A Comprehensive Study on Loss Functions for Cross-Factor Face Recognition

Gee-Sern Jison Hsu
National Taiwan University of Science and Technology
Taipei, Taiwan
jison@mail.ntust.edu.tw

Hung-Yi Wu
National Taiwan University of Science and Technology
Taipei, Taiwan
m10703435@mail.ntust.edu.tw

Moi Hoon Yap
Manchester Metropolitan University
Manchester, UK
M.Yap@mmu.ac.uk

Abstract

A significant progress has been made to face recognition in recent years. The progress includes the advancement of the deep learning solutions and the availability of more challenging databases. As the performance on previous benchmark databases, such as MPIE and LFW, saturates, more challenging databases are emerging and keep driving the development of face recognition technology. The loss function considered in a deep face recognition network plays a critical role for the performance. To better evaluate the state-of-the-art loss functions, we define four challenging factors, including pose, age, occlusion and resolution with specific databases and conduct an extensive experimental study on the latest loss functions. We select the IARPA Janus Benchmark–B (IJB-B) and IARPA Janus Benchmark–C (IJB-C) for pose, the FG-Net Aging Database (FG-Net) for age, the AR Face Database (AR Face) for occlusion, and the Surveillance Cameras Face Database (SCface) for low resolution. The loss functions include the Center Loss, the Marginal Loss, the SphereFace, the CosFace and the ArcFace. Although for most factors, the ArcFace outperforms others. However, the best performance against low-resolution is achieved by the SphereFace. Another attractive finding of this study is that the cross-age performance is the lowest among the four factors with a clear margin. This highlight possible directions for future research.

1. Introduction

Face recognition is one of the challenging problems in the fields of computer vision. As the performance reported on previous benchmark databases, such as MPIE [7] and LFW [10], saturates, more challenging databases are
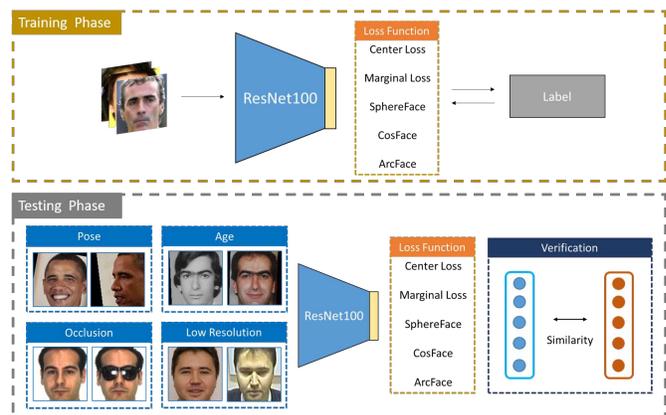

Figure 1. The flowchart of our experiments.

emerging and keep driving the development of face recognition solutions. Deep learning approaches have proven capacity of handling face recognition in a highly promising way with many breakthroughs announced in recent years. The core parts of the deep learning approaches are the deep convolutional neural networks (CNNs). Each CNN is usually composed of convolutional layers, pooling layers, normalization processing and loss functions. The loss function plays an important role in determining the performance of a CNN for face recognition.

In recent years, researchers have developed several loss functions based on the traditional softmax loss function for improving face recognition. For example, the Angular Softmax Loss [12] is defined in an angular feature space instead of the common Euclidean space so that the angular margin for measuring the inter-class variance can be computed, leading to an improvement to the recognition performance. The Large Margin Cosine Loss [16] considers a cosine margin penalty to the target logit, revealing a bet-

ter performance than the Angular Softmax Loss. The latest Additive Angular Margin Loss [4] introduces the additive angular margin penalty between the normalized features and weights, achieving a better performance than the Large Margin Cosine Loss.

The performance of the aforementioned loss functions has been verified on several benchmark databases, including the LFW, IARPA Janus Benchmark–A (IJB-A), ... Although these databases offer a wide range of variables as pose, illumination, expression and others good for performance evaluation, there are some issues deserving our attention. The first issue is the performance on different types of databases. Many facial databases contains different types facial images, such as pose, age , occlusion and low-resolution. These challenging face images cause inaccurate feature extraction and reduce the performance of the face recognition system. Understanding the impact of the above problems will help us to improve the face recognition system.

