American Journal of Neural Networks and Applications
2025, Vol. 11, No. 1, pp. 1-10
https://doi.org/10.11648/j.ajnna.20251101.11

SciencePG
Science Publishing Group

Research Article

An Improved Adaptive Angular Margin Loss Function for Deep Face Recognition

Kwang-Uk Han, Song-Jun Yun, Chol Song, Kwang-Min Kim, Chol-Jun O*

Institute of Mathematics, State Academy of Sciences, Pyongyang, Democratic People's Republic of Korea

Abstract

In recent years, there have been growing interests on deep learning based face recognition which currently produces state of the art standards in face detection, recognition and verification tasks. As is well known, loss function for extracting face feature plays a crucial role in deep face model. In this regards, margin-based loss functions which apply a fixed margin between the feature and the weight have attracted many interests. However, such margin-based losses have a somewhat limitation in enhancing the discriminative power and generalizability of the face model, since the intra-class and inter-class variations in the real face training sets are often imbalanced. In particular, the embedding feature whose angle between the feature and the weight is distributed around 90° or 180° on the hypersphere reflects the difficult embedding feature in the process of classes. These phenomena occur when one considers those class which contains few number of embedding data. In order to address this problem, in this paper we propose an improved adaptive angular margin loss that incorporates the adaptive and robust angular margin on the angular space between the feature and the corresponding weight instead of constant margin. Our new margin loss function is constructed by incorporating adaptive and more robust angular margin constraint on angular space between the embedding feature and the corresponding weight. The proposed loss function improves the feature discrimination by minimizing the intra-class variation and maximizing the inter-class variation simultaneously. We present some experimental result on LFW, CALFW, CPLFW, AgeDB and MegaFace benchmarks, which demonstrate the effectiveness of the proposed approach.

Keywords

Face Recognition, Class Imbalance, Angular Margin, Softmax Loss

1. Introduction

Face recognition is a rapidly developing and increasingly broadening field of biometric technologies [19, 24]. Its applications are wide, ranging from law enforcement to consumer applications. The recent advancement of powerful GPUs and the creation of huge face databases encouraged the development of Deep Convolutional Neural Networks (DCNNs) for face recognition tasks [19]. Deep learning based Face recognition system involves mapping the normalized face image into a feature vector (embedding). Precisely, DCNNs map the face image into a feature that having small intra-class and large inter-class distance. The design of appropriate loss function is a fundamental issue for DCNN-based face recognition. For this reason, much efforts have been devoted to creating novel loss functions to make

*Corresponding author: ocj1989@star-co.net.kp (Chol-Jun O)

Received: 10 February 2025; **Accepted:** 19 March 2025; **Published:** 17 April 2025

features not only more separable but also discriminative [3, 15, 23].

The existing loss functions which supervise the learning of network can be largely classified into two categories: one trains a multi-class classifier which can separate different identities in the training set, such as a softmax loss [2] and another learns directly an embedding, such as the triplet loss [19]. Among those, softmax loss is known to be easy to train, and have achieved excellent performances in large-scale face recognition tasks. However, the softmax loss does not explicitly optimize the feature embedding, which is not appropriate for intra-class appearance variations (e.g., pose variations [32] and age gaps [33]) and large-scale test scenarios [10].

To overcome this limitation, some researchers proposed several variants of softmax loss in the last decade (see [5, 9, 11, 12, 13, 14, 18, 22, 25, 26] and references therein). In particular, the margin-based methods which add some constant margin penalties on the angular or cosine spaces between the feature and its corresponding weight have achieved the state of the art performance. These methods aimed at enforcing intra-class compactness and inter-class discrepancy of extracted face features. The representative margin-based losses are ArcFace [3, 4], SphereFace [15] and CosFace [23]. It is remarkable that these losses make the features more discriminative over the traditional softmax loss.

However, marginal penalty losses such as ArcFace, CosFace and SphereFace assume that the samples are equally distributed in the embedding space around the class center, which is not true when dealing with largely intra-class variations. Since these methods use a constant margin, the discriminative power of deeply learned feature in such case does not become optimal.

Thus, there appeared numerous approaches to use dynamic margins. For instance, Boutros et al [1] proposed ElasticFace that relaxes the fixed single margin value by deploying a random margin drawn from the Gaussian distribution. In order to handle the class imbalance and softmax saturation issue during the training process, Zhang et al [29] proposed a class-variant margin loss (CVM-Loss) by incorporating a true-class margin and a false-class margin into the cosine space of the angle between feature vector and class weight vector. However, it should be noted that the cosine value considering the margin in CVM-Loss has an invalid value if the angle between the feature and its corresponding weight is in some range. This is clear from its mathematical expression (see (5)).

