dunaway, D., A 3d morphable model learnt from 10,000 faces, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 5543-5552.

[35] Alp Guler, R., Trigeorgis, G., Antonakos, E., Snape, P., Zafeiriou, S. and Kokkinos, I., Densereg: Fully convolutional dense shape regression in-the-wild, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp.6799–6808.

[36] Yu, R., Saito, S., Li, H., Ceylan, D. and Li, H., Learning dense facial correspondences in unconstrained images, in Proceedings of the IEEE International Conference on Computer Vision, 2017, pp.4723-4732.

[37] Bas, A., Huber, P., Smith, W. A., Awais, M. and Kittler, J., 3D morphable models as spatial transformer networks, in Proceedings of the IEEE International Conference on Computer Vision, 2017, pp.904–912.

[38] Tewari, A., Zollhofer, M., Kim, H., Garrido, P., Bernard, F., Perez, P. and Theobalt, C., Mofa: Model-based deep convolutional face autoencoder for unsupervised monocular reconstruction, in Proceedings of the IEEE International Conference on Computer Vision, 2017, pp.1274–1283.

[39] Feng, Y., Wu, F., Shao, X., Wang, Y. and Zhou, X., Joint 3d face reconstruction and dense alignment with position map regression network, in Proceedings of the European Conference on Computer Vision, 2018, pp.534–551.

[40] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S. and et al., Imagenet large scale visual recognition challenge, international journal of computer vision, vol.115, pp.211–252.

[41] Atoum, Y. , Liu, Y. , Jourabloo, A. and Liu, X., Face anti-spoofing using patch and depth-based CNNs, in The International Joint Conference on Biometrics, 2017.

[42] Parkin, Aleksandr, and O. Grinchuk, Recognizing Multi-Modal Face Spoofing With Face Recognition Networks, 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2019.

[43] Zezheng Wang., Zitong Yu., Chenxu Zhao, Xiangyu Zhu, Yunxiao Qin, Qiusheng Zhou and et al., Deep Spatial Gradient and Temporal Depth Learning for Face Anti-spoofing, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2020.