Orientation Truncated Center Learning for Deep Face Recognition

Monica M. Y. Zhang, Yifang Xu✉ and Huaming Wu

Recently, Center loss that aiming to assist Softmax loss with the objectives of both inter-class dispension and intra-class compactness simultaneously, has achieved remarkable performance on convolutional neural network based face recognition. However, its advantages highly rely on the center feature assumption, which influences the capacity of the final obtained face features. Inspired by the center loss approach, a novel Orientation Truncated Center Learning (OTCL) is proposed, which takes advantage of an orientation truncated center function to make the center feature learning have more suitable orientation for deep face recognition. Three metrics are proposed to evaluate how discriminative are the distributions of the learned features for MNIST visualization. Experimental results on several challenging benchmarks, including FGLFW, LFW, YTF, and BLUFR, show that the proposed approach can easily generate more favorable results than several state-of-the-art competitors.

Introduction: Convolutional Neural Network (CNN) based face recognition has achieved significant performance. However, how to design better supervision signals for more discriminative face features is one of the most concerned issues. Commonly used loss functions include Softmax loss, Contrastive loss [5] and Triplet loss [6]. Softmax loss is effective for multi-class classification, but the learned face features are not discriminative sometimes. Contrastive loss and Triplet loss make it more discriminative by using information of feature pairs and triplets. However, the training procedure is not straightforward and the computation complexity will increase by selecting meaningful image pairs and triplets. Recently, center loss approach [9], which is a simple and trainable method, has achieved great progress. However, its advantages highly rely on the center feature assumption. Once the center feature is not learned appropriately, the final face features may not represent the raw face images suitably. Particularly, the situation may be more serious when there exist a certain number of outliers. To this end, we propose a simple and efficient approach for more discriminative face features, called orientation truncated center learning (OTCL). Rather than defining the center feature by averaging the features of the focused class in each iteration, OTCL learns the center feature by using an orientation truncated function. The idea is to update the center feature according to its nearest feature members, instead of using the full features to avoid the disturbance of some outliers. Thus, the center feature can represent the features of the same class more efficiently to learn more suitable CNN models. Experimental results show the superiority of the proposed approach over two baselines and several state-of-the-art methods.

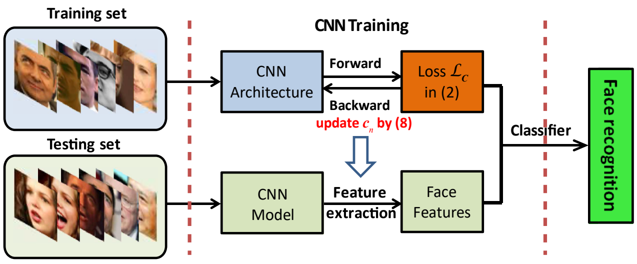

Fig. 1 *Framework of orientation truncated center learning (OTCL)*

Proposed Approach: Center loss approach for CNN model learning is based on an optimization objective, expressed as

$$\boldsymbol{\theta}^* = \min_{\boldsymbol{\theta}} \ \mathcal{L}_C(X, L, \boldsymbol{\theta}), \tag{1}$$

where $\mathcal{L}_C(X, L, \boldsymbol{\theta})$ is the joint supervision of Softmax loss $\mathcal{L}_S$ and Center loss $\mathcal{L}_c$, namely,

$$\mathcal{L}_C(X, L, \boldsymbol{\theta}) = \mathcal{L}_S(X, L, \boldsymbol{\theta}) + \lambda \mathcal{L}_c(X, L, \boldsymbol{\theta}), \tag{2}$$

and $X = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n\}$ is the training data set, $L = \{l_1, l_2, \cdots, l_n\}$ is the corresponding label set, and $\boldsymbol{\theta}$ is the parameter set, λ is a hyper-parameter to balance the two losses. Here $\mathcal{L}_c$ is Center loss which is based on the distance of the feature $\boldsymbol{x}_m$ to its corresponding center feature $\boldsymbol{c}_{l_m}$,