For pose, the change in the pose of the faces makes the face lack many important features, so it also makes the face recognition system unable to accurately extract features for recognition. For age, faces have different changes in different period. For example, in adolescents, the facial features change greatly in just a few months. However, in adulthood, it hasn't changed a lot in a few years, so it's interesting to understand the effectiveness of face recognition cross age. For occlusion, face occlusion has been a problem that has been discussed for many years in face recognition. Because people often wear accessories, such as sunglasses, masks, or objects blocked between the photographer and the object. The above situations all make the important features of the face be occluded. For low-resolution, face recognition systems are often used in surveillance system. Because of the distance and motion, frequently low in resolution or with blur quality. The above problems facial images can not be recognized.

In this paper, we have selected several latest loss functions, including the Center Loss [17], the Marginal Loss [5], the Angular Softmax Loss [12], the Large Margin Cosine Loss [16] and the Additive Angular Margin Loss [4], and compared their performance on five facial databases. The five facial databases contain the IARPA Janus Benchmark–B (IJB-B) [18], IARPA Janus Benchmark–C (IJB-C) [14], the FG-Net Aging Database (FG-Net) [11], the AR Face Database (AR Face) [13] , and the Surveillance Cameras Face Database (SCface) [2]. The above testing databases contain the different types of faces, including pose, age, occlusion and low-resolution,respectively. It will make us more clear the impact of challenging faces on loss functions by evaluating the 5 facial databases.

We show our flowchart in Fig. 1. The comparison is based on the same CNN architecture and only the loss function is replaced by each specific one. In the training phase, we change different loss functions to learn features with a large margin space and also consider the different setting on training datasets. In the testing phase, the face features are extracted from CNN to extract face features and compute the cosine similarity score to perform face verification. Note that although the performances of the selected loss functions are reported in their papers, they don't evaluate the performance on specific condition of facial databases. Therefore, the comparison will shows the evolution of the loss functions and how face recognition system performs on the different situations. In our result, an attractive finding of this study is that the cross-age performance is the lowest among the four factors with a clear margin and after adding in low-resolution augmented data the Angular Softmax Loss get the highter performance than the Additive Angular Margin loss. This highlight possible directions for future research.

2. Selected Loss Functions

The loss functions selected in this study include the Center Loss [17], the Marginal Loss [5], the Angular Softmax Loss [12], the Large Margin Cosine Loss [16] and the Additive Angular Margin Loss [4]. As these loss functions consider the Softmax Loss as a core reference, we introduce the Softmax Loss first and then the others.

The Softmax Loss function can be written as follows:

$$L_s = -\frac{1}{N_b} \sum_{i=1}^{N_b} \log \frac{e^{W_{y_i}^T x_i + b_{y_i}}}{\sum_{j=1}^{n} e^{W_j^T x_i + b_j}} \qquad (1)$$

where $x_i \in \mathbb{R}^d$ denotes the $d-$dim deep feature of the i-th sample, belonging to the y_i-th class. $W_j \in \mathbb{R}^d$ denotes the j-th column of the weight $W \in \mathbb{R}^{d \times n}$ and $b_j \in \mathbb{R}^n$ is the bias term. N_b and n are the batch size and the class number, respectively. The softmax loss is widely used in deep face recognition [1]. However, the softmax loss function does not optimize the feature embedding to enhance higher similarity for intra-class samples and diversity for inter-class samples. This motivates the developments of other loss functions.

