CosFace: Large Margin Cosine Loss for Deep Face Recognition

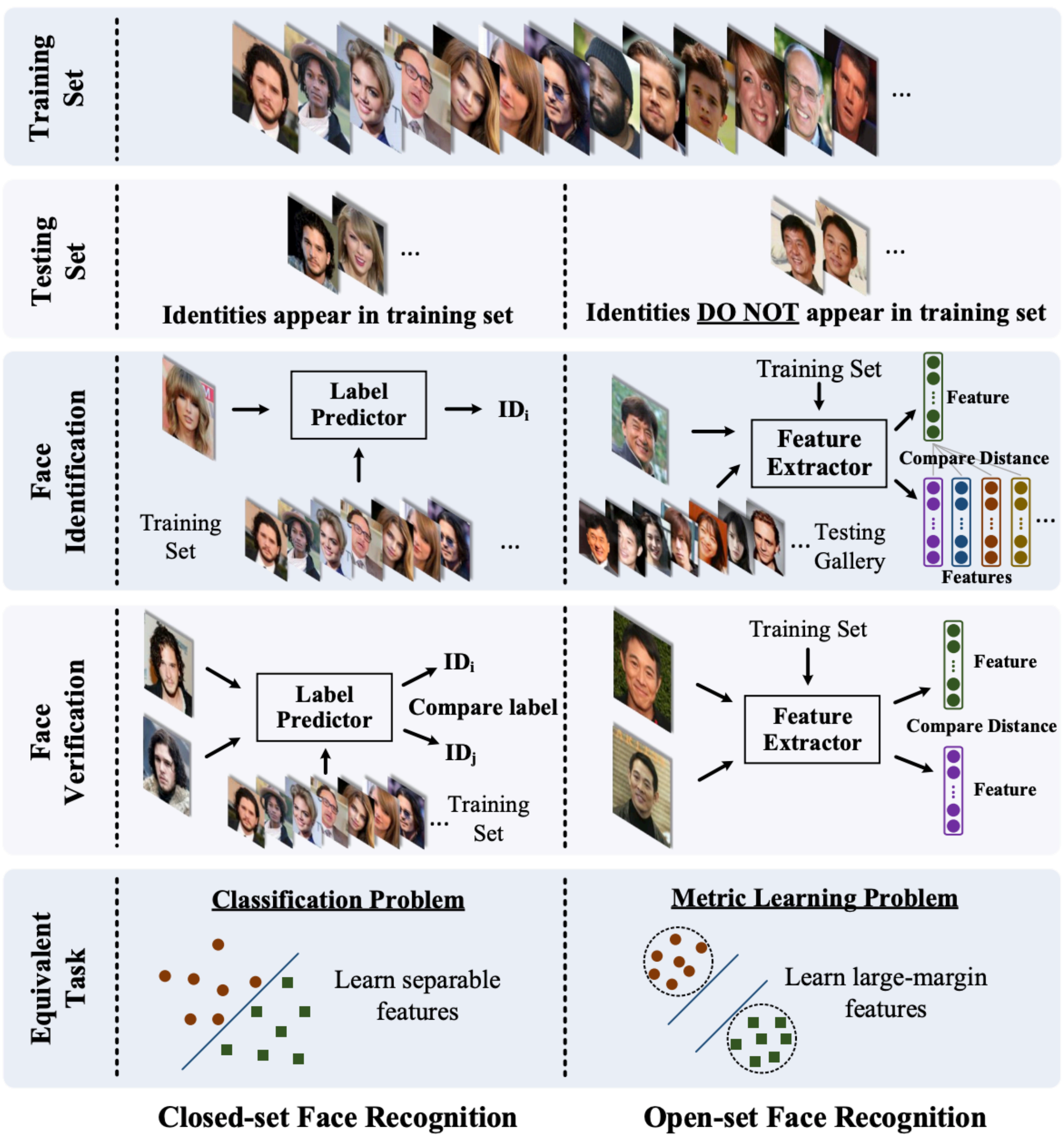
Hao Wang, Yitong Wang, Zheng Zhou, Xing Ji,
Dihong Gong, Jingchao Zhou, Zhifeng Li, and Wei Liu

Presented by Mikhail Liz

What is the task of face recognition?

The central task of face recognition, including face verification and identification, involves face feature discrimination.