Although some variants of margin-based methods are effective, there still remains some problems to be discussed in more detail. The intra-class and inter-class variations in the real face training sets are inevitable, and this leads to degradation of face recognition performance. Geometrically, the embedding feature whose angle between the feature and the weight is distributed around 90° or 180° on the hypersphere reflects the difficult embedding feature in the process of classes. These phenomena occur when we consider those class which contains few number of embedding data. Obviously, the discriminative power of face recognition system will be improved if we learn these embedding features.

The purpose of this paper is to address this problem. More precisely, in this paper we propose an adaptive angular margin loss function that incorporates the adaptive and more robust angular margin constraint on angular space between the feature and its corresponding weight. This adaptive margin constraint makes it possible embedding features in the same class more compact and those belonging different classes farther apart. The detailed information can be found in section 3.

We present some experimental result to show the effectiveness of our approach. The experiments were conducted on some face recognition benchmarks: LFW, CALFW, CPLFW, AgeDB and MegaFace. The results demonstrate (a) our method is comparable to the state of the art methods in the face recognition performance, (b) a high generalizability of the proposed approach in spite of changes in backbone architecture, training datasets, and evaluation benchmarks.

The rest of present paper is organized as follows. In Section 2, we introduce some background concepts and terminologies, and then we briefly summarize the previous research on various loss functions for deep face model. In Section 3, we shall present an adaptive angular margin loss in detail, from its motivation, intuition, and formulation to the discussion. In Section 4, we experimentally demonstrate the effectiveness of the proposed loss, including the effect of its parameters and the superiority of its performance compared to the other losses. Section 5 is devoted to concluding the main contribution of this paper.

2. Related Works

Let us briefly summarize the related research in the field of margin-based losses. We limit ourselves to the classical softmax loss and its variants. The most widely used classification loss function, softmax loss L_s being defined as a cross entropy loss between the output of the activation function and the ground-truth, is formulated as follows:

$$L_s = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{W_{y_i}^T x_i + b_{y_i}}}{\sum_{j=1}^{N} e^{W_j^T x_i + b_j}}, \tag{1}$$

where $x_i \in R^d$ denotes the deep feature of the i-th sample belonging to the y_i-th class. $W_{y_i}, W_j \in R^d$ denote the y_i-th and j-th column of the weight $W \in R^{d\times C}$, and b_{y_i}, b_j is the bias term. The batch size and the class number are N and C, respectively. In practice, we usually set $b_{y_i} = 0$ and

$b_j = 0$.

For optimimality of classification result, one often fixes the embedding feature $\| x_i \|$ by ℓ_2 normalization and rescale it to s. Then the learned embedding features are distributed on a hypersphere with a radius of s, and softmax loss is transformed into normalized softmax loss function L_{ns}, which is defined as follows.

$$L_{ns} = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{s(\cos\theta_{y_i})}}{e^{s(\cos\theta_{y_i})} + \sum_{j=1, j\neq y_i}^{C} e^{s\cos\theta_j}}. \tag{2}$$

Later on, many works have attempted to improve the softmax loss to obtain effective margin discriminative features [5, 9, 11, 12, 13, 14, 18, 22, 25, 26]. Liu et al [15] introduced a multiplicative angular margin loss named SphereFace. Wang et al [23] proposed CosFace: additive cosine margin on the cosine angle between the deep features and their corresponding weights. Deng et al [3] proposed additive angular margin termed ArcFace by deploying angular penalty margin on the angle between the deep features and their corresponding weights.

By combining SphereFace, ArcFace and CosFace, Deng et al [4] introduced an integrated loss function $L_{Integrated}$, which is formulated as.

$$L_{Integrated} = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{s(\cos(m_1\theta_{y_i}+m_2)-m_3)}}{e^{s(\cos(m_1\theta_{y_i}+m_2)-m_3)} + \sum_{j=1, j\neq y_i}^{C} e^{s\cos\theta_j}}, \tag{3}$$

where m_1, m_2 and m_3 as the hyper-parameters.

