Pattern Recognition 134 (2023) 109065

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/patcog

Meta-learning-based adversarial training for deep 3D face recognition on point clouds

Cuican Yu[a], Zihui Zhang[b], Huibin Li[a,*], Jian Sun[a], Zongben Xu[a]

[a] School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, Shaanxi, China
[b] Department of Computing, Hong Kong Polytechnic University, Hong Kong, China

ARTICLE INFO

Article history:
Received 19 April 2022
Revised 18 June 2022
Accepted 20 September 2022
Available online 23 September 2022

Keywords:
Deep 3D face recognition
Point clouds
Adversarial samples
Meta-learning

ABSTRACT

Recently, deep face recognition using 2D face images has made great advances mainly due to the readily available large-scale face data. However, deep face recognition using 3D face scans, especially on point clouds, has been far from fully explored. In this paper, we propose a novel meta-learning-based adversarial training (MLAT) algorithm for deep 3D face recognition (3DFR) on point clouds. It consists of two alternate modules: adversarial sample generating for 3D face data augmentation and meta-learning-based deep network training. In the first module, adversarial samples of given 3D face scans are dynamically generated based on current deep 3DFR model. In the second module, a meta-learning framework is designed to avoid the performance decrease caused by the generated adversarial samples. Overall, MLAT algorithm combines the adversarial sample generating and meta-learning-based network training in a uniform framework, in which adversarial samples and network parameters are optimized alternately. Thus, it can continuously generate diverse and suitable adversarial samples, and then the meta-learning framework can further improve the accuracy of 3DFR model. Comprehensive experimental results show that the proposed approach consistently achieves competitive rank-one recognition accuracies on the BU-3DFE (100%), Bosphorus (99.78%), BU-4DFE (98.02%) and FRGC v2 (98.01%) database, and thereby substantiate its superiority.

1. Introduction

Face recognition, with the characteristic of human perception and non-intrusiveness, has been one of the most popular and promising biometric modalities and has many application scenarios, such as access control, public security, healthcare, etc. Deep 2D face recognition has made remarkable achievements in both performance and robustness, mainly attributed to the readily available large-scale 2D face data. For example, the FaceNet [1] was trained on 200 million face images of 8 million identities, and 2.6 million face images were utilized to train the VGG-Face [2].

On the other side, with the fast development of 3D scanning techniques, 3D face recognition (3DFR) has attracted much attention in the past two decades. Compared with 2D face images, 3D face scans contain real 3D geometric shape information of human faces and thus have the potential to achieve more discriminative facial representations, more accurate face recognition performances, and more powerful face anti-spoofing capabilities. Traditional 3DFR methods mainly focus on extracting hand-crafted local or global facial surface geometry and shape features using facial curves [3] and iso-geodesic stripes [4], or encoding facial depth maps, normal maps, and shape index maps using wavelet coefficients [5], multi-scale local binary pattern (MS-LBP) [6], and so on. Although high performances can be achieved, these methods rely on the procedures of 3D face pre-processing, land-marking, and registration [7], which limits their timeliness and scalability. Recently, deep-learning-based methods light up the directions of 3DFR. However, collecting millions of 3D face scans from identities with great diversity is an incredibly difficult task, and has not yet been achieved in the academic community. That is the reason why most existing deep-learning-based 3DFR methods focus on using different kinds of 3D face data augmentation strategies for network training.

Existing 3D face data augmentation methods can be mainly divided into three categories: transformation-based, reconstruction-based, and synthesis-based methods. Transformation-based methods usually apply various kinds of transformations (e.g., rotation, scaling, masking, reflection, adding noises, down-sampling) to 3D face scans to generate new samples. For example, Kim et al. [8] proposed to apply random rigid transformations (i.e., yaw, pitch, row rotations, and translation) to augment 3D face scans

* Corresponding author.
E-mail address: huibinli@xjtu.edu.cn (H. Li).

https://doi.org/10.1016/j.patcog.2022.109065
0031-3203/

with pose variations. Cai et al. [9] also adopted random affine transformation (including rotation, shearing, and zooming) on 3D face scans, and twisting, and horizontal flipping on depth maps of 3D face scans for data augmentation. In general, transformation-based methods are simple, easy to implement, and have been widely used. However, this kind of data augmentation strategy can not introduce new facial expressions or identities. Reconstructing 3D face scans from single or multiple 2D face images is another promising data augmentation strategy. Although reconstruction-based methods can generate a large amount of 3D face scans from massive 2D face data, their effectiveness is heavily impacted by the reconstruction accuracy. Synthesis-based methods can generate 3D face scans with new identities or facial expressions. For example, Kim et al. [8] proposed to synthesize a large number of 3D face scans with different expressions based on the 3D Morphable Model (3DMM). Similarly, Zhang et al. [10], Yu et al. [11] proposed to generate 3D face scans using Gaussian Process Morphable Model (GPMM), i.e., a specific 3DMM, in which shape variation was modeled as a normal distribution and 3D face scans were generated by changing shape and expression coefficients. Nevertheless, due to the linearity characteristic of 3DMM, most of the generated 3D face samples are similar and thus have limited discrimination power when used for 3DFR. Similarly, Zhao et al. [12] proposed the idea of 3D-Aided Dual-Agent GANs for unconstrained face recognition. In particular, they employed an off-the-shelf 3D face model as a simulator to generate profile face images with varying poses. Gilani and Mian [13] proposed to generate 3D face scans by interpolating between facial identities and expressions, but this method relies on sophisticated 3D face dense correspondence, and meanwhile it can not guarantee the closure property of identities, which may cause serious label noise. Overall, all these existing 3D face data augmentation strategies are human-designed, and augmented 3D face scans are generated before network training and will not be updated during training. Thus, we can conclude that they are off-line or model-independent data augmentation methods.

In this paper, we propose an online and model-dependent 3D face augmentation strategy for deep 3DFR. Inspired by the success of adversarial training [14] and meta-learning [15], we propose a novel MLAT algorithm for 3D face augmentation and deep 3DFR on point clouds. Specifically, we propose to alternately generate adversarial samples of 3D face point clouds and train the deep 3DFR model with these adversarial samples. The adversarial 3D face data augmentation is based on current deep 3DFR model, and the generated adversarial samples will be dynamically updated with the update of the deep 3DFR model. Moreover, we introduce a meta-learning framework to train the deep 3DFR model using both adversarial and clean samples (i.e., the original 3D face point clouds). As shown in Fig. 1, given a set of 3D face point clouds, MLAT algorithm can generate their corresponding adversarial samples through the gradient ascent optimization algorithm. These adversarial samples together with their corresponding clean samples are then used to boost the recognition accuracy of the deep 3DFR model. Since adversarial and clean samples obey different data distributions, directly adding adversarial samples to the training set may hurt the model performance [16]. Thus, we propose to train the deep 3DFR model under a meta-learning framework. The key idea is that recognition ability learned from adversarial samples should be conducive to recognize clean samples, and further improve both the robustness and accuracy of the final deep 3DFR model. The main contributions of this paper can be summarized as follows:

- We propose the MLAT algorithm for deep 3DFR, in which the model-dependent online 3D face data augmentation strategy is designed based on the principle of adversarial training. This adaptive strategy can alternately and dynamically optimize adversarial samples and network parameters. To the best of our knowledge, this is the first work that introduces the idea of adversarial training to deep 3DFR.
- Given that adversarial samples and clean samples obey different distributions, we propose a meta-learning framework in MLAT algorithm to train the deep 3DFR model, where adversarial samples can be applied to improve the recognition accuracy of deep 3DFR model.
- Comprehensive experimental results demonstrate that our proposed approach consistently achieves competitive recognition results on public databases with a small set of training data. Moreover, our approach also improves the robustness of deep 3DFR model to the variations of facial expressions, hair noises, etc.

The remainder of this paper is organized as follows. Related works are presented in Section 2. In Section 3, we introduce the proposed MLAT algorithm for deep 3DFR. Section 4 shows our experimental results and analysis. Section 5 concludes the whole paper.

2. Related work

2.1. Deep learning on point clouds

PointNet [17] is the first deep neural network operating directly on point clouds for 3D shape analysis. Given a 3D point cloud, it computes deep features of each 3D point independently, and then applies a symmetric function to aggregate these point-wise deep features to a global representation. To capture local structures of 3D shapes, PointNet++ [18] extends PointNet using a hierarchical structure, in which deep features are extracted starting from small local regions and gradually extending to large regions. Instead of performing point-wise operations, recent methods proposed to construct convolution across the local neighborhood of each point to improve feature representation ability. For instance, the EdgeConv operation in DGCNN [19] is proposed to act on graphs dynamically computed using the K-Nearest Neighbors in feature space of the given 3D shape, rather than original coordinate space. Graphs in feature space enable EdgeConv to capture semantic characteristics over potentially long distances in coordinate space.

