ID-Softmax: A Softmax-like Loss for ID Face Recognition

Yan Kong[1], Fuzhang Wu[1], Feiyue Huang[2] and Yanjun Wu[3]
[1]*Institute of Software, Chinese Academy of Sciences, Beijing, China*
[2]*YouTu Lab, Tencent, China*
[3]*State Key Laboratory of Computer Science, Beijing, China*

Keywords: Face Recognition, Loss Function, Softmax, Convolutional Neural Network.

Abstract: The face recognition between photos from identification documents (ID, Citizen Card or Passport Card) and daily photos, which is named FRBID(Zhang et al., 2017), is widely used in real world scenarios. However, traditional Softmax loss of deep CNN usually lacks the power of discrimination for FRBID. To address this problem, in this paper, we first revisit recent progress of face recognition losses, and give the theoretical and experimental analysis on the reason why Softmax-like losses work badly on ID-daily face recognition. Then we propose an novel approach named ID-Softmax, which use ID face features as class 'agent' to guide the deep CNNs to learn highly discriminative features between ID photos and daily photos. In order to promote the ID-daily face recognition, we collect a large dataset ID74K, which includes 74,187 identities with corresponding ID photos and daily photos. To test our approach, we evaluate the feature distribution and face verification performance on dataset ID74K. In experiments, we achieve the best performance when comparing with other state-of-the-art methods, which verifies the effectiveness of the proposed ID-Softmax loss.

1 INTRODUCTION

Face recognition is a biometric identification technology based on human facial feature information, as it has been studied generally for over 50 years. In recent years, with the rapid development of big data and deep learning, there have achieved remarkable improvements in deep face recognition (verification in particular) (Sun et al., 2014; Lu and Tang, 2014). In real world applications, face recognition between ID photos and daily photos, which is known as face recognition between photos from identification documents (ID, Citizen Card or Passport Card) and daily photos (FRBID) is gaining more attention because it uses face from an ID photo as gallery and thus does not require the probe to be registered in advance (Zhang et al., 2017). We show examples of ID photo and daily photos in Figure 1. In this paper, we represent face verification between ID face photo and daily face photo as ID-daily, and represent face verification between daily face photo and daily face photo as daily-daily.

Though the previous deep face models can achieve fascinating results(Schroff et al., 2015; Wen et al., 2016; Liu et al., 2016; Liu et al., 2017; Wang et al., 2018), researchers find that the recognition performance drops dramatically when these models are applied to the real world security certificate applications (Zhou et al., 2015). There are several issues associated with FRBID.

The first challenge is the data imbalance of ID photos and daily photos in train phase. Benefiting from the dramatic increased web data, we can collect millions of daily photos easily. Due to the photo capture environment and privacy issues, the ID photos are always restricted on the Internet. Hence, the collection of large scale and pair-wised ID-daily photos is still expensive. How to apply deep learning on an unbalanced ID photos dataset remains a general problem.

The second challenge encountered is the heterogeneity of shooting environment between the gallery set and probe set (Xie et al., 2015). In real world scenarios, even though the ID photos or e-passports are captured in a very stable environment, most ID photos are compressed with low quality parameter because of ROM(Read-Only Memory) limitation. Furthermore, the probe photos are captured in a highly unstable environment using equipments such as surveillance and mobile-phone cameras. Noise, blur, arbitrary pose and age changing increase the recognition difficult between ID photos and daily photos (Hong et al., 2017).

Under the scenario of FRBID, an obvious diffe-

Kong, Y., Wu, F., Huang, F. and Wu, Y.
ID-Softmax: A Softmax-like Loss for ID Face Recognition.
DOI: 10.5220/0007370904120419
In *Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2019)*, pages 412-419
ISBN: 978-989-758-354-4

rence is the frequency of ID-daily face recognition between train phase and test phase. In train phase, the sampling frequency of ID-daily photos is low because of imbalance of ID photos and daily photos. In real world application(test phase), all recognitions are between ID photos and daily photos. Most recently face recognition algorithms (Wen et al., 2016; Liu et al., 2016; Liu et al., 2017; Wang et al., 2018) are not designed to be optimized well under imbalance FRBID scenario. To overcome the difficulty of FRBID, we propose a novel ID-Softmax loss, which aims to optimizing ID-daily face recognition directly.

Our major contributions can be summarized as follows:

(1)To train an available CNN model for real world ID-daily face recognition applications, we collect a face dataset named ID74K, which contains ID photos and the corresponding daily photos for each person.