Two ways to solve the problem



Open-set:

- Intrinsically large intra-class variation
- High inter-class similarity

Original Softmax Loss

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

where p_i denotes the posterior probability of x_i being correctly classified.
N is the number of training samples and C is the number of classes

$$f_j = W_j^T x = \|W_j\|\|x\| \cos\theta_j$$

where θ_i is the angle between W_i and x_i

Normalized version of Softmax Loss (NSL)

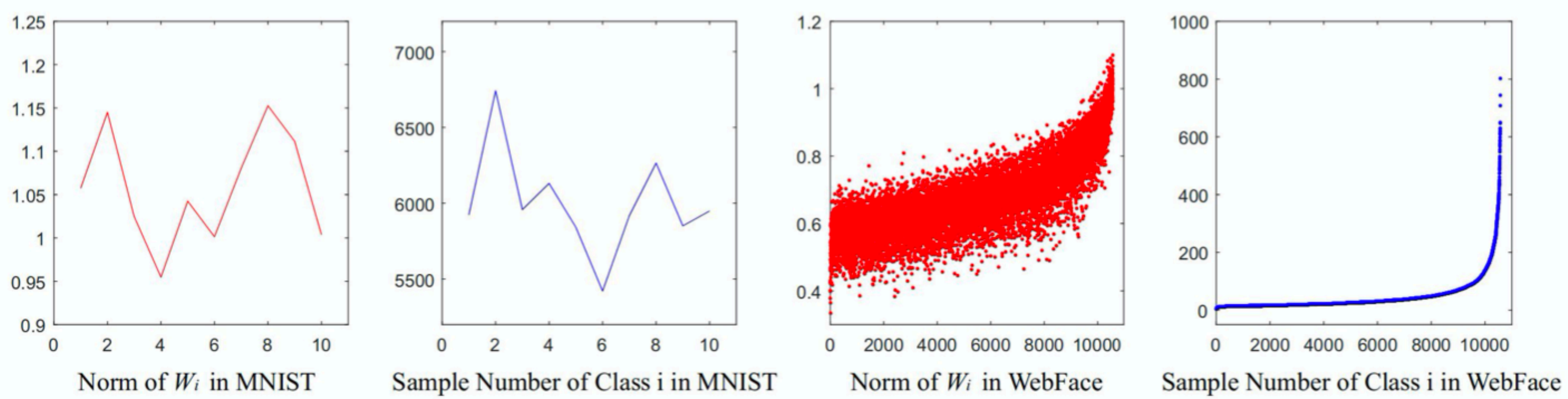
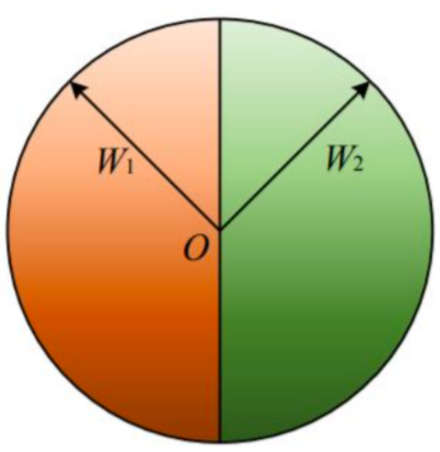


Figure 9: Norm of $\boldsymbol{W}_i$ and sample number of class i in MNIST dataset and CASIA-WebFace dataset.

$$\|\boldsymbol{W}_i\| = 1, b_i = 0$$

$$L_{\text{modified}} = \frac{1}{N} \sum_i -\log\left(\frac{e^{\|\boldsymbol{x}_i\| \cos(\theta_{y_i,i})}}{\sum_j e^{\|\boldsymbol{x}_i\| \cos(\theta_{j,i})}}\right)$$