Note that if we set $m_2 = m_3 = 0$ in (3), then it reduces to SphereFace, if we set $m_1 = m_2 = 0$ in (3), then it reduces to CosFace, and if we set $m_1 = m_3 = 0$ in (3), then it reduces to ArcFace, respectively. Furthermore, if we set $m_1 = 1$ and $m_2 = m_3 = 0$ in (3), then it reduces to normalized softmax loss function L_{ns}.

Recently, dynamic marginal penalty has proven to be more discriminative over constant marginal constraint. One of the dynamic marginal penalty is ElasticFace, which is defined as follows.

$$L_{Earc} = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{s(\cos(\theta_{y_i}+E(m,\sigma)))}}{e^{s(\cos(\theta_{y_i}+E(m,\sigma)))} + \sum_{j=1, j\neq y_i}^{C} e^{s\cos\theta_j}}, \tag{4}$$

where $E(m,\sigma)$ is a normal function that return a random value from a Gaussian distribution with the mean m and the standard deviation σ. This aims at giving the decision boundary chance to extract and retract to allow space for flexible class separability learning. This loss demonstrated the superiority of ElasticFace loss over ArcFace and CosFace losses, using the same geometric transformation, on a large set of mainstream benchmarks. However, the generated random margin in each training iteration is independent of the angle between the feature and its corresponding weight (see (4)).

Another notable dynamic marginal penalty is class-variant margin (CVM) normalized softmax loss [29], which is defined as follows.

$$L_{cvm} = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{s(\cos\theta_{y_i}-h(\theta_{y_i}))}}{e^{s(\cos\theta_{y_i}-h(\theta_{y_i}))} + \sum_{j=1, j\neq y_i}^{C} e^{s(\cos\theta_j+g(\theta_j))}}, \tag{5}$$

$$h(\theta_{y_i}) = m_1(1-\cos^2\theta_{y_i}), \tag{6}$$

$$g(\theta_j) = m_2(\cos^2\theta_j), \tag{7}$$

where $h(\theta_{y_i})$ is the margin function applied to the cosine of angle between the feature vector and the true class weight vector, named the true-class margin; $g(\theta_j)$ is the margin function added to the cosine of angle between the feature vector and the false class weight vector, named the false-class margin; m_1 and m_2 are predetermined hyperparameters; m_1 represents the upper bound of the true-class margin, and m_2 represents the upper bound of the false-class margin.

It is worth mentioning that there are other remarkable margin addptive dynamic loss functions. Xu et al [26] introduced X2-Softmax loss, which replaces the cosine function in ArcFace loss with a quadratic function. This loss is margin-adaptive and can automatically adjust the angular margin with the angle between different classes. Very recently, Khalifa el al [11] proposed a joint adaptive margins loss function termed JAMsFace that learns class-related margins for both angular and cosine spaces.

3. Proposed Approach

In this section, we propose an adaptive angular margin loss to further improve the discriminative power of deep face model. The proposed loss function aims to enhance the boundary margins whose angles are distributed around 90° or 180° on the hypersphere. In general, there exist an imbalance between different classes, especially this is true when we consider samples with large intra-class variation in training datasets or few number of embedding samples. In other words, some samples within the same class are greatly dissimilar to each other, and another samples belonging different classes are similar to each other, which strongly affects learning of features. Therefore, it is possible to enhance the intra-class compactness and inter-class discrepancy simultaneously if we design a new loss function that allows emphasize the effect of training data at the boundary between different classes.

Let us introduce new margin loss function that incorporates the adaptive and more robust angular margin constraint on angular space between the embedding feature and the corresponding weight. The proposed adaptive loss function can be formulated as follow.

$$L_{Adaptive} = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{s(\cos(\theta_{y_i}+f(\theta_{y_i})))}}{e^{s(\cos(\theta_{y_i}+f(\theta_{y_i})))}+\sum_{j=1,j\neq y_i}^{C}e^{s\cos(\theta_j)}}, \quad (8)$$

$$f(\theta_{y_i}) = m_1(1-\cos^2\theta_{y_i})+m_2, \quad (9)$$

where m_1 and m_2 are hyperparameters satisfying $m_1 \ll m_2$, whose explicit value are experimentally decided in realistic settings.

It seems that $f(\theta_{y_i})$ in (9) is similar to $h(\theta_{y_i})$ from [29], but the cosine value of angular margin is always valid, since the incorporated margin constraint depends on the angle between weights of classes and does not less than a fixed constant, namely m_2. Notice that if we set $m_1 = 0$ in (9), then it reduces to ArcFace. Furthermore if we set $m_1 = m_2 = 0$ in (9), then it reduces to the normalized softmax loss.