(2)We propose a new Softmax-like loss (ID-Softmax) as training supervision to solve the imbalance of ID-daily face photos in training datasets. By simulating real ID-daily face recognition scenario, we use ID face features as class 'agent' to guide the deep CNNs to learn highly discriminative features between ID photos and daily photos. Our ID-Softmax loss could improve the performance of FRBID obviously.

(3)Our experiments show that with the supervision of ID-Softmax, the trained CNN model achieves better recognition performance when comparing with other existing methods.

2 RELATED WORK

Deep Convolutional Neural Network. Convolutional neural networks (CNNs) have been widely used in computer vision community, and improve the state-of-the-art performance for a wide variety of computer vision task significantly. Face recognition as an important computer vision application, has achieved significant progress thanks to the great success of deep CNN models, such as VGG(Simonyan and Zisserman, 2014), GooLeNet(Szegedy et al., 2015), ResNet(He et al., 2016) and so on.

Face Recognition Loss Function. Loss function plays an important role in deep face recognition. In the early work of deep face recognition (Sun et al., 2014; Taigman et al., 2014), model is trained on a labeled facial dataset supervised by Softmax loss, and then the feature vector is taken from an intermediate layer of the network for face recognition. Since Softmax loss does not directly optimize the face feature comparison in face recognition, in order to further improve the discriminative of face feature, researchers proposed new loss functions running in Euclidean space and angular space.

In Euclidean space based loss, the Contrastive loss (Chen et al., 2014) and the Triplet loss (Schroff et al., 2015) use pair training strategy to reduce inner-class variations and increase inter-class variations. However, a good sampling method is essential for the Contrastive loss and the Triplet loss to guarantee a good model convergence. In order to reduce the optimizing difficulty, Center loss (Wen et al., 2016) learns class feature centers for each identities, which looses the constraint metric from pairwise distance to instance-center distance, but it still need to combine with Softmax loss to training recognition model.

Benefiting from better geometric interpretation, the angular space based losses are attracting more attention of researchers. Both Large Margin Softmax (Liu et al., 2016) and SphereFace (Liu et al., 2017) add angular constraints for each identities by multiplying a parameter m on feature angle. In order to make both cosine loss function can be optimized, a piecewise function is introduced to guarantee the monotonicity. Furthermore, Large Margin Softmax and SphereFace also need original Softmax to ensure the convergence. To overcome the optimization difficulty, CosFace (Wang et al., 2018) introduces margin in cosine space instead of angular space. CosFace can be implemented easily and archives the state-of-the-art performance on MegaFace (Kemelmacher-Shlizerman et al., 2016).

Normalization. Feature and weight normalization have be proved very effective for face recognition. NormFace (Wang et al., 2017) normalizes the learned deep features and weight matrix of the fully connected (FC) layer before Softmax loss layer, which forces CNN to concentrate more on the angle optimization while ignoring radial variation. SphereFace (Liu et al., 2017) and CosFace (Wang et al., 2018) also use normalization to improve face recognition performance.

3 THE PROPOSED APPROACH

In this section, we firstly introduce the dataset ID74K used for our training and testing. Then, we revisit recent progress on Softmax loss and analysis its drawback on FRBID. Finally, we introduce our ID-Softmax loss.

3.1 Data Collection

Most famous public face recognition datasets, such as LFW(Huang et al., 2008), CASIA-WebFace (Yi

(a) ID photo (b) Daily photos

Figure 1: An example of ID photo and daily photos in our ID74K dataset. (a) is an ID photo collected from IC chip embedded in Chinese Identity Card by a Card Reader, and its image quality is low because of image compression, (b) is a series of daily life photos captured by mobile-phone cameras.

et al., 2014), MegaFace (Kemelmacher-Shlizerman et al., 2016), are crawled from Internet. Due to lack of ID face photos, researchers find that the recognition performance drops dramatically when models trained on public datasets are applied on (Zhou et al., 2015). In order to overcome the data limitation, we collect an ID-daily face recognition dataset named ID74K. There are 74,187 identities in ID74K, and each identity contains 1 ID face photo and 5 daily photos (some examples showed in Figure 1) . The ID face photos are collected from IC chip embedded in Chinese Identity Card by a Card Reader, while the daily photos are captured by mobile-phone cameras from real life. All these photos are provided by volunteers with reasonable payment. In our experiments, we use 70,000 identities for training and use 4,187 identities for evaluation. It need to note that our dataset is collected from daily life, which is very different from public face recognition dataset, and the distribution of ID photos and daily photos are still heavy unbalanced.