2.1. Center Loss

Center loss [17] was proposed to improve the softmax loss for face verification. It learns a center for the features of each class and meanwhile tries to pull the deep features of the same class close to the corresponding center. Given the deep feature x_i in a batch, the center loss can be computed as:

$$L_{ce} = \frac{1}{2} \sum_{i=1}^{N} \|x_i - c_{y_i}\|_2^2 \qquad (2)$$

where $c_{y_i} \in \mathbb{R}^d$ is the center of class y_i. During training, the center loss encourages the instances of the same classes to be closer to a learnable class center. However, since the class centers are updated at each iteration based on a mini-batch instead of the whole dataset, the learning process can be very unstable. It has to be under the joint supervision of the softmax loss during training. Therefore, the following combined loss is considered when applying center loss:

$$
\begin{aligned}
L_c &= L_s + \lambda L_{ce} \\
&= -\sum_{i=1}^{N_b} \log \frac{e^{W_{y_i}^T x_i + b_{y_i}}}{\sum_{j=1}^{n} e^{W_j^T x_i + b_j}} + \frac{\lambda}{2} \sum_{i=1}^{N} \|x_i - c_{y_i}\|_2^2
\end{aligned}
\tag{3}
$$

where L_s is the softmax loss (1) and λ is a hyper-parameter that balances the two losses.

2.2. Marginal Loss

The Marginal Loss function [5] was proposed to simultaneously maximize the inter-class distances and minimize the intra-class variations. The Margin Loss function focuses on the marginal samples and is computed as follows:

$$
L_{ma} = \frac{1}{m^2 - m} \sum_{i,j,i\neq j}^{m} \left(\xi - y_{ij} \left(\theta - \left\| \frac{x_i}{\|x_i\|} - \frac{x_j}{\|x_j\|} \right\|_2^2 \right) \right)
\tag{4}
$$

The term $y_{ij} \in \{\pm 1\}$ indicates whether the faces x_i and x_j are from the same class or not, θ is the distance threshold to distinguish whether the faces are from the same person/class, and ξ is the error margin besides the classification hyperplane [5]. Similar to the center loss prone to be unstable in training, the Marginal Loss will also be unstable at training because of the batch normalization. It is thus computed with the joint supervision with the Softmax loss L_s, as given below:

$$
L_m = L_s + \lambda L_{ma}
\tag{5}
$$

The hyper-parameter λ balances the two losses. The coupling with the cross-entropy loss provides separable features and prevents the loss from degrading to zero [5].

2.3. Angular Softmax Loss

The Angular Softmax Loss was proposed to improve the issues with the bias $b_j = 0$ and $\|W_j\| = 1$ [12]. The issue of the bias $b_j = 0$ is handled by transforming the logit [15] as $W_j^T x_i = \|W_j\| \|x_i\| \cos\theta_j$, where θ_j is the angle between the weight W_j and the feature x_i. The issue with the individual weight $\|W_j\| = 1$ is handled by taking the l_2 normalization to make the prediction only depend on the angle between the feature vector and the weight vector. To make it discriminative, the authors generalize it to the following Angular Softmax (called A-Softmax in short) Loss L_{AS}, and name their solution "SphereFace".

$$
L_{as} = -\frac{1}{N} \sum_{i=1}^{N} \log \frac{e^{\|x_i\| \cos(m\theta_{y_i})}}{e^{\|x_i\| \cos(m\theta_{y_i})} + \sum_{j=1,j\neq y_i}^{n} e^{\|x_i\| \cos\theta_j}}.
\tag{6}
$$

where $\theta_{y_i} \in [0, \frac{\pi}{m}]$. A-Softmax loss has stronger requirements for a correct classification when $m \geq 2$, which generates an angular classification margin between the learned features of different classes. A-Softmax loss imposes a discriminative power to the learned features via angular margin, equivalent to learning features that are discriminative on a hypersphere manifold, while Euclidean margin losses learn features in Euclidean space.