Let us consider mathematical meaning of marginal function. By definition, it is immediate that $f(\theta_{y_i}) \geq m_2$ for all θ_{y_i}. Let us be more precise. If the marginal angle θ_{y_i} is distributed around 90°, then we have $f(\theta_{y_i}) \approx m_1 + m_2$ from (9), and if θ_{y_i} is distributed around 180°, then we have $f(\theta_{y_i}) \approx m_2$ from (9). This means that the proposed loss function actively enhances the marginal features lying on the boundary between different classes, which results in increasing extra intra-class compactness and inter-class discrepancy simultaneously. Moreover, this ensures stable training of network by adding the small constraint on the feature whose angles are distributed around 180°.

The proposed marginal loss function can be easily implemented by adding the expression (9) and its derivative calculation in forward and backward propagation of a network which is supervised by general softmax loss function. Plotted in Figure 1 shows the graph of the adaptive margin $f(\theta_{y_i})$, nonlinear mapping of angle θ_{y_i}.

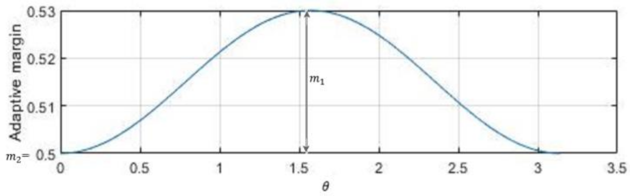
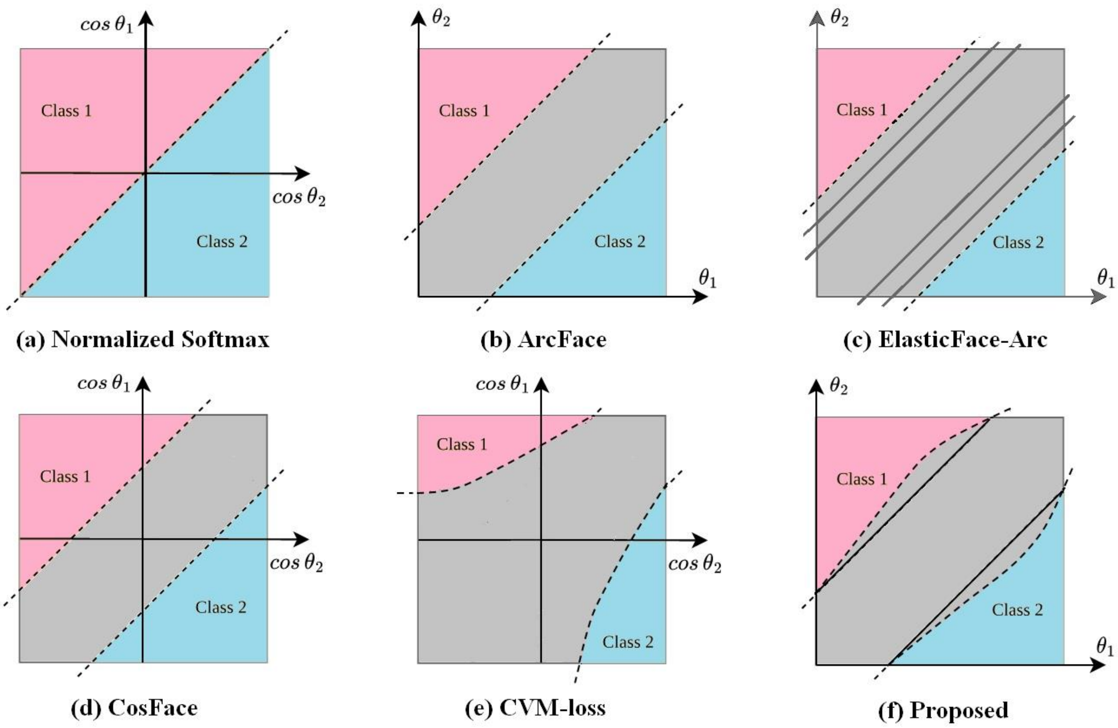


Figure 1. *Adaptive margin function* $f(\theta) = m_1(1-\cos^2\theta)+m_2$.

Now, we consider the decision boundary of adaptive angular margin loss over the previous some loss for two-class classification task. The decision boundaries of various loss functions are listed in the following Table 1.