3.2 Recent Progress on Softmax

Classical Softmax. As an important part of deep image classification, Softmax loss is existed in deep models generally. Softmax function is a generalization of the logistic function. Given an input feature vector $\mathbf{x}_i$ and its corresponding label y_i, the classical Softmax loss can be written as

$$\begin{aligned} L_s &= -\frac{1}{N}\sum_{i=0}^{N} log(p_i) \\ &= -\frac{1}{N}\sum_{i=0}^{N} log(\frac{e^{\mathbf{W}_{y_i}^T\mathbf{x}_i+\mathbf{b}_{y_i}}}{\sum_j e^{\mathbf{W}_j^T\mathbf{x}_i+\mathbf{b}_j}}). \end{aligned} \tag{1}$$

In equation 1, let d be the feature dimension of $\mathbf{x}_i$, $\mathbf{W}_j$ is the j-th column of $\mathbf{W} \in \mathbb{R}^{d\times n}$ in the last fully connected layer, $\mathbf{b} \in \mathbb{R}^n$ is the bias term, and $\mathbf{b}_j$ is the j-th element of $\mathbf{b}$. The size of mini-batch is N. We fix the bias $\mathbf{b} = 0$ for simplicity, and the classical Softmax function can be rewritten as

$$L_s = -\frac{1}{N}\sum_{i=0}^{N} log(\frac{e^{||\mathbf{W}_{y_i}||||\mathbf{x}_i||\cos(\theta_{y_i,i})}}{\sum_{j=0}^{n} e^{||\mathbf{W}_j||||\mathbf{x}_i||\cos(\theta_{j,i})}}). \tag{2}$$

where the $\theta_{j,i}$ is the angle between $\mathbf{W}_j$ and $\mathbf{x}_i$.

Normalized Softmax. Feature and weight normalization have been proved effective for face recognition (Wang et al., 2017). With L2 normalization of $\mathbf{W}_j$ and $\mathbf{x}_i$, the neural network can directly optimize the cosine similarity. After normalization, the neural network will fail to converge. In order to avoid the convergence difficulty, we follow the suggestion in NormFace (Wang et al., 2017), and introduce a scalar factor s to the normalization version of Softmax loss. Hence, the Softmax loss with cosine distance can be rewritten as

$$L_n = -\frac{1}{N}\sum_{i=0}^{N} log(\frac{e^{s\cos(\theta_{y_i,i})}}{\sum_j e^{s\cos(\theta_{j,i})}}). \tag{3}$$

subject to

$$\begin{aligned} \tilde{\mathbf{W}} &= \frac{\mathbf{W}}{||\mathbf{W}||}, \\ \tilde{\mathbf{x}} &= \frac{\mathbf{x}}{||\mathbf{x}||}, \\ \cos(\theta_{j,i}) &= \tilde{\mathbf{W}}_j^T\tilde{\mathbf{x}}_i. \end{aligned}$$

Large Margin Cosine Softmax. Deep features learned by classical Softmax and normalized Softmax are still not sufficiently discriminative because these losses only emphasize correct classification. In order to further reduce inner-class variations and increase inter-class variations of face feature, CosFace (Wang et al., 2018) add cosine margin to the classification boundary by introducing a parameter m, which is naturally incorporated with the cosine formulation of Softmax. The large margin cosine Softmax is defined as

$$L_c = -\frac{1}{N}\sum_{i=0}^{N} log(\frac{e^{s\cdot(\cos(\theta_{y_i,i})-m)}}{e^{s\cdot(\cos(\theta_{y_i,i})-m)} + \sum_{j\neq y_i} e^{s\cdot\cos(\theta_{j,i})}}). \quad (4)$$

3.3 ID-Softmax

At the beginning of this section, we analysis the Softmax-like loss theoretically. Inspired by Norm-Face (Wang et al., 2017), normalized $\tilde{\mathbf{W}}_j^T\tilde{\mathbf{x}}_i$ could be rewritten as

$$\tilde{\mathbf{W}}_j^T\tilde{\mathbf{x}}_i = 1 - \frac{1}{2}||\tilde{\mathbf{W}} - \tilde{\mathbf{x}}_i||_2^2, \quad (5)$$