2.4. Large Margin Cosine Loss

The Large Margin Cosine loss [16] was proposed to solve the issues with the A-Softmax loss. The decision boundary of the A-Softmax loss is defined over the angular space by $\cos(m\theta_1) = \cos(\theta_2)$, which can be difficult to optimize due to the non-monotonicity of the cosine function. To overcome this difficulty, the large margin cosine loss takes the normalized features as input to learn the highly discriminative features by maximizing the inter-class cosine margin. The authors define a hyper-parameter m such that the decision boundary is given by $cos(\theta_1) - m = cos(\theta_2)$, where θ_i is the angle between the feature and weight of class i. They reformulate the softmax loss as a cosine loss by applying the l_2 normalization on both the feature and weight vectors to remove radial variations, based on which a cosine margin term m is introduced to further maximize the decision margin in the angular space. The authors call their solution "CosFace". The large margin Cosine loss L_{co} is computed as follows:

$$
L_{co} = -\frac{1}{N} \sum_{i=1}^{N} \log \frac{e^{s(\cos\theta_{y_i} - m)}}{e^{s(\cos\theta_{y_i} - m)} + \sum_{j=1,j\neq y_i}^{n} e^{s\cos\theta_j}}
\tag{7}
$$

2.5. Additive Angular Margin Loss

The Additive Angular Margin Loss [4] was proposed to further improve the discriminative power of the loss considered in a face recognition model and to stabilize the training process. Following the work in [12, 16], the authors further normalize the feature and weight vectors, and coin their solution "ArcFace". The difference is that they add an additive angular margin penalty m between x_i and W_{y_i} to simultaneously enhance the intra-class compactness and inter-class discrepancy. The Additive Angular Margin Loss L_{aa} is computed as follows:

$$
L_{aa} = -\frac{1}{N} \sum_{i=1}^{N} \log \frac{e^{s(\cos(\theta_{y_i} + m))}}{e^{s(\cos(\theta_{y_i} + m))} + \sum_{j=1,j\neq y_i}^{n} e^{s\cos\theta_j}}.
\tag{8}
$$

Despite the numerical similarity between ArcFace and previous works, the proposed additive angular margin has a better geometric attribute as the angular margin has the exact correspondence to the geodesic distance. It is shown in [4] that the ArcFace has a constant linear angular margin throughout the decision boundaries when handling binary classification; however, the SphereFace and CosFace only give a nonlinear angular margin.

3. Experiments

3.1. Training/Testing Datasets

3.1.1 Training Dataset

MS-Celeb-1M. The MS-Celeb-1M [8] can be the largest face recognition dataset in the world. It is a dataset of 10 million face images for 100K subjects collected from the Internet. The majority of subjects in this dataset are American and British actors, journalists, artists, musicians and so on. However, this dataset has many mislabeled images, and we have cleaned it and extracted 5.8 million face images for 85K subjects by applying a semi-automatic procedure.

3.1.2 Testing Datasets

IJB-B and IJB-C. The IJB-B [18] and IJB-C [14] can be two of the most mainstream face databases. IJB-B [18] has 76.8K face images from 1,845 individuals that form 12,115 templates with 10,270 intra (same-person) pairs and 8M extra (different-person) pairs. IJB-C [14] has 148.8K face images from 3,531 individuals that form 23,124 templates with 19,557 intra pairs and 15,639K extra pairs. These two datasets contain face images of different conditions regardless of subject conditions (pose, expression, occlusion) or acquisition conditions (illumination, standoff, etc.). The diversity of face images can verify how a face recognition system performs in a variety of environments.

FG-Net. The FG-Net Aging Database [11] contains 1002 images from 82 different subjects with ages ranging between newborns to 69 years old subjects. Each subject has 6-18 face images at different ages. In our cross-age face verification, we randomly form 490,545 pairs, including 5,693 intra pairs and 484,852 extra pairs. The intra pairs refer to two images from same subjects, but their ages are different. The extra pairs refer to two images from different subjects and their ages can either be the same or different.

AR Face. The AR Face Database [13] contains over 4,000 color images corresponding to 126 people's faces. Each person participated in two sessions, separated by two weeks (14 days) time. There are many features like different facial expressions, illumination conditions, and occlusions (sun glasses and scarf). In this paper, we only use AR Face [13] with natural occlusions, including wearing glasses and scarfs to test the performance of occluded face recognition.

Figure 2. The samples in IJB-B [18] and IJB-C [14]. Each row denotes one subjects.

Figure 3. The samples in FG-Net [11]. Each pair denotes one cross-age subjects.

