KNOT MAGNIFY LOSS FOR FACE RECOGNITION

Qiang Rao, Bing Yu, Yun Yang, Bailan Feng

Noah's Ark Laboratory, Huawei Technologies Co., Ltd

HUAWEI

Introduction

Deep learning based face recognition

- Deep Convolutional Neural Network (DCNN) significantly improved the performance of face recognition
- Face Recognition including verification and identification is open-set recognition problem
- Discrimination and Metric learning are widely used for representation&feature learning for open-set recognition
- Softmax with cross-entropy is now a main loss function which is powerful and robust for discrimination learning

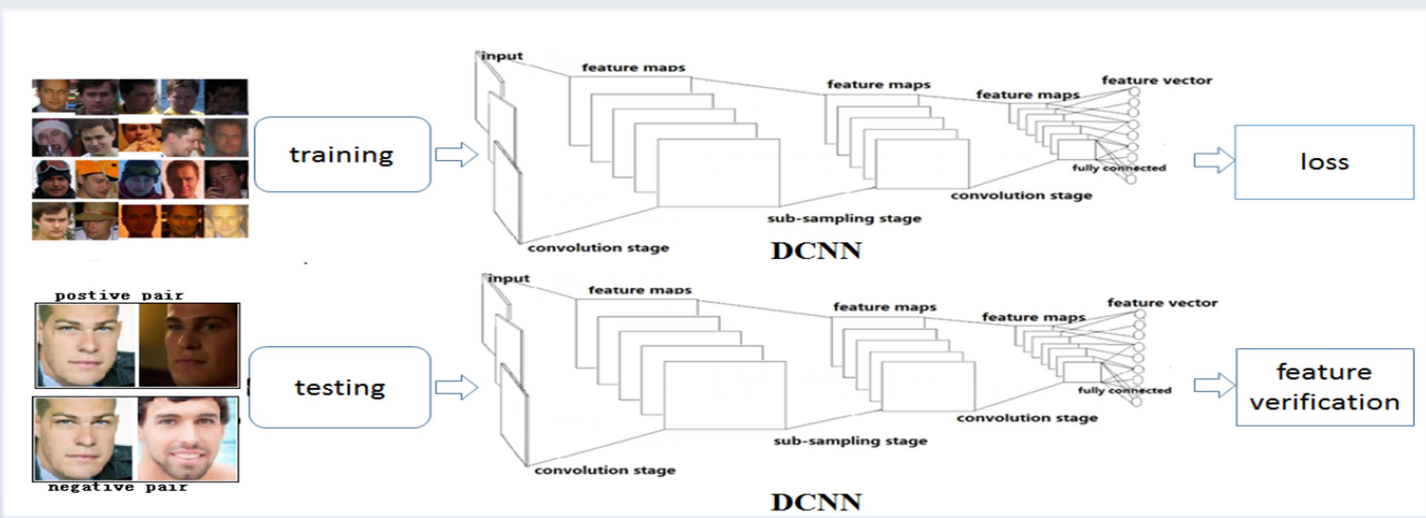

Fig1:Pipeline of DCNN for face recognition

Challenges of face recognition

- Most training dataset present imbalanced distribution
- Low quality face images(profile, occlusion, blurriness) is difficult to recognize
- Collecting training datasets is tedious and time-consuming

Motivation

Softmax with Cross-entropy

$$L_S = -\sum_{k}^{N}\sum_{i}^{c} I(y_k = i) log(\frac{e^{w_i x_k}}{\sum_{j}^{c} e^{w_j x_k}})$$

Statistical probability of each sample belonging to label

$$p_k^i = \frac{e^{w_i x_k}}{\sum_{j}^{c} e^{w_j x_k}}, y_k = i$$

- Low quality faces are prone to lower probability to its label for sparsity distribution
- Lower probabilities to some degree response the lower quality of the face images
- DCNN is able to learn more from low quality faces by relatively magnifying their loss