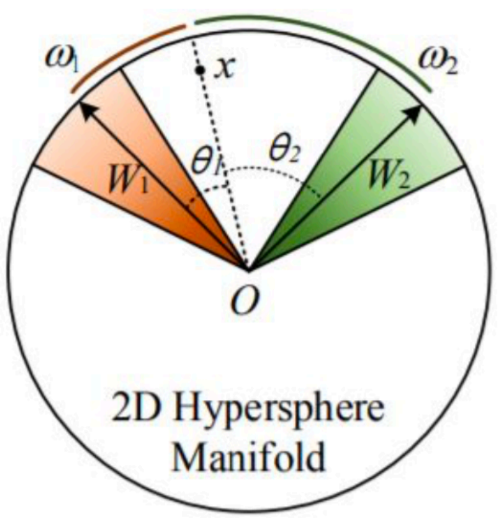
Angular Softmax (A-Softmax)

$$L_{\text{ang}} = \frac{1}{N}\sum_i -\log\left(\frac{e^{\|\boldsymbol{x}_i\|\cos(m\theta_{y_i,i})}}{e^{\|\boldsymbol{x}_i\|\cos(m\theta_{y_i,i})} + \sum_{j\neq y_i} e^{\|\boldsymbol{x}_i\|\cos(\theta_{j,i})}}\right)$$

$\theta_1 \in [0, \frac{\pi}{m}], m \geq 2$

$C_1 : \cos(m\theta_1) \geq \cos(\theta_2),$

$C_2 : \cos(m\theta_2) \geq \cos(\theta_1).$



$$L_{\text{ang}} = \frac{1}{N}\sum_i -\log\left(\frac{e^{\|\boldsymbol{x}_i\|\psi(\theta_{y_i,i})}}{e^{\|\boldsymbol{x}_i\|\psi(\theta_{y_i,i})} + \sum_{j\neq y_i} e^{\|\boldsymbol{x}_i\|\cos(\theta_{j,i})}}\right)$$

where $\psi(\theta_{y_i,i}) = (-1)^k \cos(m\theta_{y_i,i}) - 2k$ - monotonically decreasing, and

$\theta_{y_i,i} \in [\frac{k\pi}{m}, \frac{(k+1)\pi}{m}]$ and $k \in [0, m-1]$

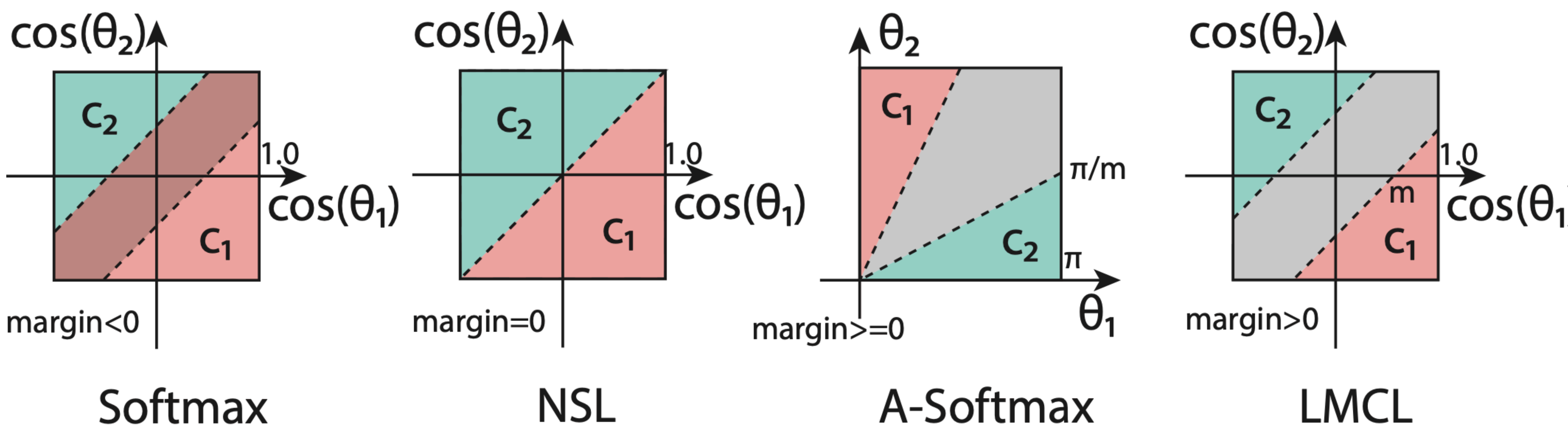
Large Margin Cosine Loss (LMCL)

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

$$\|x\| = s$$

$$cos(\theta_j, i) = W_j{}^T x_i$$

The comparison of decision margins



Results

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**

All the methods in this table are using the same training data and the same 64-layer CNN architecture.

Thank you for your attention!