Because the author didn't provide the testing protocol of face verification, we select all 2268 images with the neutral expression, sun glasses and scarf occlusion to form the 5,309 intra pairs and 832,402 extra pairs. The intra pairs refer to two images from same subjects, but the occluded-part of these faces are different.

SCface. The Surveillance Cameras Face Database [2] is the true surveillance face recognition benchmark. This dataset contains 4,160 face images of 130 identities captured in uncontrolled indoor environment using five video surveillance cameras of various qualities. In the testing set, We follow the protocols in [2] and the testing dataset contains 688 images from 43 identities which are taken at distances of 4.20, 2.60 and 1.00 meters. The numbers of the intra pairs and extra pairs are 688 and 27,090 pairs, respectively.

Figure 4. The samples in AR Face [13]. Each row denotes one subjects.

Figure 5. The samples in SCface [2]. Each pair denotes one subjects in different resolution.

3.2. Experiment Setups

As mentioned in the beginning of this paper, although the above state-of-the-art loss functions are reportedly assessed in the works that proposed them, the differences in the training/testing databases, the network architectures and settings, the pre-processing and other parameters make the comparison hard to confirm. We, therefore, consider a unified architecture with the ResNet-100 [9] as the feature embedding network with the same pre-processing and best settings as reported in [4] and only change the loss function for the experiments.

Our study focuses on the challenges from the four variables: pose, age, occlusion, and low resolution for face verification. In all experiments, we use the MS-Celeb-1M [8] as the basic training dataset. For the experiments on occlusion and low-resolution, we use the model trained on MS-Celeb-1M as a pretrained model and retrain it on the data augmented in occlusion and low-resolution. For performance evaluation, we use the IJB-B [18] and IJB-C [14] for pose, the FG-Net [11] for age, the AR Face [13] for occlusion and the SCface [2] for low resolution. Additionally, because the faces in the AR Face are all frontal. In order to better understand the impact of occluded face images, we also use IJB-B [18] as another performance evaluation criterion for occlusion.

Preprocessing. For preprocessing the face images, we employ the MTCNN [19] to detect the faces and landmarks. Five landmarks, including two eyes, nose and two mouth corners, are used to crop each face, and then normalize the size to 112×112. In the testing, we compute the cosine distance of two features to obtain the similarity score.

Pose. Large-pose face recognition is a very challenging problem in this field. Because the IJB-B [18] and the IJB-C [14] datasets have many face images with various poses, we choose these two databases for performance evaluation. For training, we use the MS-Celeb-1M [8] as the train-

ing dataset. Resenet-100 was selected as the main feature embedding network and only change the loss function. Finally, face recognition was evaluated by calculating the cosine distance between the two features.

Age. Recognizing a cross-age face can be very challenging. In the training phase, we use the original MS-Celeb-1M training set. To evaluate the performance, we choose the FG-Net [11] as the cross-age testing database.

Occlusion. Partial occlusion is a common challenging factor for face recognition. In addition to the original MS-Celeb-1M training set, we add in the occlusion-augmented MS-Celeb-1M data for training. The performance is evaluated on the IJB-B with partial occlusion and the AR database. The procedure for making the occlusion-augmented data is shown in Fig. 6. Given the MTCNN-detected 5 landmarks, we cover the eyes by a 20×60 black block and the mouth by a 20×50 black block. The same procedure is also applied to the IJB-B for making data for evaluation.

Resolution. Recognizing a low-resolution face can be very challenging, but this is often the case with the faces captured from surveillance cameras. Due to the distance and motion, the faces captured in surveillance cameras are frequently low in resolution or with blur quality. To evaluate the performance, we retrain the MS-Celeb-1M-pretrained model by adding in low-resolution augmented data. Similar to the above partial occlusion, we change the resolution of the data in the MS-Celeb-1M from 112×112 (original) to 56×56, 28×28, and 14×14 by using bilinear interpolation. The testing dataset is the SCface [6].