2.2. Adversarial training

Adversarial samples [20] refer to input samples deliberately adding slight perturbations, resulting in the model giving a wrong prediction with high confidence. Further, adversarial training is a kind of data augmentation strategy generating adversarial samples to improve the robustness of deep models against adversarial perturbations or attacks. For example, Madry et al. [21] have verified that deep neural networks can defense against adversarial attacks through reliable adversarial training. To guarantee the performance under adversarial perturbations, Sinha et al. [22] provided a principled method for efficiently guaranteeing distributional robustness with adversarial training. Recently, adversarial training has also been applied to the problem of domain generalization. In M-ADA [14], adversarial samples are generated as fictitious domains to improve the image classification performance of deep models in domain generalization. However, it is often thought that adversarial samples impair the objective classification accuracy on clean images [23]. Nowadays, Xie et al. [16] claimed that adversarial and clean samples are drawn from two different distributions, and thus they proposed to utilize Auxiliary Batch Normalization (AuxBN) during adversarial training to improve the classification accuracy on clean images. In conclusion, adversarial training is a promising

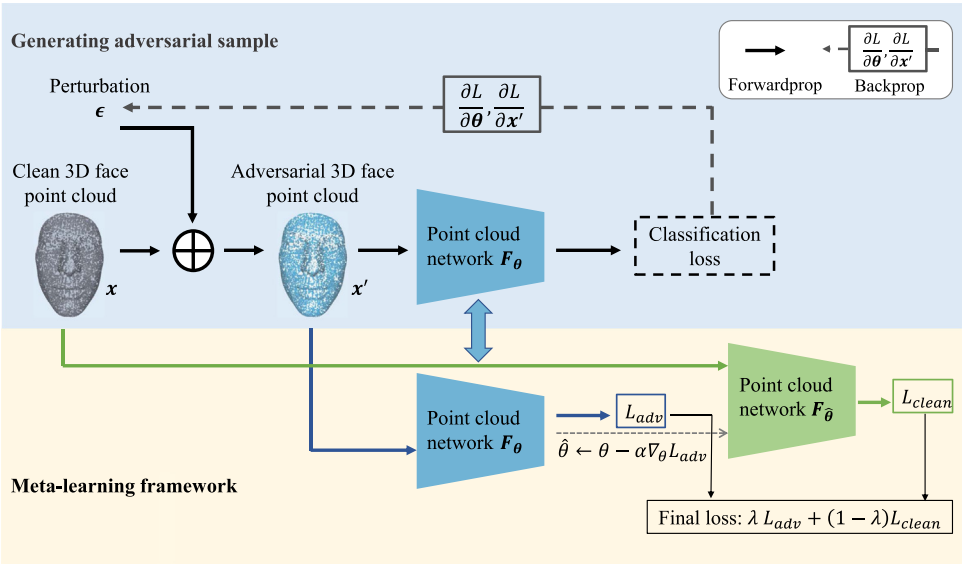

Fig. 1. The framework of the proposed meta-learning adversarial training (MLAT) algorithm for 3DFR. Given a set of 3D face point clouds and a deep point cloud network, the MLAT algorithm consists of two alternate modules: adversarial sample generating for 3D face data augmentation and meta-learning-based deep network training. With clean 3D face point clouds $\boldsymbol{x}$ as input, adversarial 3D face point clouds $\boldsymbol{x}'$ are iteratively generated by optimizing the perturbation $\boldsymbol{\epsilon}$ under the 3DFR task with the deep point cloud network $\boldsymbol{F_\theta}$. Then the clean samples $\boldsymbol{x}$ and the adversarial samples $\boldsymbol{x}'$ are used to train the deep point cloud network via a meta-training framework. In meta-train, L_{adv} is computed on adversarial 3D face point clouds, and the model parameters $\boldsymbol{\theta}$ is updated to $\hat{\boldsymbol{\theta}}$ by one or more gradient steps. In meta-test, L_{clean} on clean 3D face point clouds is computed with model parameters $\hat{\boldsymbol{\theta}}$. In meta-update, we update $\boldsymbol{\theta}$ by the gradients calculated from a combined loss of L_{adv} and L_{clean}, where meta-train and meta-test are optimized simultaneously. Finally, the deep point cloud network $\boldsymbol{F_\theta}$ has been well trained for 3DFR on point clouds.

data augmentation strategy to improve both the robustness and accuracy of deep model.

In this paper, considering that limited training data may result in over-fitting in deep 3DFR, we propose to adaptively generate adversarial samples for 3D face augmentation. In view of the potential accuracy decrease on clean 3D face scans caused by the adversarial samples, we not only adopt the strategy of AuxBN [16], but also design a meta-learning framework to train deep 3DFR models.

2.3. 3D face recognition

There exist several surveys [24,25] of 3D face recognition. Traditional 3DFR methods can be roughly categorized into holistic methods and local feature-based methods. Holistic methods [26] generally compute the dense registration errors or match the parameters of a statistical Morphable Model to achieve the distance between 3D face scans. Yu et al. [27] proposed sparse 3D directional vertices to represent and match 3D surfaces by fewer sparsely. Compared with holistic methods, local feature-based methods [28] have more advantages in facing challenges of posture, facial expression and occlusion. For example, Lei et al. [29] proposed local geometrical signatures (facial Angular Radial Signature) to effectively represent local regions of 3D face surfaces. In [30], the authors proposed to represent a 3D face with a set of Multiple Triangle Statistics, which is robust to facial expressions, partial data, and pose variations. Elaiwat et al. [31] integrated curvelet elements of different orientations to obtain rotation invariant features of 3D face scans. Nevertheless, most of these traditional methods are time-consuming owing to the complex handcrafted shape descriptors and registration pipeline.

Deep-learning-based 3DFR methods mainly adopt deep convolutional neural networks (CNNs) to extract deep features from facial geometric maps projected from 3D face scans. For example, Li et al. [32] projected 3D face scans into normal images, and used the pre-trained VGG-Face [2] to extract deep features, which partly demonstrates the effectiveness of deep CNNs for 3DFR.

Similar to 2D face recognition methods, most deep 3DFR methods adopt 3D face data augmentation. For instance, Kim et al. [8] proposed a 3D face augmentation technique to synthesize more 3D face scans with different expressions from a single 3D face scan using the 3DMM [33], and used their augmented dataset (i.e., totally 123,325 depth maps) to fine-tune the pre-trained VGG-Face [2] for 3DFR. In [13], Gilani et al. synthesized 3.1 million 3D face scans of 100 thousand identities by interpolation, and projected 3D face scans into depth, azimuth and elevation maps for deep feature extraction. Cai et al. [9] adopted multiple data augmentation strategies to enlarge the training set, and they proposed to learn deep features from local patches of facial depth maps to achieve state-of-the-art recognition results.

Recently, several methods have moved deep 3DFR beyond facial depth maps to 3D point clouds. For example, Bhople et al. [34,35] proposed to use PointNet to extract global deep features of 3D face scans over point clouds and then combined them with metric learning for 3DFR. Similarly, Jiang et al. [36] designed a siamese point cloud network and proposed a pair selection strategy to choose positive and negative pairs for network training of deep 3DFR. Zhang et al. [10], Yu et al. [11] proposed a point cloud network similar to PointNet++, in which a curvature-aware point sampling was designed specifically for deep 3DFR. The main challenge in implementing 3DFR on point clouds is the scarcity of 3D face data. Therefore, we propose to generate adversarial samples to train the point cloud network in our MLAT algorithm.

3. Meta-learning-based adversarial training for deep 3DFR

As illustrated in Fig. 1, the proposed meta-learning-based adversarial training (MLAT) method is composed of two alternately updating modules, i.e., adversarial sample generating and meta-learning-based network training. Through the above alternative modules, the optimized deep point cloud network for 3D face recognition (3DFR) is obtained.