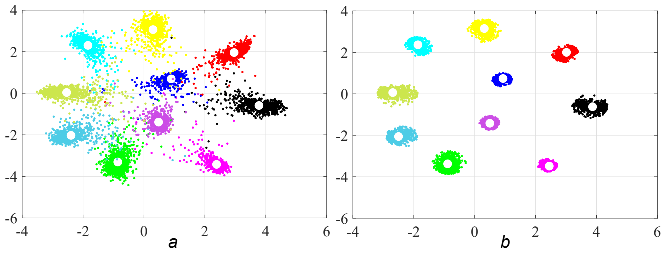

Fig. 2 *Center features (white points) for different distributions of MNIST*
a distribution of MNIST testing database by $\mathcal{L}_C$
b distribution of MNIST testing database by $\mathcal{L}_{OTC}$

formalized as

$$\mathcal{L}_c = \frac{1}{2M} \sum_{m=1}^{M} \|\boldsymbol{x}_m - \boldsymbol{c}_{l_m}\|^2, \tag{3}$$

where $\boldsymbol{c}_{l_m}$ is computed as the average of the features in the l_m-th class, and M is the mini-batch size. However, when there exist many outliers for a focused class, the corresponding center feature may not properly represent the class. As shown in Fig. 2*a*, many features are far away from their corresponding centers, which seems to have little connection with the center feature updating.

Intuitively, we can update the center feature according to those nearest features around the center feature, instead of using the full features. Suppose that there exist some nearest features $\{\boldsymbol{x}_{i_1}, \cdots, \boldsymbol{x}_{i_{N_i}}\}$ around the center feature $\boldsymbol{c}_i$, such that

$$\sum_{j=1}^{N_i} \|\boldsymbol{x}_{i_j} - \boldsymbol{c}_i\|^2 \approx f(i) = R \sum_{m=1}^{M} \mathbb{I}(l_m = i) \|\boldsymbol{x}_m - \boldsymbol{c}_{l_m}\|^2, \tag{4}$$

where N_i is the number of features in class i, $\mathbb{I}$ is the indicator function, and $R \in (0, 1)$. We want to find a suitable R to represent $\boldsymbol{c}_i$, and thus to avoid the disturbance of the outliers for center feature updating, shown in Fig. 2*b*.

For CNN training with N classes, considering all features in a mini-batch, then

$$\sum_{i=1}^{N} f(i) = R \sum_{i=1}^{N} \sum_{m=1}^{M} \mathbb{I}(l_m = i) \|\boldsymbol{x}_m - \boldsymbol{c}_{l_m}\|^2 = R \sum_{m=1}^{M} d_m, \tag{5}$$

where $d_m = \|\boldsymbol{v}_m\|^2$ and $\boldsymbol{v}_m = \boldsymbol{x}_m - \boldsymbol{c}_{l_m}$, we aim to find a smallest $\hat{M}$ ($\hat{M} \le M$) such that

$$\sum_{m=1}^{\hat{M}} d_{i_m} \ge R \sum_{m=1}^{M} d_m, \tag{6}$$

where $d_{i_1} \le d_{i_2} \le \cdots \le d_{i_M}$. Then, we propose the Orientation Truncated Center (OTC) function

$$\mathcal{L}_{OTC} = \frac{1}{2M} \sum_{m=1}^{\hat{M}} d_{i_m} = \frac{1}{2M} \sum_{m=1}^{\hat{M}} \|\boldsymbol{x}_{i_m} - \boldsymbol{c}_{l_{i_m}}\|^2, \tag{7}$$

a truncated version of Center loss, to assist the center feature updating to have more suitable orientation for CNN feature extraction. Further, we update the center feature by

$$\Delta \boldsymbol{c}_i = -\gamma \frac{\partial \mathcal{L}_{OTC}}{\partial \boldsymbol{c}_i} = \frac{\gamma}{M} \sum_{m=1}^{\hat{M}} \mathbb{I}(l_{i_m} = i) \boldsymbol{v}_{i_m}, \tag{8}$$

where γ is the center feature learning rate.

In this way, we propose orientation truncated center learning (OTCL) by changing the backward computation of center loss approach by (8), without modifying the forward computation, which can be easily optimized by the standard stochastic gradient descent.