The programs are written in Python with the MXNet deep learning framework [3]. We use the batch size 64 and train the networks on a Ubuntu 18.04 with Titan X GPU, and CUDA 9.0 with cuDNN 7.6. The learning rate starts from 0.1 and is divided by 10 at the 8th, 12th and 16th epochs.
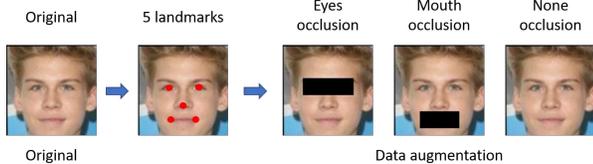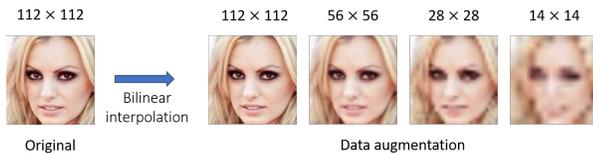
Figure 6. Data augmentation for occlusion.

Figure 7. Data augmentation for low resolution.

3.3. Evaluation Results

Pose. Table 1 and Table 2 show the verification rate in TAR on the IJB-B and IJB-C respectively. The ArcFace with the Additive Angular Margin loss outperforms all, followed by the CosFace with the Large Margin Cosine loss, then the SphereFace with the Angular Softmax loss, then the Marginal loss and the Center loss. Note that the TARs in (·) are from [4] which differ from our results in a small margin.

Age. The cross-age performance on FG-Net is shown in Table 3. The ArcFace outperforms all, followed by the SphereFace, and then the CosFace. It shows that the Angular Softmax loss can be better than the Large Margin Cosine loss for demonstrating the robustness against age-related appearance variation. An attractive finding of this study is that the cross-age performance is the lowest among the four factors with a clear margin. This highlights a potential direction for the future research in this field.

Occlusion. Table 4 shows the performance on the AR Face Database. It can be observed that the ArcFace also gets the best performance on the face occlusion dataset, but the SphereFace also has higher performance than the CosFace. After adding data augmentation, the performance of all models is slightly improved, which proves that this method of data augmentation effectively improve the recognition rate of the face recognition system under the occlude-face. In addition, we find that the SphereFace has greatly improved about 14.7% at 0.001% FAR after using data augmentation. Table 5 and Table 6 show the performance on the IJB-B. We respectively test the model on the eyes and mouth occlusion of the IJB-B. The result shows in face verification extract the feature from the eyes is more accurate than the mouth and no matter in Table 5 and Table 6 the ArcFace outperforms all.

Low-resolution. To get a understanding of face recognition system's difference on low resolution dataset, we give the detailed performance of the SCface under differ-

ent loss functions and training data in Table 5. Before adding the data augmentation, the ArcFace also gets the best performance. However, after using the data augmentation, the SphereFace unexpectedly achieves the hightest performance. The above experimental data shows we should use the SphereFace as the loss function in face recognition of low resolution and the data augmentation is effective.

Model	TAR(%)@FAR			AUC (%)
	0.01%	0.001%	0.0001%	
Center Loss [17]	88.9	80.2	68.3	98.7
Marginal Loss [5]	90.1	82.5	72.6	98.9
SphereFace [12]	94.3	91.4	81.3	99.6
CosFace [16]	95.9	92.6	89.1	99.4
ArcFace [4]	**97.4**	**94.9**	**92.6 (94.2)**	**99.5**

Table 1. Verification rate (in % TAR) for state-of-the-art loss functions tested on the IJB-B. (·) shows the performance reported in [4].

Model	TAR(%)@FAR			AUC (%)
	0.01%	0.001%	0.0001%	
Center Loss [17]	90.4	83.5	74.1	98.9
Marginal Loss [5]	92.6	87.4	79.9	99.1
SphereFace [12]	96.8	91.7	86.1	99.6
CosFace [16]	96.8	93.3	90.2	99.5
ArcFace [4]	**97.9**	**96.1**	**93.6 (95.6)**	**99.6**

Table 2. Verification rate (in % TAR) for state-of-the-art loss functions tested on the IJB-C. (·) shows the performance reported in [4].