3.1. Deep point cloud classification network

As mentioned in Section 2, different from most existing deep 3DFR methods, we use the deep point cloud network working directly on unorganized 3D face point clouds. In particular, each 3D face point cloud $\boldsymbol{x} \in \mathbb{R}^{N\times(3+d)}$ with N 3D points $\{x_i | i = 1, 2, \ldots, N\}$ is represented by point-wise 3D coordinate concatenated with an additional d-dimensional feature. For example, the normal vector is usually set as an additional input feature ($d = 3$). With $\boldsymbol{x}$ as input, the deep point cloud network $\boldsymbol{F_\theta} : \mathbb{R}^{N\times(3+d)} \rightarrow \mathbb{R}^C$ is defined as

$$\boldsymbol{F_\theta}(\boldsymbol{x}) = f(\text{MaxPool}\{g(\boldsymbol{x})\}), \tag{1}$$

where f is a classifier composed of two Fully-connected (FC) layers and a softmax layer, MaxPool is a channel-wise max-pooling layer for feature aggregation, g is a convolutional (1×1 or edge) network, and $\boldsymbol{\theta}$ denotes parameters in f and g. We respectively take the widely used PointNet [17] and PointNet++ [18] as the backbone network g for feature extraction, because their advantages of lightweight memory and computational efficiency. PointNet [17] used in this work consists of five 1×1 convolutional layers and each one is followed by a batch normalization layer and a nonlinear activation ReLU. PointNet++ [18] consists of three set abstraction modules, and each one is composed of a Sampling layer, a Grouping layer and a PointNet layer.

3.2. Adversarial sample generating

Generating adversarial sample is a task-oriented problem. For the task of 3DFR with deep point cloud network $\boldsymbol{F_\theta}$, the objective function can be formulated as:

$$\arg\min_{\boldsymbol{\theta}} \mathbb{E}_{(\boldsymbol{x},y)\sim P_{data}}[L(\boldsymbol{\theta}; \boldsymbol{x}, y)], \tag{2}$$

where $\boldsymbol{x}$ is the input 3D face point cloud with one-hot identity label y, P_{data} is the underlying data distribution, and $\boldsymbol{\theta}$ represents the network parameters. Here we chose the widely used cross-entropy as classification loss:

$$L(\boldsymbol{\theta}; \boldsymbol{x}, y) = -\sum_{i=1}^{C} y_i \log(\hat{y}_i), \tag{3}$$

where y_i represents ith dimension of the one-hot label, $\hat{y}_i$ denotes the ith dimension of the prediction of $\boldsymbol{F_\theta}$, and C is the number of categories.

An adversarial sample of a given 3D face point cloud $\boldsymbol{x}$ can be generated by adding a targeted perturbation variable, such that the deep 3DFR model $\boldsymbol{F_\theta}$ gives an incorrect prediction to the adversarial sample. More concretely, we define the adversarial sample of a given 3D face point cloud as $\boldsymbol{x'} = \boldsymbol{x} + \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon} \in \mathbb{R}^{N\times(3+d)}$ is a point-wise perturbation.

In practice, adversarial samples of 3D face point clouds can be generated by maximizing aforementioned objective function to fool the 3DFR model, that is,

$$\arg\max_{\boldsymbol{\epsilon}\in[-\delta,\delta]} \mathbb{E}_{(\boldsymbol{x},y)\sim P_{data}}[L(\boldsymbol{\theta}; \boldsymbol{x} + \boldsymbol{\epsilon}, y)], \tag{4}$$

where $\boldsymbol{\epsilon}$ is the adversarial perturbation with a range constraint $[-\delta, \delta]$. In this paper, the Fast Gradient Sign Method (FGSM) [20] is used to iteratively generate adversarial 3D face point clouds as follow,

$$\boldsymbol{x}'_k = \boldsymbol{x}'_{k-1} + \eta \cdot \text{sign}(\nabla_{\boldsymbol{\epsilon}} L(\boldsymbol{\theta}; \boldsymbol{x}'_{k-1} + \boldsymbol{\epsilon}, y)), \tag{5}$$

where $\boldsymbol{x}'_k$ and $\boldsymbol{x}'_{k-1}$ are adversarial examples with k and $k-1$ iterations, $\boldsymbol{x}'_0 = \boldsymbol{x}$ is the input clean sample, η is the step size of updating the adversarial perturbation, y is the one-hot label of $\boldsymbol{x}$, ∇ is the gradient operator, and $\boldsymbol{\theta}$ represents model parameters.

3.3. Meta-learning framework for network training

Once adversarial samples generated, how effectively combining the information that comes from adversarial and clean samples becomes a key problem. To improve both the robustness and accuracy of deep 3DFR model, we introduce a meta-learning framework for network training. That is, train a deep 3DFR model using generated adversarial 3D face point clouds in the meta-train phase, and then use the model to recognize clean 3D face point clouds in the meta-test phase. The model parameters are updated using the loss from both meta-train and meta-test phases. Formally, the meta-learning framework can be summarized as the following steps of meta-train, meta-test, and meta-update.

3.3.1. Meta-train

In the meta-train phase, the classification loss is computed on adversarial samples, and the model parameters are updated by stochastic gradient descent algorithm with a learning rate of α, that is,

$$L_{adv} = \mathbb{E}_{(\boldsymbol{x}'_k,y)\sim P'_{data}}[L(\boldsymbol{\theta}; \boldsymbol{x}'_k, y)], \tag{6}$$

$$\hat{\boldsymbol{\theta}} \leftarrow \boldsymbol{\theta} - \alpha \nabla_{\boldsymbol{\theta}} L_{adv}, \tag{7}$$

where $\boldsymbol{x}'_k$ is the adversarial sample of $\boldsymbol{x}_k$, P'_{data} is the data distribution of adversarial data, y is one-hot label of $\boldsymbol{x}$, $\boldsymbol{\theta}$ represents model parameters, α is the step-size in meta-train, and ∇ is the gradient operator.

3.3.2. Meta-test

In the meta-test phase, we calculate the classification loss on the clean 3D face point clouds with parameters $\hat{\boldsymbol{\theta}}$, that is:

$$L_{clean} = \mathbb{E}_{(\boldsymbol{x},y)\sim P_{data}}[L(\hat{\boldsymbol{\theta}}; \boldsymbol{x}, y)], \tag{8}$$

where $\boldsymbol{x}$ is the clean sample with underlying data distribution P_{data}, y is the one-hot label of $\boldsymbol{x}$, $\hat{\boldsymbol{\theta}}$ is model parameters updated in meta-train.

3.3.3. Meta-update

In the meta-update phase, to improve the model performance for both adversarial and clean samples, classification loss of the meta-train and meta-test are combined to optimize parameters of the 3DFR model:

$$\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} - \beta \nabla_{\boldsymbol{\theta}}[\lambda L_{adv} + (1 - \lambda) L_{clean}], \tag{9}$$

where β is the learning rate, ∇ is the gradient operator, and λ is trade-off between meta-train and meta-test.

3.4. 3DFR based on MLAT

To further enhance the performance on clean data, we introduce the Auxiliary Batch Normalization (AuxBN) [16] to our MLAT algorithm, which keeps separate BN layers to deep features that belong to different distributions. Specifically, the clean samples pass through the main BN layers and the adversarial samples pass through the AuxBN layers. The whole pipeline of MLAT algorithm for deep 3DFR is be summarized in Algorithm 1.

In this algorithm, for each clean mini-batch, we first attack the network using Eq. (5) and AuxBN layers to generate its corresponding adversarial mini-batch. Then, we use the adversarial mini-batch and clean mini-batch to train the network under meta-learning framework, applied with different BN layers. Finally, the network parameters are updated according to the total classification loss computed on the adversarial mini-batch and clean mini-batch. In the above meta-learning framework, model parameters

Algorithm 1: Pseudo code of MLAT algorithm for deep 3DFR.

Input: A training database of 3D face point clouds,a deep point cloud network with parameters $\boldsymbol{\theta}$,hyper-parameters α, β, λ, η and K.

Output: An optimized deep point cloud network $\boldsymbol{F_\theta}$ for 3DFR.

for A mini-batch of clean 3D face point clouds sampled from the database **do**
- Initialize the adversarial perturbation $\boldsymbol{\epsilon} \leftarrow \mathbf{0}$.
- **for** $k = 1, 2, \ldots, K$ **do**
  - Generate a mini-batch of adversarial samples using Eq.(5).
  - **Meta-train:**Update the network parameters $\boldsymbol{\theta}$ to $\hat{\boldsymbol{\theta}}$ using Eq.(7)over the mini-batch of adversarial samples with the AuxBN layers.
  - **Meta-test:**Calculate classification loss of $\boldsymbol{F}_{\hat{\boldsymbol{\theta}}}$ over the mini-batchof clean 3D face point clouds using Eq.(8) and the main BN layers.
  - **Meta-update:**Update the model $\boldsymbol{F_\theta}$ with the combination loss ofboth meta-train and meta-test using Eq.(9).
- **end**

end

$\boldsymbol{\theta}$ is optimized to $\hat{\boldsymbol{\theta}}$ with the adversarial mini-batch in the meta-train phase, which means the direction of the optimization is beneficial to recognize adversarial 3D face point clouds. With parameters $\hat{\boldsymbol{\theta}}$, classification loss on the corresponding clean mini-batch is computed in the meta-test phase. Thus, the model is required to perform well on clean samples after being updated on adversarial samples.