Table 1. *Decision boundaries of some loss functions in two-class classification case.*

Loss Function	Decision Boundaries
Normalized Softmax	$\cos\theta_1 - \cos\theta_2 = 0$
ArcFace	$\cos(\theta_1 + m) - \cos\theta_2 = 0$
CosFace	$\cos\theta_1 - \cos\theta_2 - m = 0$
ElasticFace-Arc	$\cos(\theta_1 + E(m,\sigma)) - \cos\theta_2 = 0$
CVM-Loss	$\cos\theta_1 - \cos\theta_2 - (m_1\sin^2\theta_1 + m_2\cos^2\theta_2) = 0$
The proposed loss	$\cos(\theta_1 + m_1(1-\cos^2\theta_1) + m_2) - \cos\theta_2 = 0$



a) Normalized softmax, b) ArcFace, c) Elastic Arc, d) CosFace, e) CVM-Loss, f) The proposed loss

Figure 2. *Decision boundary and margins for two class classification task.*

Figure 2 illustrates the decision boundary of Normalized Softmax, ArcFace, ElasticFace-Arc, CosFace, ElasticFace-Cos and the proposed adaptive angular margin loss. The dashed black line is the decision boundary. The gray area illustrates the decision margin.

4. Experimental Results

To demonstrate the effectiveness of our proposed adaptive angular margin loss, we conducted extensive experiments on LFW, CALFW, CPLFW, AgeDB-30 and MegaFace datasets, which are the most widely used benchmarks for face recognition.

4.1. Results on LFW, CALFW, CPLFW and AgeDB-30

First, we compared the accuracy (Acc) for the adaptive angular margin loss with the various losses. Acc is a simplified metric introduced by Labelled Faces in the Wild (LFW) [7], which represents the percentage of correct classifications. LFW is the commonly used benchmark for face recognition in unconstrained environments [8]. The original LFW protocol includes 3,000 genuine and 3,000 impostor face pairs and evaluates the mean verification accuracy on these 6,000 pairs. In addition to LFW, several other benchmarks are also used for face recognition evaluations. These include CPLFW [32], CALFW [33], and AgeDB [18].

We used the manually refined dataset from Casia-WebFace [27] as our training sets. This dataset contains images of celebrities which consist of 10554 identities and 446996 images. We used a similarity transformation based on 5 facial landmarks (eyes, a nose, and mouth corners) detected by MTCNN [28] to normalize the face image. The normalized face is then cropped and resized to 112×112 [1, 4]. During training, the RGB values are normalized from [0, 255] to [-1, 1]. These normalized faces are used as an input of network.

For a fair comparison with other losses, we employed slightly modified version of well-known CNN architecture such as ResNet50, the widely used DCNN architecture as a backbone. The proposed loss in this paper are implemented using Caffe framework. We follow the common setting [4] to set the scale parameter s to 64. We set the dimension of feature embedding to 512. All models are trained on NVIDIA Tesla P100 using the stochastic gradient descent (SGD) algorithm with a momentum of 0.9 and a weight decay of 5e-4. The learning rate starts at 0.01 and is decreased by multiplying 0.8 at the 18th and 28th epoch, and training is stopped after the 35 epochs. We also set the hyper parameters $m_1 = 0.03$ and $m_2 = 0.5$, respectively. In Table 2, we presented the changes of Acc with various selection of hyperparameters. Relying on this table, we chose $m_1 = 0.03$ and $m_2 = 0.5$ as the best possible hyperparameter.

Table 2. *Parameter selection.*

m_1	m_2=0.45			m_2=0.5			m_2=0.55		
	0.01	0.03	0.05	0.01	0.03	0.05	0.01	0.03	0.05
LFW (Acc) (%)	99.39	99.41	99.37	99.48	99.53	99.43	99.35	99.39	99.42

Table 3. *Verification comparison with state-of-the-art methods on four benchmarks (LFW, CALFW, CPLFW and AgeDB-30) reported in terms of accuracy (%).*

Loss functions	LFW	CALFW	CPLFW	AgeDB-30
Softmax	99.08	91.57	87.17	92.33
ArcFace (0.5) [4]	99.53	93.93	89.99	95.15
ArcFace (0.5)	99.46	93.89	89.98	95.05
CosFace (0.35) [23]	99.51	93.83	89.70	94.56
CosFace (0.35)	99.37	94.12	89.66	94.93
Combined Margin (1, 0.3, 0.2)	99.43	92.23	90.06	95.11
ElasticFace-Arc [1]	99.53	93.65	90.88	95.23
The proposed (0.03, 0.5)	99.53	93.81	91.03	95.25

