Cosface

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

Tencent AI Lab, University of Columbia

Published on: 3rd April 2018

Outline

Motive
- Why Cosface?
- Previous Works

Results
- Main Results
- Basic Ideas for Proofs/Implementation

What is Cosface?

- The main challenge of face recognition involves face feature discrimination.
- The traditional softmax loss in deep CNNs lacks the power of discrimination. Hence, several loss functions such as center loss, large margin softmax loss, and angular softmax loss have been proposed.
- All these improved losses share the same idea: maximizing inter-class variance and minimizing intra-class variance.
- In cosface, the discriminative face features are learned with a large margin between different classes.

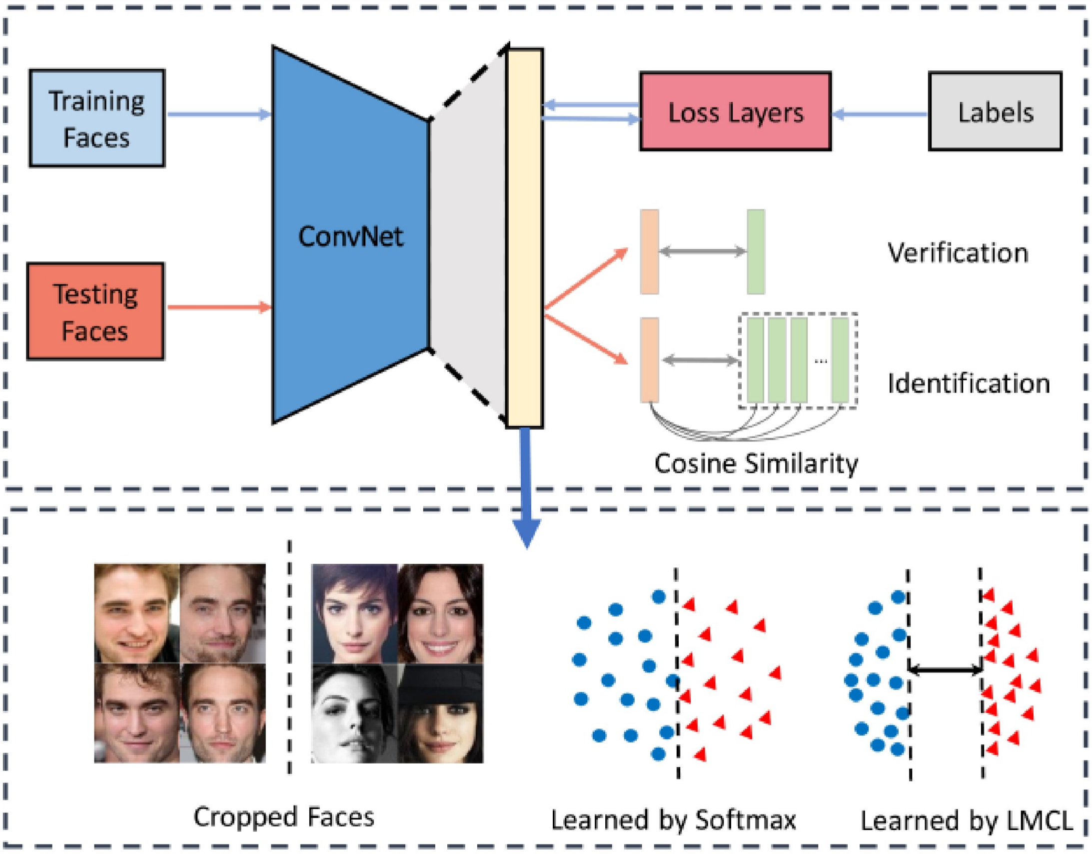

Figure: Illustration of Cosface

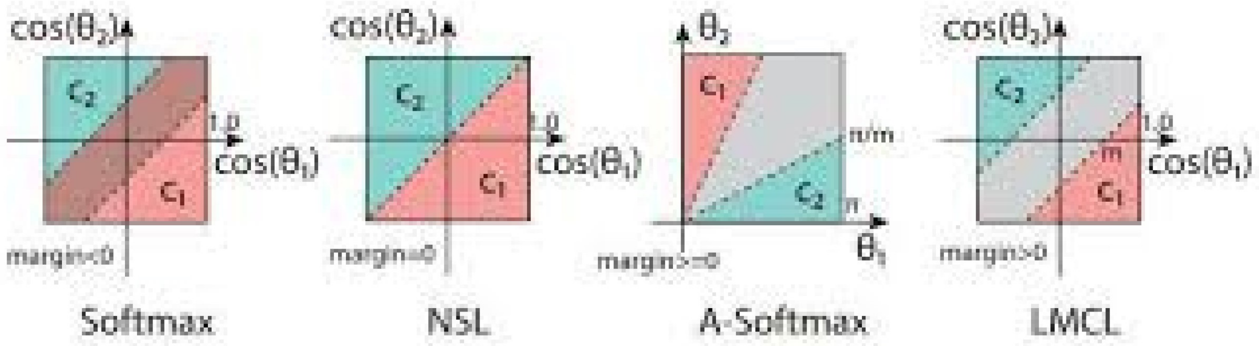

Figure 2. The comparison of decision margin for different loss functions for two classes. The dashed line represents decision boundary, and the gray areas are decision margins.

