CosFace: Large Margin Cosine Loss for Deep Face Recognition

Wang H, Wang Y, Zhou Z, et al. CosFace: Large Margin Cosine Loss for Deep Face Recognition[J]. 2018.

Abstract

- Proposed a novel loss function named large margin cosine loss (LMCL), to realize maximize inter-class variance and minimize intra-class variance.

Wang H, Wang Y, Zhou Z, et al. CosFace: Large Margin Cosine Loss for Deep Face Recognition[J]. 2018.

Softmax Loss:

$$L_s = \frac{1}{N}\sum_{i=1}^{N} -\log p_i = \frac{1}{N}\sum_{i=1}^{N} -\log \frac{e^{f_{y_i}}}{\sum_{j=1}^{C} e^{f_j}}, \qquad (1)$$

fix the bias Bj = 0 for simplicity

$$f_j = W_j^T x = \|W_j\|\|x\| \cos\theta_j, \qquad (2)$$

Normalized version of Softmax Loss (NSL):

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

The formula of Large Margin Cosine Loss (LMCL):

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

subject to

$$\begin{aligned} W &= \frac{W^*}{\|W^*\|}, \\ x &= \frac{x^*}{\|x^*\|}, \\ cos(\theta_j, i) &= {W_j}^T x_i, \end{aligned} \tag{5}$$

Wang H, Wang Y, Zhou Z, et al. CosFace: Large Margin Cosine Loss for Deep Face Recognition[J]. 2018.

Scaling Parameter s

Given the normalized learned features x and unit weight vectors W, we denote the total number of classes as C where $C > 1$. Suppose that the learned features separately lie on the surface of a hypersphere and center around the corresponding weight vector. Let P_w denote the expected minimum posterior probability of the class center (*i.e.*, W). The lower bound of s is formulated as follows:

$$s \geq \frac{C-1}{C} \ln \frac{(C-1)P_W}{1-P_W}$$

Cosine Margin m

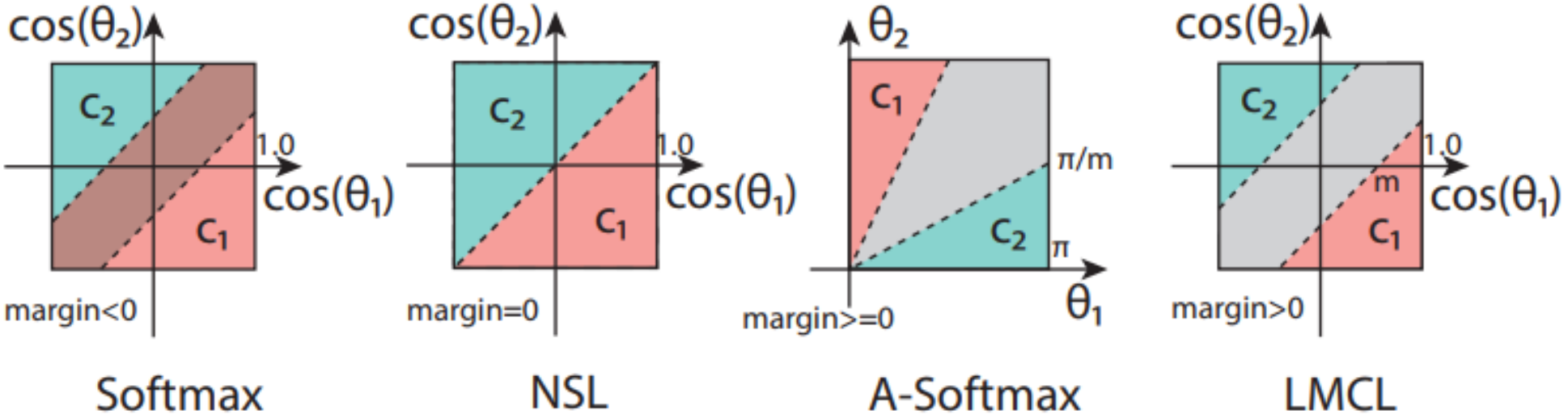
Suppose that the weight vectors are uniformly distributed on a unit hypersphere. The variable scope of the introduced cosine margin m is formulated as follows :

$$0 \leq m \leq 1 - \cos\frac{2\pi}{C}, \quad (K = 2)$$

$$0 \leq m \leq \frac{C}{C-1}, \quad (K > 2, C \leq K + 1)$$

$$0 \leq m \ll \frac{C}{C-1}, \quad (K > 2, C > K + 1)$$

where C is the total number of training classes and K is the dimension of the learned features.