It is worth noting that in Algorithm 1, generating adversarial samples and meta-learning-based network training are carried out alternately. After each iteration of network training, the optimized network is used to generating a new adversarial mini-batch. Since the deep model is updated constantly, adversarial samples are generated adaptively to current deep model and input samples. Thus, the performance of deep 3DFR model can be improved gradually.

Once the MLAT algorithm has been done, the deep 3DFR model with the main BN layers can be used to extract deep features of 3D face point clouds. In particular, we extract deep features of each probe and gallery 3D face scans by the final deep 3DFR model, and then the Euclidean distances between deep features of each probe 3D face scan and gallery scans can be computed and then used for 3D face identification.

4. Experiments

We introduce 3D face datasets and implementation details in Section 4.1 and Section 4.2 respectively. In Section 4.3, we evaluate the effectiveness of adversarial 3D face data augmentation strategy, the meta-learning-based network training framework, and their combination for 3D face identification. Hyper-parameters tuning is also conducted. In Section 4.4, we compare the proposed method with the prior state-of-the-art methods. The identification results are evaluated by rank-one score.

4.1. 3D face databases and evaluation protocols

BU-3DFE. This dataset [37] includes 100 subjects (56 females, 44 males) with a variety of racial ancestries (e.g., White, Black, East-Asian), and their ages ranging from 18 to 70 years old. Each subject has one sample of the neutral expression, and 24 samples of six prototypical expressions (i.e., happiness, disgust, fear, anger, surprise and sadness). Among them, each expression contains four levels of intensity: level 1 and level 2 are low intensity, level 3 and level 4 are high intensity. As a result, this database consists of totally 2500 3D face scans. Because of the diversity of expression types and variability of expression intensities, the BU-3DFE dataset can be utilized to evaluate the expression-robustness of 3DFR methods. We take the standard experimental protocol, in which the neutral expression scan of each individual is selected to form the gallery (100 scans), and the rest 3D face scans are selected as the probes (2400 scans).

Bosphorus. This dataset [38] has been widely used for 3DFR, which contains a total of 4666 3D facial scans of 105 individuals (60 males and 45 females). These scans cover rich pose changes (e.g., yaw and pitch rotations), various facial expressions, and typical occlusions. In this paper, we adopt a subset containing 2902 scans with different expressions (basic emotions and action units). The first neutral scan of each subject forms the gallery (105 scans), and the remaining scans are used as the probes (2797 scans).

FRGC v2. This dataset [39] is the largest public high-quality 3D face dataset that has been widely used for 3DFR in recent years. It consists of 4007 3D face scans of 466 subjects with different facial expressions (1642 samples). These 3D face scans were collected using the Minolta laser sensors under controlled lighting conditions, and over different sets of sessions: Spring 2003 subset, Fall 2003 subset, and Spring 2004 subset. For evaluation, the standard experimental protocol is that the first scan of each subject forms the gallery (466 scans), and the remaining 3541 scans are used as the probes.

BU-4DFE. This dataset [40] contains 101 subjects (58 females, 43 males) with expressions of anger, happiness, fear, disgust, sadness, surprise, and neutral. Each expression is represented by a 3D sequence with a length of around 100 frames. Each sequence is supposed to begin with neutral expression, then change to the labeled expression with strong expression intensity, finally return to neutral expression. To compare fairly, we adopt the same strategy as [13] that retains five frames equally spaced apart for each sequence. Among retained frames, the first frame per identity is selected to form the gallery (101 scans) and others as the probes (2929 scans).

XJTU-3DFace. We collected a 3D face dataset, named as XJTU-3DFace. It contains 2980 high-quality 3D face scans of 535 individuals, and all the 3D faces are scanned by the Facego Pro, a 3D scanning equipment developed by Chishine3D. In this dataset, each subject has four kinds of expressions, including anger, happiness, surprise, and neutral. In addition, each person has five or six 3D face scans. All the 3D face scans of the XJTU-3DFace are used as training data in Section 4.4.

Figure 2 shows several examples of our self-collected XJTU-3DFace dataset, and Fig. 3 shows several examples of the BU-3DFE, Bosphorus, FRGC v2, and BU-4DFE datasets.

4.2. Implementation details

Each 3D face scan is firstly cropped within a radius of 90mm according to the nose tip to preserve the main face region, and then is aligned to a 3D face template using the rigid-ICP algorithm [41], similar to Kim et al. [8]. The rigid-ICP algorithm can also be replaced by the recently proposed 3D facial head pose estimation method [42]. Each cropped and aligned 3D face scan is then downsampled to 6000 points by Farthest Point Sampling [18] and normalized to be zero mean within a unit ball.

We use the Adam as the optimizer with a learning rate of $1e-3$. The batch size is 24, and the number of epochs is 500. Hyper-parameters α and β are set to $1e-3$, and the hyper-parameter tuning of λ and K is explored in Section 4.3.4. The models were trained on two GeForce RTX 2080 Ti. After training, the last two

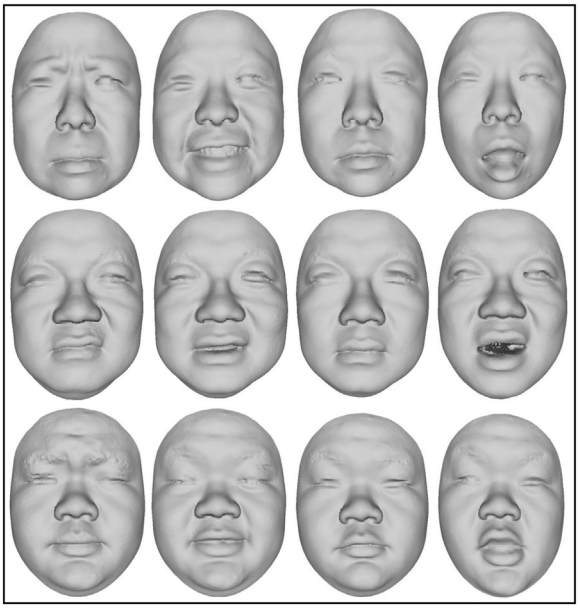

Fig. 2. Illustration of the 3D face scans sampled from the datasets of XJTU-3DFace.

layers (FC layer and soft-max layer) of the 3DFR model are removed, and the remaining layers with frozen parameters are used to extract deep features of 3D face point clouds for evaluation. During evaluation, to avoid the contingency of test results caused by random initialization, we run our method three times with different random seeds, and report the mean accuracy.

4.3. Ablation study

In this section, we merge the BU-3DFE and Bosphorus datasets as the training set and the FRGC v2 dataset as the test set. Then conduct experiments with the backbones of PointNet and PointNet++, with the input modality of Points and Points+Normals respectively.

4.3.1. Effectiveness of the adversarial 3D face samples

Table 1 compares the rank-one scores achieved by the adversarial training (AT) based models and the baseline models (PointNet and PointNet++) on the FRGC v2 database. Specifically, the deep networks of AT-PointNet and AT-PointNet++ are trained across the clean mini-batch and the adversarial mini-batch of 3D face samples alternately. While the baseline models are trained using only the clean 3D face samples. We can conclude that: 1) With different backbones (PointNet and PointNet++) and different input modalities (Points and Points+Normals), AT-based models can generally perform better than the baseline models. 2) Furthermore, additional input modality of normal vectors can generally boost the rank-one scores of both baseline models and AT-based models. These results indicate the effectiveness of adversarial 3D face samples for deep 3DFR.

Table 1
The effectiveness of adversarial 3D face samples on the FRGC v2 database.

Method	Modality	Accuracy (%)
PointNet	Points	77.53
AT-PointNet	Points	81.96 (↑ **4.43**)
PointNet+	Points	92.03
AT-PointNet+	Points	93.33 (↑ **1.30**)
PointNet	Points+Normals	91.60
AT-PointNet	Points+Normals	93.59 (↑ **1.99**)
PointNet+	Points+Normals	94.10
AT-PointNet+	Points+Normals	95.66 (↑ **1.56**)

Table 2
The effectiveness of meta-learning framework on the FRGC v2 database.