Hence, we reformat formula (3) as

$$\begin{aligned} L_n^{'} &= -\frac{1}{N}\sum_{i=0}^{N} log(\frac{e^{s(1-\frac{1}{2}||\tilde{\mathbf{W}}_{y_i}-\tilde{\mathbf{x}}_i||_2^2)}}{\sum_j e^{s(1-\frac{1}{2}||\tilde{\mathbf{W}}_j-\tilde{\mathbf{x}}_i||_2^2)}}) \\ &= -\frac{1}{N}\sum_{i=0}^{N} log(\frac{e^s\cdot e^{-\frac{s}{2}||\tilde{\mathbf{W}}_{y_i}-\tilde{\mathbf{x}}_i||_2^2}}{\sum_j e^s\cdot e^{-\frac{s}{2}||\tilde{\mathbf{W}}_j-\tilde{\mathbf{x}}_i||_2^2}}). \end{aligned} \quad (6)$$

Due to $\frac{s}{2}||\tilde{\mathbf{W}}_j - \tilde{\mathbf{x}}_i||_2^2 \geq 0$ and $f(x) = e^x$ is monotonously, the learning process of Softmax for CNN can be formulated as

$$\begin{aligned} &\arg\min_{\mathbf{W},\mathbf{x}} ||\tilde{\mathbf{W}}_j - \tilde{\mathbf{x}}_i||_2^2, \quad if\ y_i = j \\ &\arg\max_{\mathbf{W},\mathbf{x}} ||\tilde{\mathbf{W}}_j - \tilde{\mathbf{x}}_i||_2^2, \quad if\ y_i \neq j, \end{aligned} \quad (7)$$

which is similar to triplet loss,

$$\begin{aligned} L = max(0, m + ||\mathbf{x}_i - \mathbf{x}_p||_2^2 - ||\mathbf{x}_i - \mathbf{x}_n||_2^2) \\ y_i = y_p, y_i \neq y_n. \end{aligned} \quad (8)$$

Hence, the $\mathbf{W}_i$ can be treated as the 'agent' of the i-th class. During the training phase, Softmax loss minimizes the angle between $\mathbf{W}_{y_i}$ and $\mathbf{x}_i$, and maximizes the angle between $\mathbf{W}_{j\neq y_i}$ and $\mathbf{x}_i$. After network convergence, the $\mathbf{W}_i$ will roughly correspond to the means of features of the i-th class, because Softmax assumes that individual samples of classes are equally important.

However, in real world FRBID scenario, all face comparison are between ID photos and daily photos. As mentioned in section 1, there are obvious difference between ID photos and daily photos, which leads that the feature of i-th class ID photos is far away from $\mathbf{W}_i$.

As illustrated in Figure 2, under classical Softmax loss, $\mathbf{W}_0$ and $\mathbf{W}_1$ will converge to the means of features of class 0 and class 1 respectively. Since there is a large margin between $\mathbf{W_0}$ and $\mathbf{z_0}$, the learned ID face feature $\mathbf{z_0}$ is not a good representation of class 0. The mean θ reported in Table 1 reflects the margin between $\mathbf{W_0}$ and $\mathbf{z_0}$. The cosine margin introduced by CosFace can reduce the difference between $\mathbf{W_0}$ and $\mathbf{z_0}$, but it cannot solve the problem intrinsically.

According to the above theoretical and experimental analysis, the optimization goals of recent proposed Softmax losses are different from the FR-BID scenario. So, how to develop an effective loss function to improve the discriminative power in FR-BID scenario ? It is intuitive to replace the 'agent' of i-th class ($\mathbf{W}_i$) with the feature of i-th class ID photo($\mathbf{z}_i$). Let's rewrite the formula (3) as

$$\begin{aligned} L_{id} &= -\frac{1}{N}\sum_{i=0}^{N} log(\frac{e^{s||\tilde{\mathbf{z}}_{y_i}||||\tilde{\mathbf{x}}_i||\cos(\theta^{'}_{y_i,i})}}{\sum_j e^{s||\tilde{\mathbf{z}}_j||||\tilde{\mathbf{x}}_i||\cos(\theta^{'}_{j,i})}}), \\ &= -\frac{1}{N}\sum_{i=0}^{N} log(\frac{e^s\cdot e^{-\frac{s}{2}||\tilde{\mathbf{z}}_{y_i}-\tilde{\mathbf{x}}_i||_2^2}}{\sum_j e^s\cdot e^{-\frac{s}{2}||\tilde{\mathbf{z}}_j-\tilde{\mathbf{x}}_i||_2^2}}) \end{aligned} \quad (9)$$

where the $\theta^{'}_{j,i}$ is the angle between $\mathbf{z}_j$ and $\mathbf{x}_i$. The formulation effectively characterizes the ID-daily face feature variations. Ideally, we need to update $\mathbf{z}$ and $\mathbf{x}$ as the deep feature changed. In other words, we need to extract ID photo features of every class in every iteration, which is inefficient even impractical.