Comparison of decision margins for different loss functions the binary-classes scenarios. Dashed line represents decision boundary, and gray areas are decision margins.

Wang H, Wang Y, Zhou Z, et al. CosFace: Large Margin Cosine Loss for Deep Face Recognition[J]. 2018.

Comparison on Different Loss Functions (definition of decision boundary)

- Softmax loss: $\|W_1\|\cos(\theta_1) = \|W_2\|\cos(\theta_2).$
- NSL: $\cos(\theta_1) = \cos(\theta_2).$
- A-Softmax: $C_1 : \cos(m\theta_1) \geq \cos(\theta_2),$
  $C_2 : \cos(m\theta_2) \geq \cos(\theta_1).$
- LMCL: $C_1 : \cos(\theta_1) \geq \cos(\theta_2) + m,$
  $C_2 : \cos(\theta_2) \geq \cos(\theta_1) + m.$

Wang H, Wang Y, Zhou Z, et al. CosFace: Large Margin Cosine Loss for Deep Face Recognition[J]. 2018.

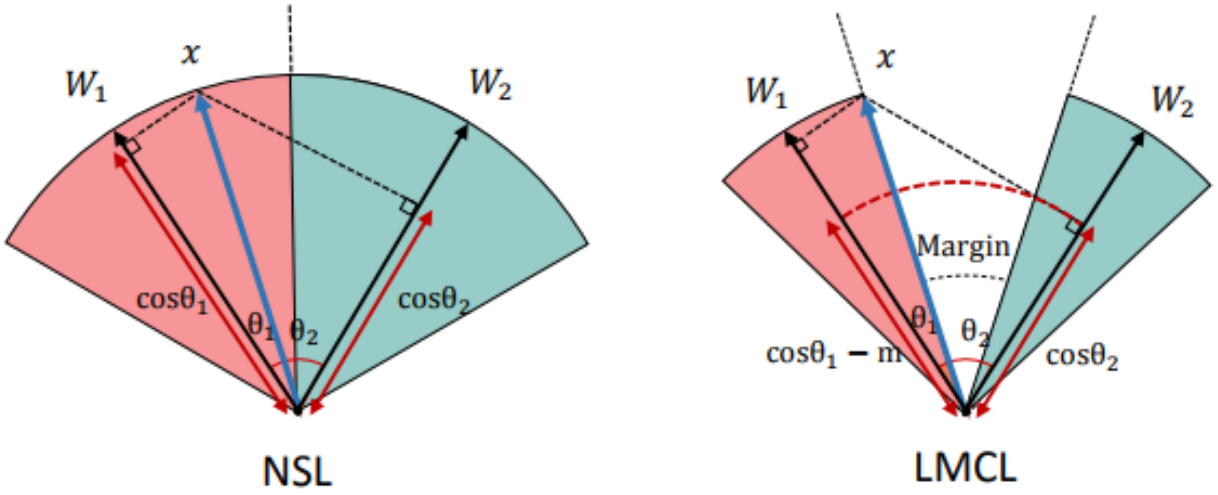
Normalization on Features

- Simultaneously normalized both weight vectors and feature vectures.