Method	Modality	Accuracy (%)
AT-PointNet	Points	81.96
MLAT-PointNet	Points	83.83 (↑ **1.87**)
AT-PointNet+	Points	93.33
MLAT-PointNet+	Points	94.67 (↑ **1.34**)
AT-PointNet	Points+Normals	93.59
MLAT-PointNet	Points+Normals	94.84 (↑ **1.25**)
AT-PointNet+	Points+Normals	95.66
MLAT-PointNet+	Points+Normals	96.96 (↑ **1.30**)

In Fig. 4, we plot the accuracy curves of AT-based models and baseline models on the training set and test set respectively. We can find that adversarial 3D face samples can significantly reduce the accuracy gaps of deep models achieved on the training set and test set, which demonstrates the phenomenon of over-fitting can be alleviated by the adversarial 3D face data augmentation.

4.3.2. Effectiveness of the meta-learning framework

Compared with AT, MLAT expects that the network updated with adversarial samples also performs well on clean samples. From Table 2, with different backbones (PointNet and PointNet++) and different input modalities (Points, Points+Normals), MLAT-based models can generally outperform the AT-based ones with significant margins. These improvements confirm that the proposed meta-learning framework also plays an important role to improve the accuracy of deep 3DFR. In the following experiments, we use both points (i.e., coordinates) and normal vectors of the 3D face scans as the input modality.

The purpose of introducing the meta-learning framework is to make use of the adversarial samples to improve recognition results

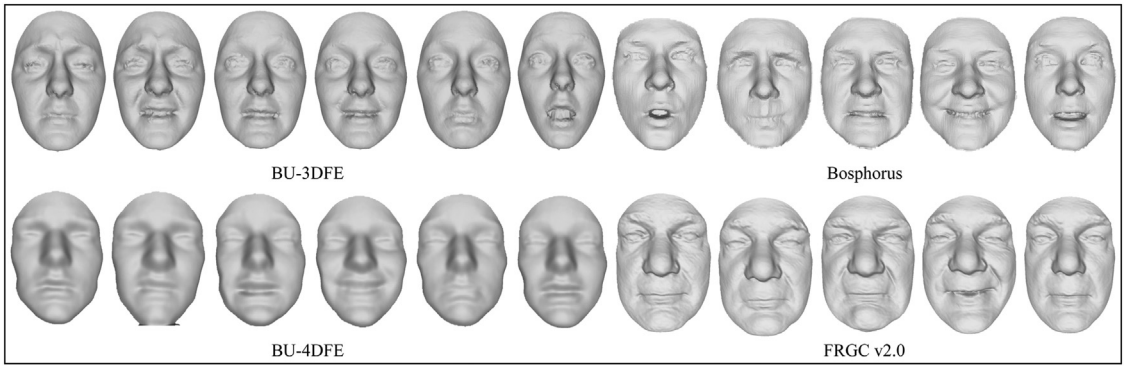

Fig. 3. Illustration of the 3D face scans sampled from the datasets of BU-3DFE, Bosphorus, BU-4DFE, and FRGC v2.

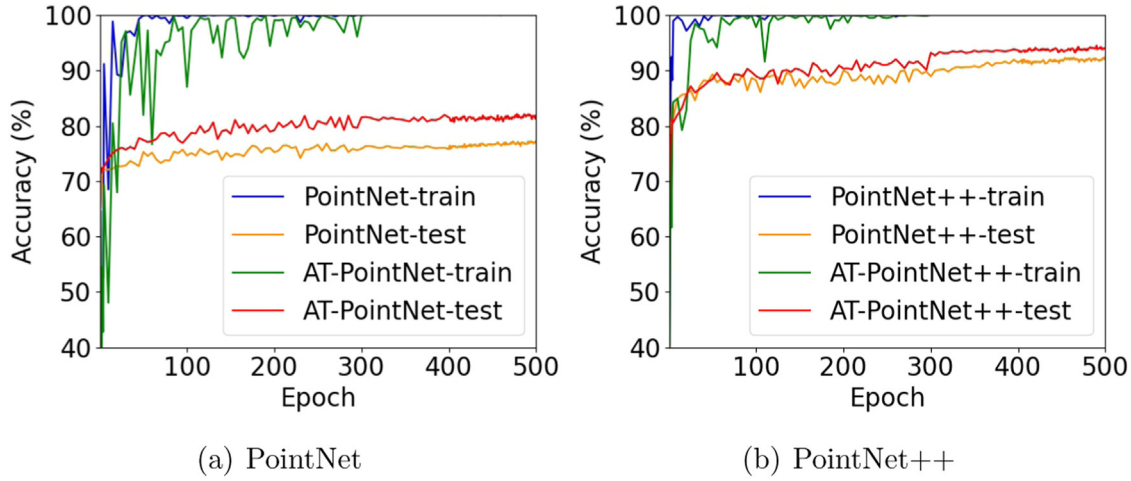

Fig. 4. Illustration of the accuracy curves on the training and test sets of PointNet and PointNet++ backbones with and without the adversarial 3D face samples. We can easily find that adding adversarial samples can relief over-fitting and improve the accuracies on the test set.

Table 3
The effectiveness of the meta-learning framework without AuxBN on the FRGC v2 database.

Method	Accuracy (%)
PointNet	91.60
AT-PointNet w/o AuxBN	90.84 (↓)
AT-PointNet w AuxBN	93.59 (↑)
MLAT-PointNet w/o AuxBN	93.53 (↑)
MLAT-PointNet w AuxBN	94.84 (↑)
PointNet+	94.10
AT-PointNet+ w/o AuxBN	93.45 (↓)
AT-PointNet+ w AuxBN	95.66 (↑)
MLAT-PointNet+ w/o AuxBN	96.00 (↑)
MLAT-PointNet+ w AuxBN	96.96 (↑)

on clean samples (i.e., test samples). To verify above claim, we remove AuxBN [16] from AT and MLAT, and thus clean samples and adversarial samples are normalized by the same BN layers. From Table 3, we can find that: 1) AuxBN is important for AT to improve the performance on clean samples. Without AuxBN, AT-based models hurt the recognition accuracy on clean samples. 2) Even without AuxBN, MLAT-PointNet can still improve the rank-one score by 1.93% compared with PointNet, and MLAT-PointNet++ still outperform PointNet++ by 1.90%. These results indeed indicate the effectiveness of the meta-learning framework in alleviating the negative impact of the generated adversarial samples in deep 3DFR.

4.3.3. Effectiveness of the MLAT algorithm

In this section, we present some visualization results to further show the effectiveness of the proposed MLAT algorithm for deep 3DFR.

In Fig. 5, deep features extracted by different 3DFR models are projected into a 2D embedding space by t-SNE [43]. We can see that: compared with PointNet and PointNet++, deep features extracted by MLAT-PointNet and MLAT-PointNet++ are clustered more compactly, which indicates that MLAT-based models can extract more discriminative features.

In Fig. 6, we visualize the convex hulls of deep features of three identities in 2D embedding space. We can find that: 1) These deep features are not well-clustered by PointNet, and the intra-class distances of them are even larger than the inter-class distances. 2) By contrast, since MLAT algorithm can generate adversarial samples dynamically and update model parameters online under the meta-learning framework, during the training phase of MLAT-PointNet, the outlier samples can be gradually gathered to their corresponding class centers. Moreover, inter-class distances of deep features become larger and intra-class distances are decreased gradually.

Figure 7 illustrates a part of examples from FRGC v2 dataset that cannot be correctly identified by PointNet++, but can be identified correctly by MLAT-PointNet++. These examples almost contain hair occlusions or exaggerated expressions, which partly demonstrates that MLAT can improve the robustness of deep 3DFR model against hair occlusions and large expression variations.

4.3.4. Hyper-parameters tuning

We also analyze the effect of two important hyper-parameters in MLAT algorithm: the trade-off between classification loss on clean samples and adversarial samples (i.e., λ), iterations in adversarial sample generation (i.e., K).

Experimental result in Fig. 8(a) indicates that once the number of adversarial samples exceed a certain threshold, it will increase the instability and degrade the performance of 3DFR model. Figure 8(b) shows the accuracy curve under different λ. The accuracy reaches the summit when $\lambda = 0.4$ and drops slightly when λ increases. This is because a small λ will reduce the effect of adversarial samples, while a large λ will make the optimization direction tend to where the model performs better on adversarial samples. Thus, for all the experiments in this paper, we set $K = 2$ and $\lambda = 0.4$.

4.4. Comparison with the prior art

According to ablation studies in Section 4.3, we take MLAT algorithm with backbone PointNet++ (i.e., MLAT-PointNet++), and input modality of Points+Normal as our final model. In this section, we use our final model to compare with the prior state-of-the-art methods on the public 3D face datasets BU-3DFE, Bosphorus, BU-4DFE, and FRGC v2. It is worth noting that when one of them is set as the test set, the other three public datasets and our self-collected XJTU-3DFace dataset are merged as the training set.