To address this problem, one solution is using pair training strategy like contrastive loss and triplet loss, which is not easy enough to training. In our solution, we make necessary modification for Formula 9. We replace the ID face feature $\mathbf{z}_i$ with the snapshot of ID face feature $\mathbf{z}_i^{'}$. In each iteration, the $\mathbf{z}_i^{'}$ is replaced by the ID face feature of the corresponding classes in mini-batch, which means 'agent' of some classes may not update. In other words, if there are ID card features in one mini-batch, then only $\mathbf{z}_i^{'}$s corresponding these ID cards are updated. Formally, we adopt Large Margin Cosine Softmax and define ID-Softmax as

$$L_{id} = -\frac{1}{N}\sum_{i=0}^{N} log(\frac{e^{s(\cos^{'}(\theta_{y_i,i})-m)}}{e^{s(\cos^{'}(\theta_{y_i,i})-m)} + \sum_{j\neq y_i} e^{s(\cos^{'}(\theta_{j,i}))}}). \quad (10)$$

subject to

$$\begin{aligned} \tilde{\mathbf{z}}^{'} &= \frac{\mathbf{z}^{'}}{||\mathbf{z}^{'}||}, \\ \tilde{\mathbf{x}} &= \frac{\mathbf{x}}{||\mathbf{x}||}, \\ \cos^{'}(\theta_{j,i}) &= \tilde{\mathbf{z}}_j^{'T}\tilde{\mathbf{x}}_i. \end{aligned}$$

The learning procedure of ID-Softmax can be summarized as Algorithm 1.

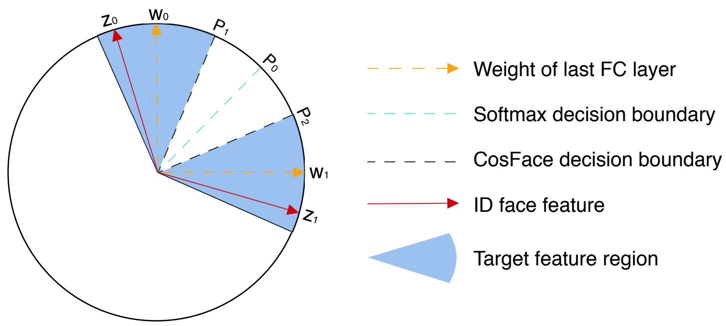


Figure 2: Geometrical interpretation of Softmax loss under feature perspective. In FRBID scenario, $\mathbf{z_0}$ is used as the representation of class 0. However, since there is a large margin between $\mathbf{W_0}$ and $\mathbf{z_0}$, the learned ID face feature $\mathbf{z_0}$ is not a good representation of class 0.

Algorithm 1: The ID-Softmax feature learning algorithm.

Input: Training data $\mathbf{I}_i$. Nerual network feature function $f(\theta_C, \mathbf{I}_i)$. Initialized face feature network parameter θ_C. Feature parameter of last fully connected layer $\mathbf{z}'_{y_i}$. (The learning rate of $\mathbf{z}'_{y_i}$ is set to 0.) Learing rate μ^t. The number of iterator t = 0.
Output: face feature network parameter θ_C
while *network not converge* **do**
  t = t + 1;
  Foward network and compute the ID-Softmax loss L_{id};
  For each $\mathbf{I}_i$ in current iteration, replace the parameter $\mathbf{z}'_{y_i}$ by $f(\theta_C, \mathbf{I}_i)$ if $\mathbf{I}_i$ is ID photo. ;
  Backward network and update parameter θ_C.