Model	TAR(%)@FAR			AUC (%)
	0.1%	0.01%	0.001%	
SphereFace [12]	86.1	65.6	43.6	95.1
CosFace [16]	84.1	56.6	33.7	94.2
ArcFace [4]	**89.7**	**71.3**	**52.3**	**96.3**

Table 3. Verification rate (in % TAR) for state-of-the-art loss functions tested on the FG-Net.

4. Conclusion

As new loss functions and benchmark databases keep emerging, there must be an evaluation that reports the latest assessment every once in a short period of time. Verified on five latest benchmark datasets, we have compared the four challenging factors, including pose, age, occlusion and resolution with the state-of-the-art loss functions. Although for most factors, the ArcFace outperforms others. However, the best performance against low-resolution is achieved by the SphereFace. Another attractive finding of this study is that

Model	TAR(%)@FAR			AUC (%)
	0.1%	0.01%	0.001%	
SphereFace [12]	97.2	83.3	71.4	98.9
CosFace [16]	94.3	79.6	72.7	98.2
ArcFace [4]	99.5	92.7	81.3	99.6
SphereFace$_{occ}$	99.6	95.3	86.1	99.7
CosFace$_{occ}$	96.7	83.7	75.1	98.8
ArcFace$_{occ}$	**99.8**	**98.0**	**91.3**	**99.9**

Table 4. Verification rate (in % TAR) for state-of-the-art loss functions tested on the AR Face. $_{occ}$ means the model fine-tune on the MS-Celeb-1M of occluded data augmentation.

Model	TAR(%)@FAR			AUC (%)
	0.01%	0.001%	0.0001%	
SphereFace [12]	85.9	55.4	20.8	99.1
CosFace [16]	89.2	76.7	59.7	98.6
ArcFace [4]	94.9	89.8	81.4	99.3
SphereFace$_{occ}$	94.0	83.1	60.3	99.5
CosFace$_{occ}$	92.2	82.6	68.4	98.8
ArcFace$_{occ}$	**96.7**	**93.4**	**89.9**	**99.5**

Table 5. Verification rate (in % TAR) for state-of-the-art loss functions tested on the eyes occlusion of the IJB-B. $_{occ}$ means the model fine-tune on the MS-Celeb-1M of occluded data augmentation.

Model	TAR(%)@FAR			AUC (%)
	0.01%	0.001%	0.0001%	
SphereFace [12]	93.8	80.8	56.8	99.5
CosFace [16]	93.9	87.1	77.8	99.3
ArcFace [4]	96.4	92.3	87.8	99.5
SphereFace$_{occ}$	96.2	90.1	79.8	**99.6**
CosFace$_{occ}$	94.9	89.1	81.7	99.3
ArcFace$_{occ}$	**97.1**	**94.8**	**91.0**	99.5

Table 6. Verification rate (in % TAR) for state-of-the-art loss functions tested on the mouth occlusion of the IJB-B. $_{occ}$ means the model fine-tune on the MS-Celeb-1M of occluded data augmentation.

Model	TAR(%)@FAR			AUC (%)
	0.1%	0.01%	0.001%	
SphereFace [12]	91.2	75.8	65.4	97.2
CosFace [16]	93.2	79.2	69.6	97.6
ArcFace [4]	95.2	84.3	74.3	98.2
SphereFace$_{low}$	**97.9**	**89.2**	74.7	**99.3**
CosFace$_{low}$	97.7	88.4	78.5	99.1
ArcFace$_{low}$	97.5	88.2	**79.1**	99.1

Table 7. Verification rate (in % TAR) for state-of-the-art loss functions tested on the SCface. $_{low}$ means the model fine-tune on the MS-Celeb-1M of low resolution data augmentation.

the cross-age performance is the lowest among the four factors with a clear margin. This highlight possible directions for future research. The above result shows that analyze the loss functions in different factors, which may lead to a better design of novel loss function and better use of the existing ones.