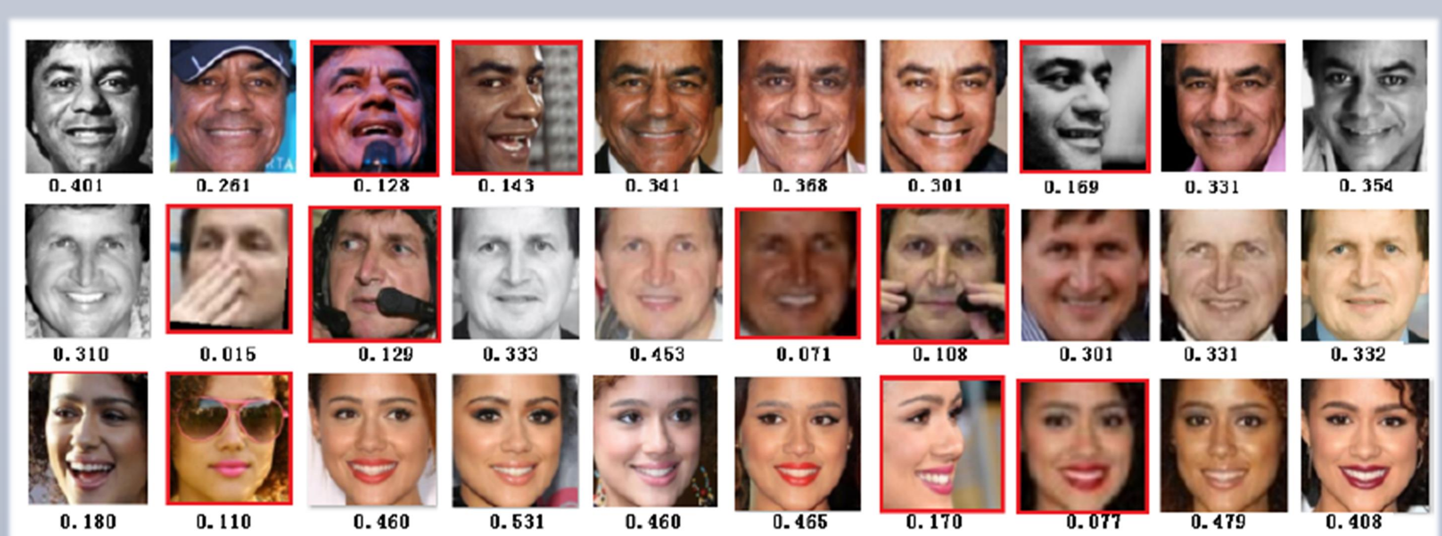

Fig2: p_k^i of some examples in the training images set MS-1M[1]

Proposed Method

Relatively magnify the loss of low quality images

The proposed Knot Magnify(KM) loss is multiplying a factor $\frac{1}{(\gamma p+1)^2}$ with softmax loss. As we can see, setting γ to a suitable value, the KM loss will suppress the loss of easy samples and magnify that of hard samples relatively; When $\gamma = 0$, the KM loss degenerates into softmax loss. Fig.3 shows that modifying the loss weight of each sample will have different impacts on the easy samples that have larger softmax output p_k^i and hard samples that have smaller p_k^j

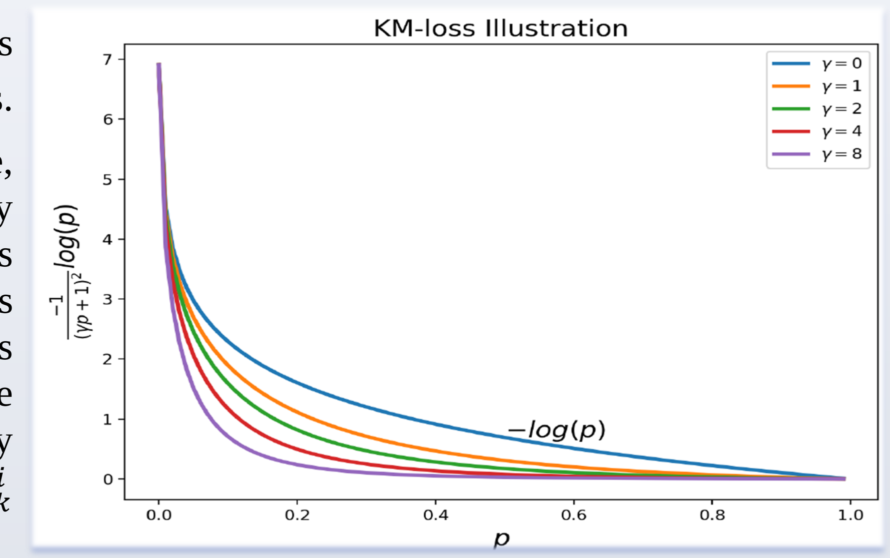

Fig3:Illustration of Knot Magnify Loss

Knot Magnify Loss based on Softmax

- The proposed KM loss definition

$$L_k = -\sum_{k}^{n} \frac{1}{(\gamma p_k^i + 1)^2} log(p_k^i)$$

- The Component of KM Loss

$$K_\gamma(p_k^i) = \frac{1}{(\gamma p_k^i + 1)^2} log(p_k^i)$$

The KM loss adds different weight to each sample's loss during training, such that hard samples would have more effect on optimizing DCNNs. In other words, we assign larger weights to the rare hard samples and smaller weights to the mostly easy samples in terms of recognition loss

Theoretic Analysis

Normalized Loss

We quantitatively analyzed the loss effect by considering its cumulative loss corresponding to softmax output p_k^i which ranges in [0,1].