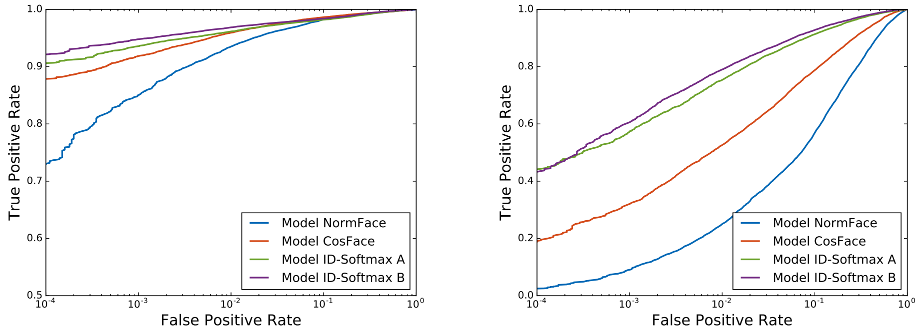
4 EXPERIMENTS

4.1 Implementation Details

Preprocessing. For all face photos, we use public available MTCNN (Zhang et al., 2016) open source implementation to detect and align faces. The 5 facial points generated by MTCNN are used to perform similarity transformation. All face photos are resized to 120x120 size. Each pixel of RGB photos is normalized by subtracting 127.5 then dividing by 128.
Training. Since CNN have achieved the outstanding performance in the face recognition tasks, we use ResNet-18 (He et al., 2016) with 512 output of fully connected layer as CNN architecture. Our model and training code are implemented on MxNet framework (Chen et al., 2015). The CNN models are trained by SGD with momentum. We set momentum to 0.9, initial learning rate to 0.01, weight decay to 0.0005, batch size to 128. The networks are trained on 4 Nvidia Tesla P40 GPUs. For all models, we train CNN by 120 epochs, and the learning rate is divided by 10 at the 40, 80, 120 epochs. The training faces are horizontally flipped for data augmentation. As mentioned in Section 3.1, in our experiments, we use 70,000 identities for training and use 4,187 identities for evaluation. There is no overlap between training dataset and evaluation dataset.

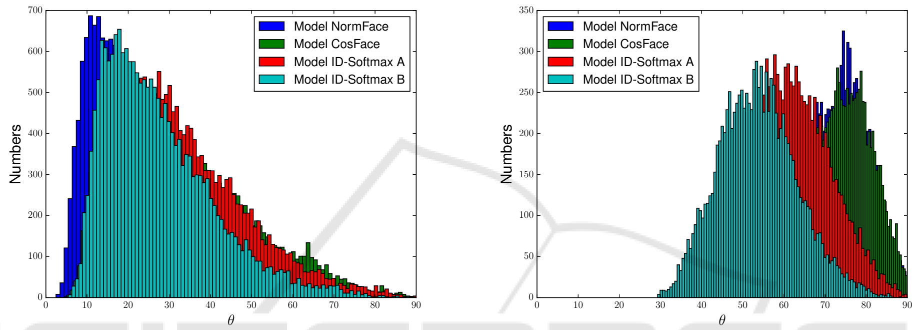
4.2 Evaluation

In this section, we compare the face recognition performance of different loss functions on FRBID scenario, and evaluate the influence of different update strategies. For fair comparison, we respectively train four kind of models under the supervision of Normalized Softmax(Model NormFace), Large Margin Cosine Softmax loss (Model CosFace) and ID-Softmax loss(Model ID-Softmax A, Model ID-Softmax B). Due to lack of ID face photos in public available datasets, such as LFW, CSAIA-WebFace and so on, we only evaluate different methods on our ID74K dataset. The experiment results of Table 2 show that our method is also competitive in “daily vs daily” face recognition scenario.
Model NormFace: We use NormFace as baseline model. The training procedure is described in section 4.1. Furthermore, we set scaling parameter s to 64, which is used by CosFace paper (Wang et al., 2018). It takes 45 hours to train this model.
Model CosFace: The CosFace has been proven effective to learn compact face feature for face recog-



(a) ROC curves of daily-daily scenario. (b) ROC curves of ID-daily scenario.

Figure 3: We draw the ROC curves of Model NormFace, Model CosFace, Model ID-Softmax A, Model ID-Softmax B in ID-daily/daily-daily scenario respectively.



(a) θ distribution of daily-daily scenario. (b) θ distribution of ID-daily scenario.

Figure 4: We draw the θ distribution of Model NormFace, Model CosFace, Model ID-Softmax A, Model ID-Softmax B in ID-daily/daily-daily scenario respectively.

nition. The training procedure is described in section 4.1. Furthermore, we set scaling parameter s to 64, and the margin parameter m is set to 0.35 which is suggested by CosFace paper (Wang et al., 2018). It takes 45 hours to train this model.