MNIST visualization: We use LeNet++ [9] and MNIST database for feature visualization. Three metrics are proposed to characterize the discrimination of the features: the average cosine distance between each sample and its corresponding center feature (CD1), the average cosine distance between the center features (CD2), the average cosine distance between each sample and its inter-class center feature (CD3), where

$$\mathrm{CD1} = \sum_{i=1}^{N} \sum_{j=1}^{N_i} \frac{1}{N_i N} \frac{\boldsymbol{c}_i^T \boldsymbol{x}_{ij}}{\|\boldsymbol{c}_i\| \, \|\boldsymbol{x}_{ij}\|}, \tag{9}$$

$$\mathrm{CD2} = \frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j=1}^{N} \mathbb{I}(i \neq j) \frac{\boldsymbol{c}_i^T \boldsymbol{c}_j}{\|\boldsymbol{c}_i\| \, \|\boldsymbol{c}_j\|}, \tag{10}$$

$$\mathrm{CD3} = \frac{1}{\sum_i^N N_i(N-1)} \sum_{i=1}^{N} \sum_{j=1}^{N_i} \sum_{n=1}^{N} \mathbb{I}(n \neq i) \frac{\boldsymbol{c}_n^T \boldsymbol{x}_{ij}}{\|\boldsymbol{c}_n\| \, \|\boldsymbol{x}_{ij}\|}, \tag{11}$$

$\boldsymbol{c}_i$ is the center feature for class i, $\boldsymbol{x}_{ij}$ is the feature in class i, N_i is the number of features for class i, and N is the number of classes. By the above definitions, feature distribution with larger CD1, smaller CD2 and smaller CD3 is treated as more discriminative. As shown in Fig. 3, our proposed OTCL performs better than the center loss approach.

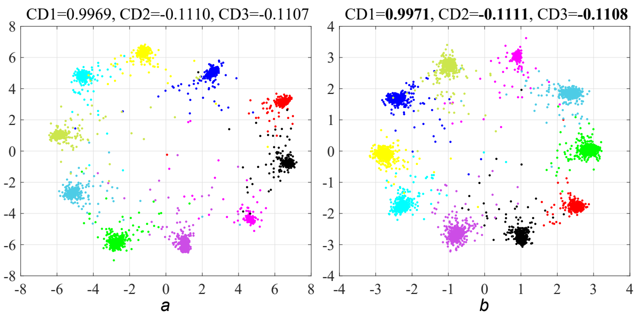

Fig. 3 *Final feature distributions corresponding to $\mathcal{L}_C$ and $\mathcal{L}_{OTC}$*

a the diameter of a class cluster is about 2

b the diameter of a class cluster is about 1

Experimental results: The proposed approach is used for face feature extraction without fine-tuning operations on CASIA-WebFace database [11] and ResNet-27 [9]. The initial learning rate is 0.1 and is divided by 10 at 30K, 50K iterations, until reaching the maximum iteration 60K. We set $\lambda = 0.003$ and $\gamma = 0.5$ according to [9], and range R in $[0.1, 0.2, \cdots, 0.9]$.

Table 1: Comparing performance on FGLFW

Method	#Train	FGLFW (%)
Noisy Softmax [1]	0.5M	94.50
Human [2]	n/a	92.00
DCMN [2]	0.5M	91.00
VGG [4, 2]	2.6M	85.78
DeepFace [7, 2]	0.5M	78.78
DeepID2 [5, 2]	0.2M	78.25
Softmax	0.44M	90.87
Softmax + Center	0.44M	94.28
OTCL-0.5	0.44M	95.38
OTCL-0.7	0.44M	**95.45**

Table 2: Comparing performance on LFW and YTF

Method	#Train	LFW (%)	YTF (%)
SphereFace [3]	0.49M	99.42	95.0
SphereFace	0.44M	99.12	92.98
NormFace [8]	0.49M	99.19	94.72
NormFace	0.44M	98.63	93.26
Softmax + Center [9]	0.7M	99.28	94.9
Softmax + Center	0.44M	99.03	93.3
OTCL-0.5	0.44M	**99.17**	93.94
OTCL-0.7	0.44M	**99.17**	**94.18**