Normalization	LFW	YTF	MF1 Rank 1	MF1 Veri.
No	99.10	93.1	75.10	88.65
Yes	99.33	96.1	77.11	89.88

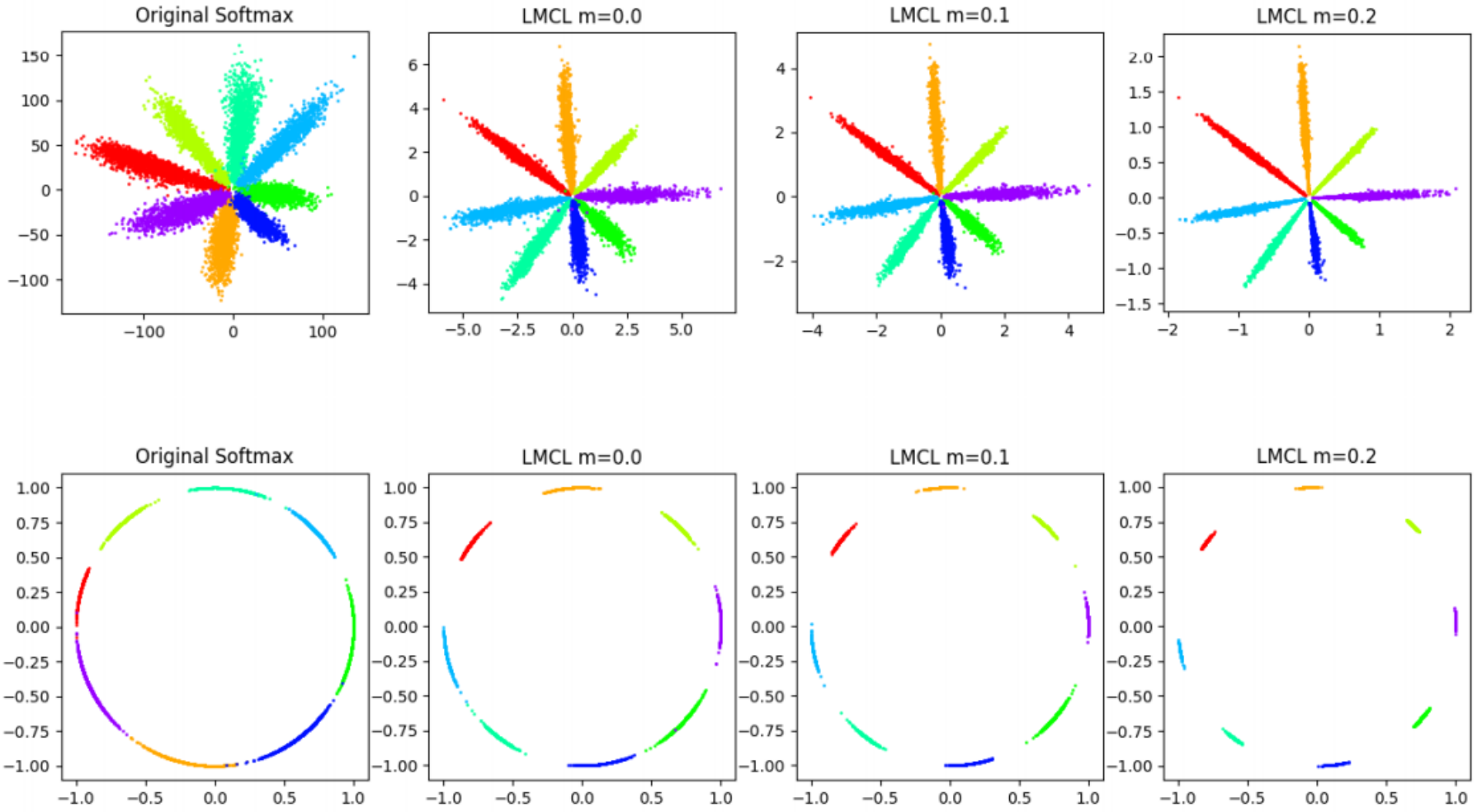
Table 1. Comparison of our models with and without feature normalization on Megaface Challenge 1 (MF1). "Rank 1" refers to rank-1 face identification accuracy and "Veri." refers to face verification TAR (True Accepted Rate) under 10^{-6} FAR (False Accepted Rate).

Wang H, Wang Y, Zhou Z, et al. CosFace: Large Margin Cosine Loss for Deep Face Recognition[J]. 2018.