- Definition of softmax loss and KM loss

$$L_S^N = \frac{log(p_k^i)}{\int_0^1 log(p_k^i)} = log(p_k^i) \qquad L_{K_\gamma}^N = \frac{K_\gamma(p_k^i)}{\int_0^1 K_\gamma(p_k^i)} = \frac{-\gamma \log(p_k^i)}{\log(\gamma+1)(\gamma p_k^i + 1)^2}$$

Analysis of factor γ

- Ratio of normalized KM loss over normalized softmax loss

$$R(\gamma) = \frac{L_{K_\gamma}^N}{L_S^N} = \frac{-\gamma}{\log(\gamma+1)(\gamma p_k^i + 1)^2}$$

We expect $R_\gamma(p) \geq 1$ when $p \geq p_c$ and $R_\gamma(p) \leq$ when $p \leq p_c$. Considering $\gamma > 0, p > 0$, letting $R_\gamma(p) = 1$, we get the critical point:

$$p_c(\gamma) = \sqrt{\frac{1}{\gamma log(1+\gamma)}} - \frac{1}{\gamma}$$

γ	0.1	1	2	4	6	8
$p_c(\gamma)$	0.244	0.202	0.175	0.145	0.126	0.114

Results

Face verification and identification performance

- The performance of our proposed method is compared with a serial of methods in both verification and identification
- KM loss outperforms softmax a large margin in face verification and identification on LFW data
- KM loss combined with center loss outperforms a serial of methods both on LFW and CFP data

Tab1:Verification accuracy on LFW

Methods	Data	#loss	#Net	mAcc(%)
DeepFace[2]	4M	2	3	97.35
VGGFace[3]	2.6M	1	2	98.95
Facenet[4]	200M	1	1	99.63
DeepID2[5]	300k	2	25	99.47
Center Loss[6]	700k	2	1	99.28
Sphereface[7]	500k	1	1	99.42
softmax	3.7M	1	1	99.10
KM loss	3.7M	1	1	99.31
Center+KM loss	3.7M	2	1	99.53

Tab3:Verification accuracy on CFP

Data	Loss	mAcc(%)
CFP-FF	softmax	99.19
	KM loss	99.43
	KM+center loss	99.46
CFP-FP	softmax	91.27
	KM loss	91.71
	KM+center loss	93.39

Tab2:Rank-1 identification accuracy on LFW dataset

Methods	#Net	Protocol	Rank1(%)
DeepFace[2]	7	unrestricted	97.35
Web-Scale[8]	4	unrestricted	98.95
DeepID2[5]	25	unrestricted	99.63
VGGFace[3]	1	unsupervised	99.47
Center Loss[6]	1	unsupervised	99.28
softmax	1	unsupervised	99.10
KM loss	1	unsupervised	99.31
Center+KM loss	1	unsupervised	99.53

Conclusion

- To address the problem of lacking learning ability from hard samples, we modified the widely adopted softmax loss by proposing KM loss which assigns weights to training samples according to its softmax output in order to suppress the influence of easy samples and magnify the effect of rare hard samples.
- Our approach is simple and easy to implement and can be easily combined with other auxiliary losses which benefit for getting a more robust model
- We demonstrated our method's effectiveness on the well-known LFW dataset and the challenging CFP dataset.

Reference

[1] Guo Y, Zhang L, Hu Y, et al. Ms-celeb-1m: A dataset and benchmark for large-scale face recognition[C]//ECCV. 2016: 87-102.

[2] Taigman Y, Yang M, Ranzato M A, et al. Deepface: Closing the gap to human-level performance in face verification[C]//CVPR.. 2014: 1701-1708.

[3] Parkhi O M, Vedaldi A, Zisserman A. Deep face recognition[C]//BMVC. 2015, 1(3): 6.

[4] Schroff F, Kalenichenko D, Philbin J. Facenet: A unified embedding for face recognition and clustering[C] //CVPR.. 2015: 815-823.

[5] Sun Y, Wang X, Tang X. Deeply learned face representations are sparse, selective, and robust[C]// CVPR. 2015: 2892-2900.

[6] Wen Y, Zhang K, Li Z, et al. A discriminative feature learning approach for deep face recognition[C]//ECCV. 2016: 499-515.

[7] Liu W, Wen Y, Yu Z, et al. Sphereface: Deep hypersphere embedding for face recognition[C]//CVPR. 2017, 1: 1.

[8] Taigman Y, Yang M, Ranzato M A, et al. Web-scale training for face identification[C]//CVPR. 2015: 2746-2754.