Results on BU-3DFE. From Table 4, we can find that our proposed method achieves the state-of-the-art rank-one score of 100%, which demonstrates the expression-robustness of our method. It is worth noting that 26K samples of 1.2K identities in our training set contain real samples and their corresponding adversarial samples, while all 20K scans of 3.7K identities in the training set of Cai et al. [9] are real data. Furthermore, the training sets of Gilani and Mian [13] and Kim et al. [8] contain 100K identities and 123K samples respectively, which are far more than ours.

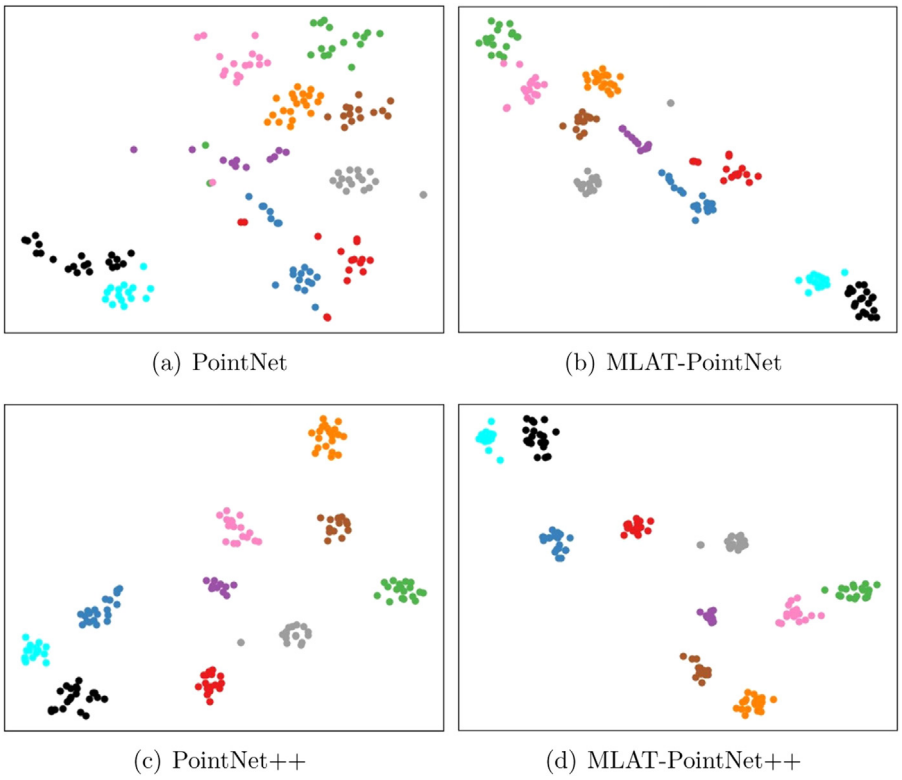

Fig. 5. t-SNE visualization of the deep features achieved by the baseline model PointNet and PointNet++, and the proposed models MLAT-PointNet and MLAT-PointNet++. (Different colors indicate different identities.).

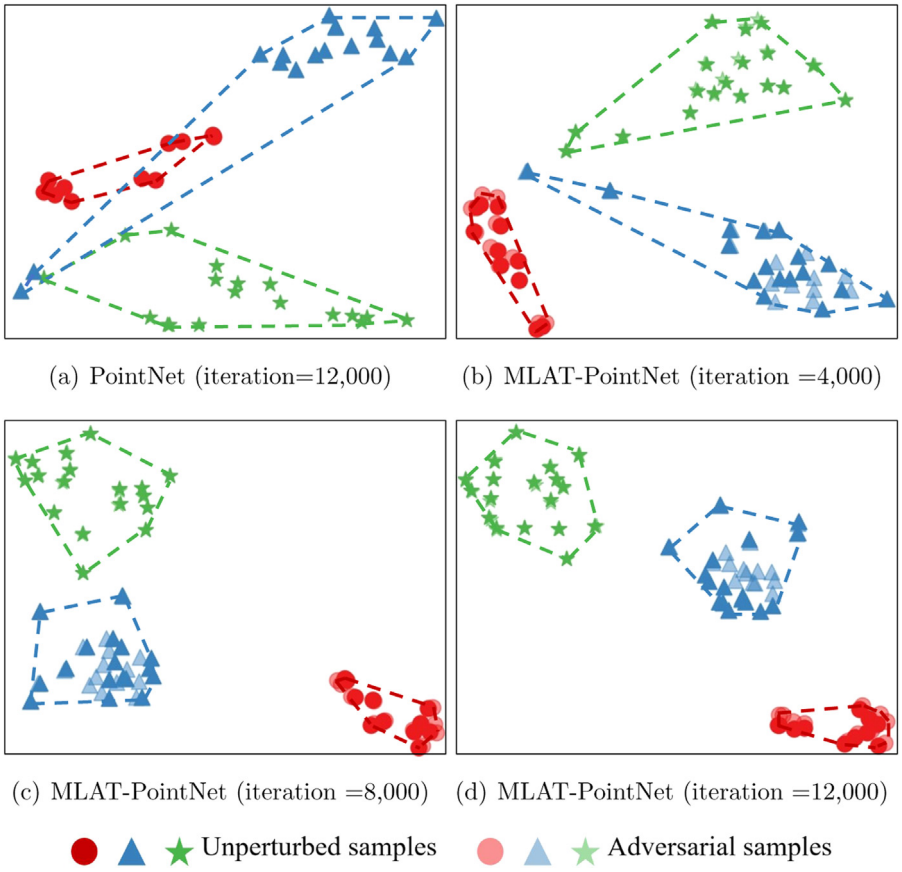

Fig. 6. Visualization of deep features extracted from unperturbed and adversarial samples belonging to the same three 3D face identities. We can find that since the adversarial sample generation and network training are optimized alternatively, our proposed MLAT algorithm can gradually gather the outlier samples to their corresponding class centers, and decrease the intra-class distances.

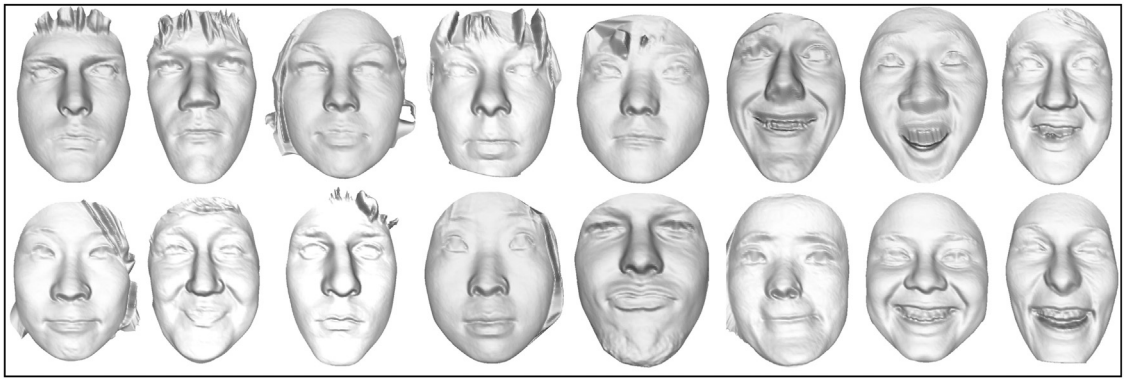

Fig. 7. A part of examples that cannot be correctly identified by PointNet++ but can be correctly identified by MLAT-PointNet++ in the FRGC v2 database.

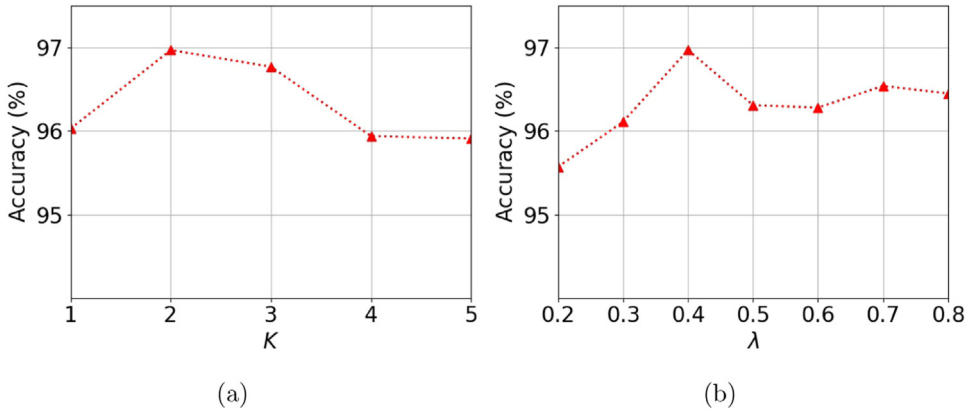

Fig. 8. Illustration of the hyper-parameter tuning of iterations in adversarial sample generation K, and the trade-off between classification loss on clean and adversarial samples λ.