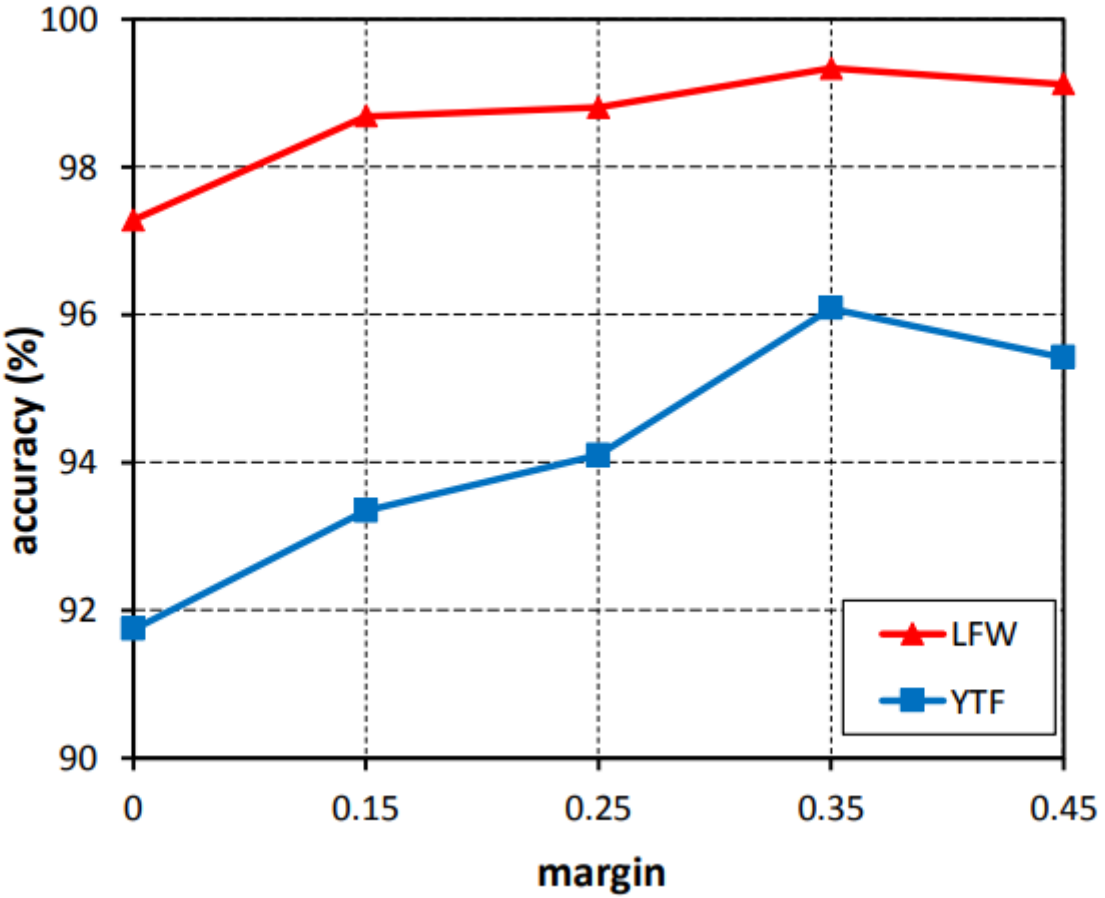
A geometrical interpretation of LMCL from feature perspective. Different color areas represent feature space from distinct classes. LMCL has a relatively compact feature region compared with NSL.

Wang H, Wang Y, Zhou Z, et al. CosFace: Large Margin Cosine Loss for Deep Face Recognition[J]. 2018.



A toy experiment of different loss functions on 8 identities with 2D features. The first row maps the 2D features onto the Euclidean space, while the second row projects the 2D features onto the angular space. The gap becomes evident as the margin term m increases.

Wang H, Wang Y, Zhou Z, et al. CosFace: Large Margin Cosine Loss for Deep Face Recognition[J]. 2018.



Accuracy (%) of CosFace with different margin parameters m on LFW and YTF.