Model ID-Softmax A: This model is trained by ID-Softmax loss. We update parameter $\mathbf{z}'$ in every mini-batch iteration, and only the ID photo features existed in current mini-batch are adopted for update. The learning rate of Model ID-Softmax A is set to 0. It takes 45 hours to train this model.

Model ID-Softmax B: This model is trained by ID-Softmax loss. We update parameter $\mathbf{z}'$ in every 60 mini-batch iteration. Compared with Model ID-Softmax A, we extract ID features of ID photos in training dataset for update **all at the once**. We update $\mathbf{z}'$ every 60 mini-batch iteration instead of every iteration because updating all ID's $\mathbf{z}'$s is computing expensive. We choose 60 as update frequency empirically to accelerate network convergence. The learning rate of Model ID-Softmax B is set to 0. It takes 120 hours to train this model.

4.2.1 Experiments on the Feature Distribution

In this section, we evaluate the effectiveness of minimizing the intra-class distances between the ID face photos and the daily photos for all the compared models. In order to reflect the difference visually and quantitatively, we calculate the average angle between ID photos and daily photos for 4000 individuals of ID74K dataset. In Figure 4(b), we visualize the distribution of θ between ID features and daily features. It's easy to find that the average angle of our ID-Softmax model is the smallest, which intuitively proves that ID-Softmax loss is able to narrow the angle between the ID face photos and daily face photos in the feature space. The average angles are reported in Table 1. We can note that the average feature angle generated by our model is the smallest when comparing with other models in ID-daily scenario. The difference of θ distribution in daily-daily scenario is small(Figure 4(a)), and the mean θ of Model NormFace is the smallest in daily-daily scenario.

Table 1: mean θ of different models.

	ID vs Daily	Daily vs Daily
Model NormFace	72.72	**26.81**
Model CosFace	74.30	32.40
Model ID-Softmax A	61.67	33.43
Model ID-Softmax B	**53.64**	27.57

Table 2: Face verification performance of different models.

	ID vs Daily			Daily vs Daily		
TPR@FAR	1%	0.1%	0.01%	1%	0.1%	0.01%
Model NormFace	25.18%	9.23%	2.56%	93.55%	85.04%	73.20%
Model CosFace	52.73%	32.16%	19.34%	95.95%	91.87%	87.92%
Model ID-Softmax A	75.53%	57.48%	**44.32%**	96.15%	93.61%	90.67%
Model ID-Softmax B	**79.17%**	**60.88%**	43.54%	**96.86%**	**94.80%**	**92.18%**

4.2.2 Experiments on the Feature Verification

Face verification is one of the most widely used application of face recognition. For face verification, the algorithm should decide a given pair of photos is the same person or not. Generally, we use True Accept Rate(TAR) and False Accept Rate(FAR) to evaluate the performance of face verification. In our experiments, we follow the common protocol that is used for face verification evaluation. Specifically, in ID-daily scenario, we random sample ID photos and daily photos from same individual as positive pair, sample ID photos and daily photos from different individuals as negative pair. In daily-daily scenario, we random sample daily photos from same individual as positive pair, sample daily photos from different individuals as negative pair. We use cosine distance of L2 normalized face feature as comparison method.

In Figure 3, we report the Receiver Operating Characteristic (ROC) curves of different models at different scenarios. In Table 2, we report TARs under 1%, 0.1%, 0.01% FAR separately at different scenario. From these results we have following observations. Firstly, not surprisingly, the performance of daily-daily scenario is obviously better than ID-daily scenario for all models. A model that works well in daily-daily scenario may not be qualified for the ID-daily scenario. We have show the theoretical and experimental analysis in Section 3.3. Secondly, models trained by ID-Softmax(Model ID-Softmax A, Model ID-Softmax B) have large advantage in ID-daily scenario. There is a large performance gap between baseline model (Model NormFace) and ID-Softmax models. Compared with Model NormFace, the intra-class variation of Model CosFace is smaller, because the cosine margin improves the discriminative of deep model. Surprisingly, in the daily-daily scenario, models trained by ID-Softmax are better than others. Thirdly, the Model ID-Softmax B has a small advantage over the Model ID-Softmax A. The reason for the better performance may be that we update the weights synchronously during the training process of Model ID-Softmax B. It is worth noting that the performance of Model ID-Softmax A and Model ID-Softmax B is still quite comparable, and the training speed of Model ID-Softmax A is faster (same with Model CosFace). Through the experiment, we can conclude that the performance of ID-Softmax loss is more competitive, especially in the ID-daily scenario.