Table 4
Comparison of the rank-one scores on the BU-3DFE database.

Method	Training data		Accuracy(%)
	Ids	Scans	
Li et al. [44]	-	-	92.20
Lei et al. [30]	-	-	93.20
Gilani et al. [45]	-	-	96.20
Li et al. [32]	-	-	96.10
Kim et al. [8]	0.7K	123K	95.00
Gilani and Mian [13]	100K	3.1M	98.64
Cai et al. [9]	3.7K	20K	99.88
MLAT-PointNet+ (ours)	1.2K	26K	**100**

Table 5
Comparison of the rank-one scores on the Bosphorus database.

Method	Traning data		Accuracy (%)
	Ids	Scans	
Li et al. [44]	-	-	95.40
Li et al. [46]	-	-	98.80
Lei et al. [30]	-	-	98.90
Gilani et al. [45]	-	-	98.60
Li et al. [32]	-	-	97.89
Kim et al. [8]	0.7K	123K	99.20
Cai et al. [9]	3.7K	20K	99.75
Zhang et al. [10]	10K	500K	99.68
Yu et al. [11]	10K	2000K	99.33
MLAT-PointNet+ (ours)	1.2K	25K	**99.78**

Results on Bosphorus. As shown in Table 5, our proposed method achieves a state-of-the-art rank-one score of 99.78%, and results of [8–11] are close to ours. Among them, the authors of [8] adopted a pre-trained 2DFR model and generated 123K 3D face scans for data augmentation. Cai et al. [9] used 20K real 3D face scans from 3.7K identities as training data, and adopted four deep networks to extract features. Zhang et al. [10] generated 500K 3D face samples of 10K identities by GPMM as the training data and utilized the angular loss and triplet loss. Moreover, the training set of Yu et al. [11] even contains up to 2000K 3D face scans. Compared with them, we only use 3D face scans of 1.2K subjects for training, and the point cloud network has not been pre-trained. Moreover, the proposed adversarial 3D face data augmentation is more convenient.

Table 6
Comparison of the rank-one scores on the BU-4DFE database.

Method	Traning data		Accracy (%)
	Ids	Scans	
Gilani et al. [45]	-	-	96.00
Gilani and Mian [13]	100K	3.1M	95.53
MLAT-PointNet+ (ours)	1.2K	25K	**98.02**

Results on BU-4DFE. From Table 6, we can find that our method achieves a rank-one score of 98.02%, which outperforms [13] by 2.49%. Gilani et al. [45] achieves a comparable rank-one score of 96.0%. Nevertheless, it relies on the sophisticated dense correspondence algorithm between 3D face scans.

Results on FRGC v2. From Table 7, we can find that our method achieves a comparable rank-one score of 98.01% on the FRGC v2. Similar to above cases, the scales of training sets of Gilani and Mian [13] (3.1M), Zhang et al. [10] (500K) and Yu et al. [11] (2000K) are far more than ours (24K). Meanwhile, real 3D face

Table 7
Comparison of the rank-one scores on the FRGC v2 database.

Method	Traning data		Accuracy (%)
	Ids	Scans	
Li et al. [46]	-	-	96.30
Lei et al. [30]	-	-	96.30
Gilani et al. [45]	-	-	98.50
Li et al. [32]	-	-	98.01
Gilani and Mian [13]	100K	3.1M	97.06
Gilani and Mian [13]	100K	3.1M	99.88
Cai et al. [9]	3.2K	19K	100
Zhang et al. [10]	10K	500K	99.46
Yu et al. [11]	10K	2000K	98.85
MLAT-PointNet+ (ours)	0.8K	24K	98.01

scans of 3.2K identities are used in Cai et al. [9], while only 0.8K identities are used in our training set. Moreover, the 3D face data augmentation methods in Gilani and Mian [13], Yu et al. [11] and Zhang et al. [10] are offline, and thus they need augment their face scans before network training. In contrast, the proposed adversarial 3D face data augmentation strategy is more convenient since it is an online method.

5. Conclusion

In this paper, we propose a novel Meta-Learning-based Adversarial Training algorithm for deep 3DFR, that generates adversarial 3D face point clouds, and trains deep 3DFR models using these adversarial samples as well as clean samples under a meta-learning framework. Compared with existing methods, the proposed data augmentation strategy is online and model-dependent, in which generating adversarial samples and updating model parameters are carried out alternately, and thus adversarial samples are dynamically and adaptively updated. Moreover, another advantage of our method is that compared with other methods, it can use relatively less real 3D face data to obtain competitive results.

Although our method achieves comparable results on the representative BU-3DFE, Bosphorus, BU-4DFE, and FRGC v2 databases, it still exists the following problems for future research. On one hand, how to choose a more efficient algorithm to generate adversarial 3D face scans for deep 3DFR is an important research direction. In this work, we only adopt the FGSM algorithm, the effectiveness of other adversarial sample generation algorithms is still open. On the other hand, the MAML-based framework proposed in this paper is an initial and beneficial attempt. How to improve the accuracy of deep 3DFR using the generated adversarial 3D face samples is also an important issue for further exploitation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

We thank the following professors and their group members: Yin et al., Savra et al., Phillips et al. and Zhang et al. for the BU-3DFE, Bosphorus, FRGC v2, and BU-4DFE datasets. The authors are supported in part by the National Natural Science Foundation of China (NSFC) under Grant (No. 61976173), the Fundamental Research Funds for the Central Universities (No. xzy012019041), the National Key Research and Development Program of China (No. 2018AAA0102200), the MoE-CMCC Artificial Intelligence Project (No. MCM20190701), and the National Natural Science Foundation of China (NSFC) under Grant (No. 61721002, No. U1811461).