Second, we evaluated the models on LFW, CALFW, CPLFW and AgeDB-30 benchmarks. The DCNN used in this paper is the slight modification of ResNet50 and the adaptive angular margin loss is used. We selected the datasets as our training sets: MS1MV2 [3] refined from MS-Celeb-1M dataset [6]. MS1MV2 datasets include 5.8M images of 85K different identities. We also selected custom dataset consist of 27709 identities and 219384 images. By combining these dataset, the total dataset contains 87709 identities and 6019384 images. The results evaluated in terms of recognition on LFW, CALFW, CPLFW and AgeDB-30 benchmarks are listed in Table 3. The hyperparameters used in training process are set as above. The results of the various SOTA methods evaluated on LFW, CALFW, CPLFW and AgeDB-30 benchmarks are listed in Table 3. As shown in Table 4, the proposed adaptive angular margin loss have shown the slight improvement to the previous face recognition methods.

Table 4. *Evaluation of verification accuracy for face recognition methods on LFW, CALFW, CPLFW and AgeDB-30 (%).*

Methods	#Images of Training data	LFW	CALFW	CPLFW	AgeDB-30
DeepID [20]	0.2M	99.47	-	-	-
Dynamic-AdaCos [30]	6.3M	99.73	-	-	-
FaceNet [19]	200M	99.63	-	-	-
SphereFace [15]	0.5M	99.42	-	-	-
CosFace [23]	5M	99.73	-	-	-
UniformFace [5]	6.1M	99.8	-	-	-
CircleLoss [21]	3.6M	99.73	-	-	-
CoCo Loss [16, 24]	3M	99.86	-	-	-
GroupFace [12]	5.8M	99.85	96.20	93.13	98.28

Methods	#Images of Training data	LFW	CALFW	CPLFW	AgeDB-30
ArcFace [4, 3]	5.8M/11.96M	99.83	95.45	92.08	98.15
MagFace [17]	5.8M	99.83	96.15	92.87	98.17
ElasticFace-Arc [1]	5.8M	99.82	96.17	93.28	98.35
BroadFace [13]	11.96M	99.85	-	-	-
The proposed method	6.0M	99.87	96.33	93.35	98.41

4.2. Results on MegaFace

The MegaFace dataset [10] includes 1M images of 690K different individuals as the gallery set and 100K photos of 530 unique individuals from FaceScrub as the probe set. MegaFace and MegaFace(R) [11] are used for even more rigorous evaluations. MegaFace(R) benchmark refers to the refined version of MegaFace which is manually selected. These benchmarks are publicly available. We evaluate the Rank-1 identification accuracy and the TAR (@FAR=1e-6) on both benchmarks. Rank-1 is based on what percentage of probe searches return the probe's gallery mate within the top 1 rank-ordered results. TAR (@FAR=1e-6) denotes true acceptance rates (TAR) at false acceptance rates (FAR) of 1e-6. Then, we compare the proposed method with state-of-the-art face recognition approaches and verify its effectiveness. Note that "*" indicates our implementations and the best results are indicated in bold.

We use MS1MV2 as the training dataset, which is manually refined dataset containing images of celebrities which consist of 87709 identities and 6019384 images. We employ the DCNN architecture such as ResNet100. Loss function and the hyper parameters set as above, and the learning rate initially starts at 0.01 and is decreased by multiplying 0.8 at 20th and 32nd epoch, and training is stopped after 43 epochs.

The evaluation results are shown in Table 5.

Table 5. *The evaluation results for face recognition methods on MegaFace and MegaFace (R).*

Methods	#Images of Training data	MegaFace (%)		MegaFace (R) (%)	
		Rank-1	TAR (@FAR=1e-6)	Rank-1	TAR (@FAR=1e-6)
Dynamic-AdaCos [30]	6.3M	-	-	97.41	-
FaceNet [19]	200M	70.49	86.47	-	-
SphereFace [15]	0.5M	72.729	85.561	-	-
CosFace [23]	5M	82.72	96.65	-	-
UniformFace [5]	6.1M	79.98	95.36	-	-
CircleLoss [21]	3.6M	-	-	98.50	98.73
AdaptiveFace [14]	5M	-	-	95.02	95.61
GroupFace [12]	5.8M	81.31	97.35	98.74	98.79
ArcFace [4, 3]	5.8M/	81.03/	96.98/	98.35/	98.48/
	11.96M	81.43	97.63	98.98	99.08
MC-FaceGraph [31]	18.8M	-	-	99.02	98.94
ElasticFace-Arc[1]	5.8M	80.76	97.30	98.81	98.92
BroadFace [13]	11.96M	-	-	98.70	98.95
The proposed method*	6.0M	84.11	93.50	98.82	98.91