Table 3: Comparing performance on BLUFR protocol

Method	Verification (%)		Identification (%)	
	FAR=0.1%	FAR=1%	FAR=1%	FAR=10%
NormFace [8]	95.83	-	77.18	-
Softmax + Center [8, 9]	93.35	-	67.86	-
LightenedCNN [10]	89.12	-	61.79	-
WebFaceCNN [11]	80.26	-	28.9	-
Softmax	82.22	93.5	56.81	73.3
Softmax + Center	93.64	98.12	70.73	86.91
OTCL-0.5	94.15	98.15	75.29	88.73
OTCL-0.7	**94.88**	**98.42**	**77.28**	**89.12**

Fig. 4 *Example images for MINST, LFW, and YTF*

The performances of our best models OTCL-0.5 ($R = 0.5$) and OTCL-0.7 ($R = 0.7$) are reported on FGLFW in Table 1, LFW and YTF in Table 2, and BLUFR protocol in Table 3, respectively. Experimental results show that the proposed approach outperforms two baselines: Softmax, and Softmax + Center. Specifically, our proposed approach achieves the 1st place on FGLFW, and performs better than most of the compared methods on LFW, YTF, and BLUFR protocol. Note that it also surpasses NormFace and SphereFace with the same training data in Table 2. These all show the superiority of the proposed approach to characterize the center features for more discriminative face features.

Conclusions: In this letter, we propose a simple and more efficient algorithm for CNN-based face features learning, referred to as orientation truncated center learning. By adopting an orientation truncated center function to restrict the clustering degree in a mini-batch for the center feature definition, we make the center feature represent the features of the same class more efficiently to learn CNN models. Feature visualization with three metrics on real-world dataset MNIST shows that the proposed approach can make the features more discriminative. Various evaluation implementations on face recognition tasks show that the proposed approach is effective and can easily generate more favorable results than the baseline center loss approach and related state-of-the-art methods.

Monica M. Y. Zhang and Yifang Xu (*Center for Combinatorics, LPMC, Nankai University, No. 94, Weijin Road, Tianjin 300071, People's Republic of China*)
E-mail: xyf@mail.nankai.edu.cn

Huaming Wu (*Center for Applied Mathematics, Tianjin University, No. 92, Weijin Road, Tianjin 300072, People's Republic of China*)

References

1 Chen, B. H., Deng, W. H. and Du, J. P.: 'Noisy softmax: Improving the generalization ability of dcnn via postponing the early softmax saturation', *Proc. of IEEE CVPR*, 2017

2 Deng, W. H., Hu, J. N., Zhang, N. H., *et al.*: 'Fine-grained face verification: FGLFW database, baselines, and human-dcmn partnership', *Pattern Recognition*, 2017, **66**, pp. 63–73

3 Liu, W. Y., Wen, Y. D., Yu, Z. D., *et al.*: 'Sphereface: Deep hypersphere embedding for face recognition', *Proc. of IEEE CVPR*, 2017, pp. 212–220

4 Parkhi, O., Vedaldi, A., Zisserman, A., *et al.*: 'Deep face recognition.' Proc. of BMVC, **1**, 2015, pp. 6

5 Sun, Y., Chen, Y. H., Wang, X. G., *et al.*: 'Deep learning face representation by joint identification-verification', *Proc. of NIPS*, 2014, pp. 1988–1996

6 Schroff, F., Kalenichenko, D. and Philbin, J.: 'Facenet: A unified embedding for face recognition and clustering', *Proc. of IEEE CVPR*, 2015, pp. 815–823

7 Taigman, Y., Yang, M., Ranzato, M., *et al.*: 'Deepface: Closing the gap to human-level performance in face verification', *Proc. of IEEE CVPR*, 2014, pp. 1701–1708

8 Wang, F., Xiang, X., Cheng, J., *et al.*: 'Normface: L2 hypersphere embedding for face verification', Proc. of ACM Multimedia, 2017

9 Wen, Y. D., Zhang, K. P., Li, Z. F., *et al.*: 'A discriminative feature learning approach for deep face recognition', *Proc. of ECCV*, 2016, pp. 499–515

10 Wu, X., He, R., and Sun, Z. N.: 'A lightened cnn for deep face representation', *Proc. of IEEE CVPR*, **4**, 2015

11 Yi, D., Lei, Z., Liao, S. C., *et al.*: 'Learning face representation from scratch', *arXiv preprint arXiv:1411.7923*, 2014