5 CONCLUSIONS

In this paper, we propose an novel approach named ID-Softmax to guide the deep CNNs to learn highly discriminative features in FRBID scenario, which can boost the performance of face recognition. We first revisit recent progress of face recognition losses, then give the theoretical and experimental analysis on the reason why Softmax-like losses work badly on ID-daily face recognition. In order to promote the ID-daily face recognition, we collect a large dataset ID74K, which includes 74,187 identities with corresponding ID photos and daily photos. Through the experiments, we verify the effectiveness of the proposed ID-Softmax loss. In the future, we intend to further analyze the impact of different training strategies, and study the reason why our algorithm performs better than CosFace in the daily-daily scenario.

REFERENCES

Chen, T., Li, M., Li, Y., Lin, M., Wang, N., Wang, M., Xiao, T., Xu, B., Zhang, C., and Zhang, Z. (2015). Mxnet: A flexible and efficient machine learning library for heterogeneous distributed systems. *arXiv preprint arXiv:1512.01274*.

Chen, Y., Chen, Y., Wang, X., and Tang, X. (2014). Deep learning face representation by joint identification-verification. In *International Conference on Neural Information Processing Systems*, pages 1988–1996.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

Hong, S., Im, W., Ryu, J., and Yang, H. S. (2017). Ssppdan: deep domain adaptation network for face recognition with single sample per person. In *Image Processing (ICIP), 2017 IEEE International Conference on*, pages 825–829. IEEE.

Huang, G. B., Mattar, M., Berg, T., and Learned-Miller, E. (2008). Labeled faces in the wild: A database forstudying face recognition in unconstrained environments. In *Workshop on faces in'Real-Life'Images: detection, alignment, and recognition*.

Kemelmacher-Shlizerman, I., Seitz, S. M., Miller, D., and Brossard, E. (2016). The megaface benchmark: 1 million faces for recognition at scale. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4873–4882.

Liu, W., Wen, Y., Yu, Z., Li, M., Raj, B., and Song, L. (2017). Sphereface: Deep hypersphere embedding for face recognition. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, page 1.

Liu, W., Wen, Y., Yu, Z., and Yang, M. (2016). Large-margin softmax loss for convolutional neural networks. In *ICML*, pages 507–516.

Lu, C. and Tang, X. (2014). Surpassing human-level face verification performance on lfw with gaussianface. *Computer Science*.

Schroff, F., Kalenichenko, D., and Philbin, J. (2015). Facenet: A unified embedding for face recognition and clustering. pages 815–823.

Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Sun, Y., Wang, X., and Tang, X. (2014). Deep learning face representation from predicting 10,000 classes. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1891–1898.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9.

Taigman, Y., Yang, M., Marc, and Wolf, L. (2014). Deepface: Closing the gap to human-level performance in face verification. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1701–1708.

Wang, F., Xiang, X., Cheng, J., and Yuille, A. L. (2017). Normface: l 2 hypersphere embedding for face verification. In *Proceedings of the 2017 ACM on Multimedia Conference*, pages 1041–1049. ACM.

Wang, H., Wang, Y., Zhou, Z., Ji, X., Li, Z., Gong, D., Zhou, J., and Liu, W. (2018). Cosface: Large margin cosine loss for deep face recognition. *arXiv preprint arXiv:1801.09414*.

Wen, Y., Zhang, K., Li, Z., and Qiao, Y. (2016). *A Discriminative Feature Learning Approach for Deep Face Recognition*. Springer International Publishing.

Xie, X., Cao, Z., Xiao, Y., Zhu, M., and Lu, H. (2015). Blurred image recognition using domain adaptation. In *IEEE International Conference on Image Processing*.

Yi, D., Lei, Z., Liao, S., and Li, S. Z. (2014). Learning face representation from scratch. *arXiv preprint arXiv:1411.7923*.

Zhang, K., Zhang, Z., Li, Z., and Qiao, Y. (2016). Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, 23(10):1499–1503.

Zhang, S., He, R., Sun, Z., and Tan, T. (2017). Demeshnet: Blind face inpainting for deep meshface verification. *IEEE Transactions on Information Forensics & Security*, 13(3):637–647.

Zhou, E., Cao, Z., and Yin, Q. (2015). Naive-deep face recognition: Touching the limit of lfw benchmark or not? *Computer Science*.