References

[1] F. Schroff, D. Kalenichenko, J. Philbin, FaceNet: A unified embedding for face recognition and clustering, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2015, pp. 815–823.
[2] O.M. Parkhi, A. Vedaldi, A. Zisserman, Deep face recognition, in: Proc. BMVC, 2015, pp. 41.1–41.12.
[3] C. Samir, A. Srivastava, M. Daoudi, Three-dimensional face recognition using shapes of facial curves, IEEE Trans. Pattern Anal. Mach. Intell. 28 (11) (2006) 1858–1863.
[4] S. Berretti, A.D. Bimbo, P. Pala, 3D face recognition using isogeodesic stripes, IEEE Trans. Pattern Anal. Mach. Intell. 32 (12) (2010) 2162–2177.
[5] I.A. Kakadiaris, G. Passalis, G. Toderici, M.N. Murtuza, Y. Lu, N. Karampatziakis, T. Theoharis, Three-dimensional face recognition in the presence of facial expressions: An annotated deformable model approach, IEEE Trans. Pattern Anal. Mach. Intell. 29 (4) (2007) 640–649.
[6] D. Huang, G. Zhang, M. Ardabilian, Y. Wang, L. Chen, 3D face recognition using distinctiveness enhanced facial representations and local feature hybrid matching, in: Proc. BTAS, 2010, pp. 1–7.
[7] L.J. Spreeuwers, Fast and accurate 3D face recognition - Using registration to an intrinsic coordinate system and fusion of multiple region classifiers, Int. J. Comput. Vis. 93 (3) (2011) 389–414.
[8] D. Kim, M. Hernandez, J. Choi, G.G. Medioni, Deep 3D face identification, in: Proc. Int. Joint Conf. on Biometrics (IJCB), 2017, pp. 133–142.
[9] Y. Cai, Y. Lei, M. Yang, Z. You, S. Shan, A fast and robust 3D face recognition approach based on deeply learned face representation, Neurocomputing 363 (2019) 375–397.
[10] Z. Zhang, F. Da, Y. Yu, Learning directly from synthetic point clouds for "in--the-wild" 3D face recognition, Pattern Recognit. 123 (2022) 108394.
[11] Y. Yu, F. Da, Z. Zhang, Few-data guided learning upon end-to-end point cloud network for 3d face recognition, Multim. Tools Appl. 81 (9) (2022) 12795–12814.
[12] J. Zhao, L. Xiong, J. Li, J. Xing, S. Yan, J. Feng, 3d-aided dual-agent GANs for unconstrained face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 41 (10) (2019) 2380–2394.
[13] S.Z. Gilani, A. Mian, Learning from millions of 3D scans for large-scale 3D face recognition, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2018, pp. 1896–1905.
[14] F. Qiao, L. Zhao, X. Peng, Learning to learn single domain generalization, in: in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2020, pp. 12553–12562.
[15] C. Finn, P. Abbeel, S. Levine, Model-agnostic meta-learning for fast adaptation of deep networks, in: Proc. ICML, volume 70, 2017, pp. 1126–1135.
[16] C. Xie, M. Tan, B. Gong, J. Wang, A.L. Yuille, Q.V. Le, Adversarial examples improve image recognition, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2020, pp. 816–825.
[17] C.R. Qi, H. Su, K. Mo, L.J. Guibas, PointNet: Deep learning on point sets for 3D classification and segmentation, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2017, pp. 77–85.
[18] C.R. Qi, L. Yi, H. Su, L.J. Guibas, PointNet++: Deep hierarchical feature learning on point sets in a metric space, in: Proc. NIPS, 2017, pp. 5099–5108.
[19] Y. Wang, Y. Sun, Z. Liu, S.E. Sarma, M.M. Bronstein, J.M. Solomon, Dynamic graph CNN for learning on point clouds, ACM Trans. Graph. 38 (5) (2019) 146:1–146:12.
[20] I.J. Goodfellow, J. Shlens, C. Szegedy, Explaining and harnessing adversarial examples, in: Proc. ICLR, 2015.
[21] A. Madry, A. Makelov, L. Schmidt, D. Tsipras, A. Vladu, Towards deep learning models resistant to adversarial attacks, in: Proc. ICLR, 2018.
[22] A. Sinha, H. Namkoong, J.C. Duchi, Certifying some distributional robustness with principled adversarial training, in: Proc. ICLR, 2018.
[23] C. Xie, Y. Wu, L. van der Maaten, A.L. Yuille, K. He, Feature denoising for improving adversarial robustness, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2019, pp. 501–509.
[24] M. Li, B. Huang, G. Tian, A comprehensive survey on 3D face recognition methods, Eng. Appl. Artif. Intell. 110 (2022) 104669.
[25] S. Soltanpour, B. Boufama, Q.M.J. Wu, A survey of local feature methods for 3D face recognition, Pattern Recognit. 72 (2017) 391–406.
[26] P.J. Neugebauer, Reconstruction of real-world objects via simultaneous registration and robust combination of multiple range images, Int. J. Shape Model. 3 (1-2) (1997) 71–90.
[27] X. Yu, Y. Gao, J. Zhou, Sparse 3D directional vertices vs continuous 3D curves: Efficient 3D surface matching and its application for single model face recognition, Pattern Recognit. 65 (2017) 296–306.
[28] S. Soltanpour, B. Boufama, Q. Jonathan Wu, A survey of local feature methods for 3D face recognition, Pattern Recognit. 72 (2017) 391–406.
[29] Y. Lei, M. Bennamoun, M. Hayat, Y. Guo, An efficient 3D face recognition approach using local geometrical signatures, Pattern Recognit. 47 (2) (2014) 509–524.
[30] Y. Lei, Y. Guo, M. Hayat, M. Bennamoun, X. Zhou, A two-phase weighted collaborative representation for 3D partial face recognition with single sample, Pattern Recognit. 52 (2016) 218–237.
[31] S. Elaiwat, M. Bennamoun, F. Boussaid, A. El-Sallam, A curvelet-based approach for textured 3D face recognition, Pattern Recognit. 48 (4) (2015) 1235–1246.
[32] H. Li, J. Sun, L. Chen, Location-sensitive sparse representation of deep normal patterns for expression-robust 3D face recognition, in: Proc. Int. Joint Conf. on Biometrics (IJCB), 2017, pp. 234–242.

[33] V. Blanz, T. Vetter, A morphable model for the synthesis of 3D faces, in: Proc. SIGGRAPH, 1999, pp. 187–194.
[34] A.R. Bhople, A.M. Shrivastava, S. Prakash, Point cloud based deep convolutional neural network for 3D face recognition, Multim. Tools Appl. 80 (20) (2021) 30237–30259.
[35] A.R. Bhople, S. Prakash, Learning similarity and dissimilarity in 3d faces with triplet network, Multim. Tools Appl. 80 (28-29) (2021) 35973–35991.
[36] C. Jiang, S. Lin, W. Chen, F. Liu, L. Shen, PointFace: Point set based feature learning for 3D face recognition, in: Proc. Int. Joint Conf. on Biometrics (IJCB), 2021, pp. 1–8.
[37] L. Yin, X. Wei, Y. Sun, J. Wang, M. Rosato, A 3D facial expression database for facial behavior research, in: Proc. on Automatic Face and Gesture Recognition, 2006.
[38] A. Savran, N. Alyüz, H. Dibeklioglu, O. Çeliktutan, B. Gökberk, B. Sankur, L. Akarun, 3D face recognition benchmarks on the Bosphorus database with focus on facial expressions, in: Proc. Workshop on Biometrics and Identity Management, 2008.
[39] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, W. Worek, Overview of the face recognition grand challenge, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2005.
[40] X. Zhang, L. Yin, J.F. Cohn, S.J. Canavan, M. Reale, A. Horowitz, P. Liu, A high--resolution spontaneous 3D dynamic facial expression database, in: Proc. FG, 2013, pp. 1–6.
[41] H. Mohammadzade, D. Hatzinakos, Iterative closest normal point for 3D face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 35 (2) (2013) 381–397.
[42] Y. Xu, C. Jung, Y. Chang, Head pose estimation using deep neural networks and 3D point clouds, Pattern Recognit. 121 (2022) 108210.
[43] L. van der Maaten, G. Hinton, Visualizing data using t-sne, Journal of Machine Learning Research 9 (86) (2008) 2579–2605.
[44] H. Li, D. Huang, J. Morvan, L. Chen, Y. Wang, Expression-robust 3D face recognition via weighted sparse representation of multi-scale and multi-component local normal patterns, Neurocomputing 133 (2014) 179–193.
[45] S.Z. Gilani, A.S. Mian, F. Shafait, I. Reid, Dense 3D face correspondence, IEEE Trans. Pattern Anal. Mach. Intell. 40 (7) (2018) 1584–1598.
[46] H. Li, D. Huang, J. Morvan, Y. Wang, L. Chen, Towards 3D face recognition in the real: A registration-free approach using fine-grained matching of 3D keypoint descriptors, Int. J. Comput. Vis. 113 (2) (2015) 128–142.

Cuican Yu received the bachelor's degree from the School of Mathematics and Applied Mathematics, University of Electronic Science and Technology of China, China, in 2017. She is currently pursuing the Ph.D. degree with the School of Mathematics and Statistics, Xi'an Jiaotong University. Her research interests include 3D computer vision, 2D and 3D face recognition.

Zihui Zhang received the bachelor's degree from the School of Mathematics and Applied Mathematics, University of Electronic Science and Technology of China in 2018, and the master's degree in mathematics from Xi'an Jiaotong University in 2021. He is currently working toward the Ph.D. degree in the department of Computing in Hong Kong Polytechnic University. His research interests lie in 3D computer vision, point clouds analysis and 3D face recognition.

Huibin Li received the bachelor's degree in mathematics from Shaanxi Normal University in 2006, the master's degree in mathematics from Xi'an Jiaotong University, in 2009, and the Ph.D. degree in mathematics and computer science from École Centrale de Lyon, LIRIS, Université de Lyon, CNRS, Lyon, France. He is currently an Associate Professor with the School of Mathematics and Statistics, Xi'an Jiaotong University. His research interests include 2D and 3D computer vision, machine learning, and medical image processing.

Jian Sun received the Ph.D. degree in applied mathematics from Xi'an Jiaotong University. He worked as a visiting student in Microsoft Research Asia (Nov. 2005 - March 2008), a post-doctoral researcher with University of Central Florida in USA (August 2009 - April 2010), and a post-doctoral researcher in willow team of Ecole Normale Supérieure de Paris/INRIA (Sept. 2012 - August 2014). He now serves as a professor with the school of mathematics and statistics of Xi'an Jiaotong University. His research interests include the mathematics and machine learning-based approaches for image processing/recognition, and medical image analysis.

Zongben Xu received his Ph.D. degree in mathematics from Xi'an Jiaotong University, China, in 1987. He now serves as the Chief Scientist of National Basic Research Program of China (973 Project), and Director of the Institute for Information and System Sciences of the university. He is owner of the National Natural Science Award of China in 2007, and winner of CSIAM Su Buchin Applied Mathematics Prize in 2008. He delivered a 45 minute talk on the International Congress of Mathematicians 2010. He was elected as member of Chinese Academy of Science in 2011. His current research interests include intelligent information processing and applied mathematics.