References

[1] Qiong Cao, Li Shen, Weidi Xie, Omkar M Parkhi, and Andrew Zisserman. Vggface2: A dataset for recognising faces across pose and age. In *FG*. IEEE, 2018. 2

[2] Rafael Ceschin, Alexandria Zahner, William Reynolds, Jenna Gaesser, Giulio Zuccoli, Cecilia W Lo, Vanathi Gopalakrishnan, and Ashok Panigrahy. A computational framework for the detection of subcortical brain dysmaturation in neonatal mri using 3d convolutional neural networks. *NeuroImage*, 178:183–197, 2018. 2, 4, 5

[3] Tianqi Chen, Mu Li, Yutian Li, Min Lin, Naiyan Wang, Minjie Wang, Tianjun Xiao, Bing Xu, Chiyuan Zhang, and Zheng Zhang. Mxnet: A flexible and efficient machine learning library for heterogeneous distributed systems. *arXiv preprint arXiv:1512.01274*, 2015. 5

[4] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *CVPR*, 2019. 2, 3, 4, 5, 6, 7

[5] Jiankang Deng, Yuxiang Zhou, and Stefanos Zafeiriou. Marginal loss for deep face recognition. In *CVPRW*, 2017. 2, 3, 6

[6] Mislav Grgic, Kresimir Delac, and Sonja Grgic. Scface–surveillance cameras face database. *Multimed. Tools. Appl.*, 51(3):863–879, 2011. 5

[7] Ralph Gross, Iain Matthews, Jeffrey Cohn, Takeo Kanade, and Simon Baker. Multi-pie. *Image and Vision Computing*, 28(5):807–813, 2010. 1

[8] Yandong Guo, Lei Zhang, Yuxiao Hu, Xiaodong He, and Jianfeng Gao. Ms-celeb-1m: A dataset and benchmark for large-scale face recognition. In *ECCV*. Springer, 2016. 4, 5

[9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, 2016. 5

[10] Gary B Huang, Marwan Mattar, Tamara Berg, and Eric Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. 2008. 1

[11] Andreas Lanitis, Christopher J. Taylor, and Timothy F Cootes. Toward automatic simulation of aging effects on face images. *TPAMI*, 24(4):442–455, 2002. 2, 4, 5

[12] Weiyang Liu, Yandong Wen, Zhiding Yu, Ming Li, Bhiksha Raj, and Le Song. Sphereface: Deep hypersphere embedding for face recognition. In *CVPR*, 2017. 1, 2, 3, 6, 7

[13] Aleix M Martinez. The ar face database. *CVC Technical Report24*, 1998. 2, 4, 5

[14] Brianna Maze, Jocelyn Adams, James A Duncan, Nathan Kalka, Tim Miller, Charles Otto, Anil K Jain, W Tyler Niggel, Janet Anderson, Jordan Cheney, et al. Iarpa janus benchmark-c: Face dataset and protocol. In *ICB*. IEEE, 2018. 2, 4, 5

[15] Gabriel Pereyra, George Tucker, Jan Chorowski, Łukasz Kaiser, and Geoffrey Hinton. Regularizing neural networks by penalizing confident output distributions. *arXiv preprint arXiv:1701.06548*, 2017. 3

[16] Hao Wang, Yitong Wang, Zheng Zhou, Xing Ji, Dihong Gong, Jingchao Zhou, Zhifeng Li, and Wei Liu. Cosface: Large margin cosine loss for deep face recognition. In *CVPR*, 2018. 1, 2, 3, 6, 7

[17] Yandong Wen, Kaipeng Zhang, Zhifeng Li, and Yu Qiao. A discriminative feature learning approach for deep face recognition. In *ECCV*. Springer, 2016. 2, 6

[18] Cameron Whitelam, Emma Taborsky, Austin Blanton, Brianna Maze, Jocelyn Adams, Tim Miller, Nathan Kalka, Anil K Jain, James A Duncan, Kristen Allen, et al. Iarpa janus benchmark-b face dataset. In *CVPRW*, 2017. 2, 4, 5

[19] Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, and Yu Qiao. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, 23(10):1499–1503, 2016. 5