As Figured in Table 5, the evaluation result on MegaFace shows that proposed face recognition method achieves the similar performance to the state of the art methods. Furthermore, the ArcFace [33] (TAR@ FAR=1e-6: 99.08%) and MC-FaceGraph [12] (Rank-1 Acc: 99.02%) whose size of training datasets are 11.96M and 18.8M scored the highest performance on MegaFace (R) datasets. Compared to these methods, the proposed method has as less as 2 or 3 times of that datasets, and the size of network is 1.5 times as less as the size of those network, while archives the similar performance to that methods.

5. Conclusions

As stated above, we proposed an adaptive angular margin loss function, which can effectively enhance the discriminative power of feature embedding learned via DCNNs for face recognition. We demonstrate the effectiveness of our approach through some experiments. The experimental result reveals that the proposed adaptive angular margin loss actively enhances the marginal features lying on the boundary between different classes, which results in increasing intra-class compactness and inter-class discrepancy simultaneously. Moreover, this ensures stable training of network by adding the small constraint on the feature whose angles are distributed around 180°. Specifically our method advances the state of the art face recognition performance on LFW, CALFW, CPLFW and AgeDB-30, and achieves comparable results on MegaFace benchmarks.

In the future, we will explore deep learning models for face recognition such as Swin transformer, which outperforms DCNNs by a large margin loss in the image classification. In addition, we will also explore novel loss function which is more robust to facial poses and expressions.

Abbreviations

DCNN	Deep Convolutional Neural Network
CVM-Loss	Class-variant Margin Loss
LFW	Labeled Faces in the Wild
CALFW	Cross-age Labeled Faces in the Wild
CPLFW	Cross-pose Labeled Faces in the Wild
TAR	True Acceptance Rate
FAR	False Acceptance Rate
Acc	Accuracy

Acknowledgments

The authors would like to express their sincere thanks to the editor and the anonymous referees for their valuable comments concerning the paper.

Statements and Declarations

Ethics Approval and Consent to Participate

Not applicable.

Consent for Publication

Not applicable.

Author Contributions

All authors contributed to this work in collaboration. Dr. Han and Dr. Yun designed the study, conceptual approach and proposed the adaptive angular margin loss. Dr. Song made a comparison between the theoretical and experimental results and analyzed the data. Dr. Kim and Dr. O carried out the experiment and wrote the first draft of paper. All authors read and approved the final version of manuscript.

Funding

The authors declare that no funds, grants or other supports were received during the perpetration of this research.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Boutros F, Damer N, Kirchbuchner K, Kuijper A (2022) ElasticFace: Elastic margin loss for deep face recognition. preprint arXiv:2109.09416v4.

[2] Cao Q, Shen L, Xie W, et al (2018) Vggface2: a dataset for recognizing faces across pose and age. In: 2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018), IEEE, p 67-74 https://doi.org/10.1109/FG.2018.00020

[3] Deng J, Guo J, Xue N, et al (2019) Arcface: additive angular margin loss for deep face recognition. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, p 4690-4699 https://doi.org/10.1109/CVPR.2019.00482

[4] Deng J, Guo J, Yang J, et al (2022), ArcFace: additive angular margin loss for deep face recognition, preprint arXiv: 1801.07698v4.

[5] Duan Y, Lu J, Zhou J (2019) Uniformface: learning deep equidistributed representation for face recognition. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, p3415~3424.

[6] Guo Y, Zhang L, Hu Y, et al (2016) Ms-celeb-1m: a dataset and benchmark for large-scale face recognition. In: European conference on computer vision, Springer, p 87-102.

[7] Huang GB, Mattar M, Berg T, et al (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, p 1-15.

[8] Huang GB, Ramesh M, Berg T, Learned-Miller E (2007) Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst, Tech. Rep.

[9] Jiao J, Liu W, Mo Y et al (2021) Dyn-arcface: dynamic additive angular margin loss for deep face recognition. Multimed Tools Appl 80(17): 25741-25756.

[10] Kemelmacher-Shlizerman I, Sitz SM, Miler D, 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.