Figure: Loss Comparison

What is the Use?

- Loss function is an important part in artificial neural networks, which is used to measure the inconsistency between predicted value and actual label.
- For each prediction that we make, our loss function will simply measure the absolute difference between our prediction and the actual value.
- Most machine learning algorithms use some sort of loss function in the process of optimization, or finding the best parameters (weights) for your data.

What is the Use?

- In Cosface, unlike softmax, the classes are mapped on to a hypersphere, where our sole motive is to reduce the intra class distance and to increase the inter class distance.
- Here, we also represent angles apart from only distances. In our case, the distances are geodesic in nature.
- It is very useful in dealing with very large and tricky datasets, especially when they are intended to confuse the neural network.

Formulas

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

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

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

subject to

$$W = \frac{W^*}{\|W^*\|},$$
$$x = \frac{x^*}{\|x^*\|},$$
$$cos(\theta_j, i) = {W_j}^T x_i,$$

Hyperspace Example

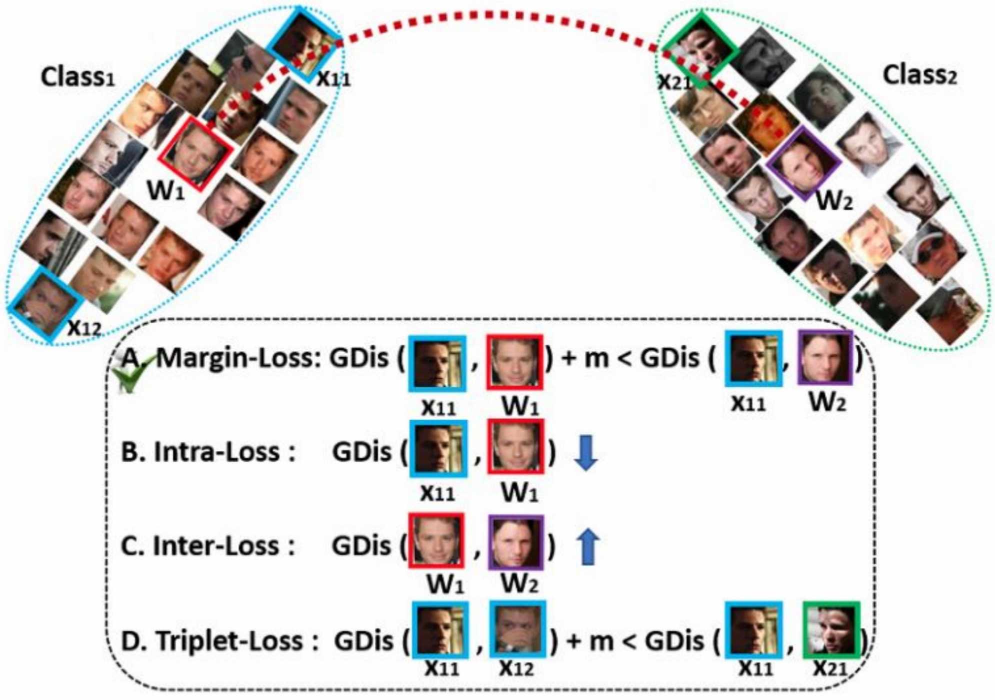

Figure: Geodesic distance and class seperation

Previous Works

CASIA and LFW, respectively.

Method	#Image	LFW	YTF
DeepID [30]	0.2M	99.47	93.20
Deep Face [31]	4.4M	97.35	91.4
VGG Face [22]	2.6M	98.95	97.30
FaceNet [27]	200M	99.63	95.10
Baidu [15]	1.3M	99.13	-
Center Loss [36]	0.7M	99.28	94.9
Range Loss [43]	5M	99.52	93.70
Marginal Loss [8]	3.8M	99.48	95.98
SphereFace [17]	0.5M	99.42	95.0
SphereFace+ [16]	0.5M	99.47	-
CosFace [35]	5M	99.73	97.6
MS1MV2, R100, ArcFace	5.8M	**99.83**	**98.02**

Table 4. Verification performance (%) of different methods on LFW and YTF.

Figure: Results Comparison

Main Results

- Cosface is second after Arcface as per the results published. However, we only have one publication of Arcface, Sphereface and Cosface existing.

Links/References

http://openaccess.thecvf.com/content_cvpr_2018/CameraReady/1797.pdf