Wang H, Wang Y, Zhou Z, et al. CosFace: Large Margin Cosine Loss for Deep Face Recognition[J]. 2018.

Overall Benchmark Comparison

Method	LFW	YTF	MF1 Rank1	MF1 Veri.
Softmax Loss [23]	97.88	93.1	54.85	65.92
Softmax+Contrastive [30]	98.78	93.5	65.21	78.86
Triplet Loss [29]	98.70	93.4	64.79	78.32
L-Softmax Loss [24]	99.10	94.0	67.12	80.42
Softmax+Center Loss [42]	99.05	94.4	65.49	80.14
A-Softmax [23]	**99.42**	95.0	72.72	85.56
A-Softmax-NormFea	99.32	95.4	75.42	88.82
LMCL	99.33	**96.1**	**77.11**	**89.88**

Table 2. Comparison of the proposed LMCL with state-of-the-art loss functions in face recognition community. All the methods in this table are using the same training data and the same 64-layer CNN architecture.

Wang H, Wang Y, Zhou Z, et al. CosFace: Large Margin Cosine Loss for Deep Face Recognition[J]. 2018.

Method	Training Data	#Models	LFW	YTF
Deep Face[35]	4M	3	97.35	91.4
FaceNet[29]	200M	1	99.63	95.1
DeepFR [27]	2.6M	1	98.95	97.3
DeepID2+[33]	300K	25	99.47	93.2
Center Face[42]	0.7M	1	99.28	94.9
Baidu[21]	1.3M	1	99.13	-
SphereFace[23]	0.49M	1	99.42	95.0
CosFace	5M	1	**99.73**	**97.6**

Table 3. Face verification (%) on the LFW and YTF datasets. "#Models" indicates the number of models that have been used in the method for evaluation.

Wang H, Wang Y, Zhou Z, et al. CosFace: Large Margin Cosine Loss for Deep Face Recognition[J]. 2018.

Method	Protocol	MF1 Rank1	MF1 Veri.
SIAT_MMLAB[42]	Small	65.23	76.72
DeepSense - Small	Small	70.98	82.85
SphereFace - Small[23]	Small	75.76	90.04
Beijing FaceAll V2	Small	76.66	77.60
GRCCV	Small	77.67	74.88
FUDAN-CS_SDS[41]	Small	77.98	79.19
CosFace(Single-patch)	Small	77.11	89.88
CosFace(3-patch ensemble)	Small	**79.54**	**92.22**
Beijing FaceAll_Norm_1600	Large	64.80	67.11
Google - FaceNet v8[29]	Large	70.49	86.47
NTechLAB - facenx_large	Large	73.30	85.08
SIATMMLAB TencentVision	Large	74.20	87.27
DeepSense V2	Large	81.29	95.99
YouTu Lab	Large	83.29	91.34
Vocord - deepVo V3	Large	**91.76**	94.96
CosFace(Single-patch)	Large	82.72	96.65
CosFace(3-patch ensemble)	Large	84.26	**97.96**

Table 4. Face identification and verification evaluation on MF1. "Rank 1" refers to rank-1 face identification accuracy and "Veri." refers to face verification TAR under 10^{-6} FAR.

Wang H, Wang Y, Zhou Z, et al. CosFace: Large Margin Cosine Loss for Deep Face Recognition[J]. 2018.

Method	Protocol	MF2 Rank1	MF2 Veri.
3DiVi	Large	57.04	66.45
Team 2009	Large	58.93	71.12
NEC	Large	62.12	66.84
GRCCV	Large	75.77	74.84
SphereFace	Large	71.17	84.22
CosFace (Single-patch)	Large	74.11	86.77
CosFace(3-patch ensemble)	Large	**77.06**	**90.30**

Table 5. Face identification and verification evaluation on MF2. "Rank 1" refers to rank-1 face identification accuracy and "Veri." refers to face verification TAR under 10^{-6} FAR .

Wang H, Wang Y, Zhou Z, et al. CosFace: Large Margin Cosine Loss for Deep Face Recognition[J]. 2018.