[11] Khalifa A, Al-Hamadi A (2023) JAMsFace: joint adaptive margins loss for deep face recognition. Neural Computing and Applications, 35: 19025-19037. https://doi.org/10.1007/s00521-02308732-5

[12] Kim Y, Park W, Roh M, et al (2020) Groupface: learning latent groups and constructing group-based representations for face recognition, In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, p 5620~5629 https://doi.org/10.1109/CVPR42600.2020.00566

[13] Kim Y, Park W, Shin J (2020) Broadface: looking at tens of thousands of people at once for face recognition, In: European conference on computer vision https://doi.org/10.48550/arXiv.2008.06674

[14] Liu H, Zhu X, Lei Z, et al (2019b) Adaptiveface: adaptive margin and sampling for face recognition. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, p 11947-11956. https://doi.org/10.1109/CVPR.2019.01222

[15] Liu W, Wen Y, Yu Z, et al (2017) Sphereface: deep hypersphere embedding for face recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, p 212-220. https://doi.org/10.1109/CVPR.2017.713

[16] Liu Y, Li H, Wang X (2017) Rethinking feature discrimination and polymerization for large-scale recognition, preprint arXiv:1710.00870.

[17] Meng Q, Zhao S, Huang Z, Zhou F (2021) Magface: a universal representation for face recognition and quality assessment, In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. https://doi.org/10.1109/CVPR46437.2021.01400

[18] Moschoglou S, Papaioannou A, Sagonas C, et al (2017) Agedb: the first manually collected, in-the-wild age database. In: proceedings of the IEEE conference on computer vision and pattern recognition workshops, p 51-59. https://doi.org/10.1109/CVPRW.2017.250

[19] Schroff F, Kalenichenko D, Philbin J (2015) Facenet: a unified embedding for face recognition and clustering. In: Proceedings of the IEEE conference on computer vision and pattern recognition, p 815-823. https://doi.org/10.1109/CVPR.2015.7298682

[20] Sun Y, Chen Y, Wang X, et al (2014) Deep learning face representation by joint identification-verification, In: NeurIPS, p 1988~1996.

[21] Sun Y, Cheng C, Zhang Y, et al (2020) Circle loss: a unified perspective of pair similarity optimization. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, p 6397~6406. https://doi.org/10.1109/CVPR42600.2020.00643

[22] Taigman Y, Yang M, Ranzato M, et al (2014) Deepface: closing the gap to human-level performance in face verification, In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. https://doi.org/10.1109/CVPR.2014.220

[23] Wang H, Wang Y, Zhou Z, et al (2018) Cosface: large margin cosine loss for deep face recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, p 5265-5274. https://doi.org/10.48550/arXiv.1801.09414

[24] Wang M, Deng W (2021) Deep face recognition: a survey. Neurocomputing 429: 215-244. https://doi.org/10.1016/j.neucom.2020.10.081

[25] Wen Y, Zhang K, Li Z, et al (2016) A discriminative feature learning approach for deep face recognition. In: European conference on computer vision, Springer, p 499-515.

[26] Xu J, Liu X, Zhang X, et al (2023) "X2-softmax: margin adaptive loss function for face recognition", preprint arXiv:2312.05281v2.

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

[28] Zhang K, Zhang Z, Li Z et al (2016) Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Signal Process Lett 23(10): 1499-1503. https://doi.org/10.1109/ICISCE.2017.95

[29] Zhang W, Chen Y, Yang W, et al (2020) Class-variant margin normalized softmax loss for deep face recognition, IEEE Trans on Neural Networks and Learning Systems. https://doi.org/10.1109/TNNLS.2020.3017528

[30] Zhang X, Zhao R, Qiao Y, et al (2019) Adacos: adaptively scaling cosine logits for effectively learning deep face representations. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, p 10823~10832. https://doi.org/10.1109/CVPR.2019.01108

[31] Zhang Y, Deng W, Wang M, et al (2020) Global-local gcn: large-scale label noise cleansing for face recognition. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. https://doi.org/10.1109/CVPR42600.2020.00775

[32] Zheng T, Deng W (2018) Cross-pose lfw: a database for studying cross-pose face recognition in unconstrained environments. Beijing Univ Posts Telecommun Tech Rep 5:7. https://doi.org/10.48550/arXiv.1708.08197

[33] Zheng T, Deng W, Hu J (2017) Cross-age lfw: a database for studying cross-age face recognition in unconstrained environments. arXiv preprint arXiv:1708.08197.