A Discriminative Feature Learning Approach for Deep Face Recognition

Yandong Wen, Kaipeng Zhang, Zhifeng Li*, Yu Qiao

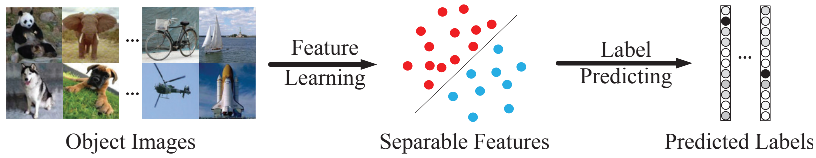
Shenzhen Institutes of Advanced Technology, CAS, China

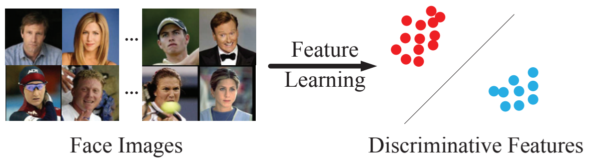
The Chinese University of Hong Kong, Hong Kong, China

Introduction

- For generic object, scene or action recognition. The deeply learned features need to be separable. Because the classes of the possible testing samples are within the training set, the predicted labels dominate the performance.
- For face recognition task, the deeply learned features need to be not only separable but also discriminative. Since it is impractical to pre-collect all the possible testing identities for training, the label prediction in CNNs is not always applicable.
- The deeply learned features are required to be generalized enough for identifying new unseen classes without label prediction.

Overview





Discriminative Feature Learning

- SOFTMAX LOSS: encouraging the separability of features.
- CENTER LOSS: simultaneously learning a center for deep features of each class and penalizing the distances between the deep features and their corresponding class centers.
- JOINT SUPERVISION: minimizing the intra-class variations while keeping the features of different classes separable.

$$\mathcal{L} = \mathcal{L}_S + \lambda \mathcal{L}_C$$

$$= \underbrace{-\sum_{i=1}^{m} \log \frac{e^{W_{y_i}^T \boldsymbol{x}_i + b_{y_i}}}{\sum_{j=1}^{n} e^{W_j^T \boldsymbol{x}_i + b_j}}}_{\text{Inter-class Separability}} + \underbrace{\frac{\lambda}{2} \sum_{i=1}^{m} \|\boldsymbol{x}_i - \boldsymbol{c}_{y_i}\|_2^2}_{\text{Intra-class Compactness}}$$

Detailed Discussion on Center Loss

- **Easy-to-Implement.** The gradient and update equation are easy to derive and the resulting CNN model is trainable.

backward computation

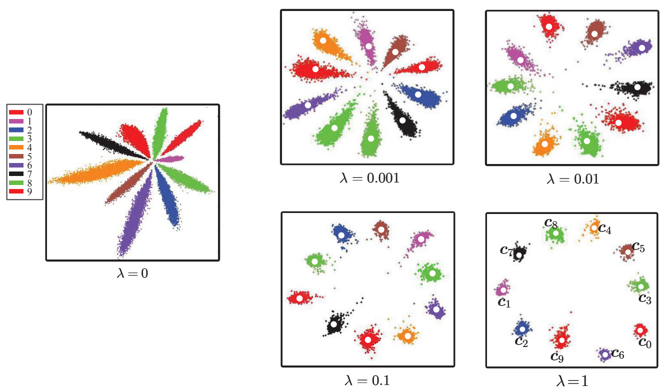
$$\frac{\partial \mathcal{L}_C}{\partial \boldsymbol{x}_i} = \boldsymbol{x}_i - \boldsymbol{c}_{y_i}$$

$$\Delta \boldsymbol{c}_j = \frac{\sum_{i=1}^{m} \delta(y_i = j) \cdot (\boldsymbol{c}_j - \boldsymbol{x}_i)}{1 + \sum_{i=1}^{m} \delta(y_i = j)}$$

- **Easy-to-Train.** Centers are updated based on mini-batch with an adjustable learning rate.
- **Easy-to-Input.** Center loss enjoys the same requirement as the softmax loss and needs no complex sample mining and recombination, which is inevitable in contrastive loss and triple loss.
- **Easy-to-Converge.** Under the joint supervision, our DeepIDNet trained by 0.7M face images is converged at 28k iterations, within 14 hours.

A Visualization Example on MNIST

	stage 1		stage 2		stage 3		stage 4
Layer	conv	pool	conv	pool	conv	pool	FC
LeNets	$(5,20)_{/1,0}$	$2_{/2,0}$	$(5,50)_{/1,0}$	$2_{/2,0}$			500
LeNets++	$(5,32)_{/1,2} \times 2$	$2_{/2,0}$	$(5,64)_{/1,2} \times 2$	$2_{/2,0}$	$(5,128)_{/1,2} \times 2$	$2_{/2,0}$	**2**



- With only softmax loss (λ=0), the deeply learned features are separable, but not discriminative (significant intra-class variations).
- With proper λ, the discriminative power of deep features can be significantly enhanced, which is crucial for face recognition

Experimental Results

- **Labeled Face in the Wild (LFW) & Youtube Face (YTF)**
  — The proposed model is trained on 0.7M face images, termed as model C.

Method	Images	Networks	Acc. on LFW	Acc. on YTF
DeepFace [33]	4M	3	97.35%	91.4%
DeepID-2+ [32]	-	1	98.70%	-
DeepID-2+ [32]	-	25	99.47%	93.2%
FaceNet [27]	200M	1	99.65%	95.1%
Deep FR [25]	2.6M	1	98.95%	97.3%
Baidu [21]	1.3M	1	99.13%	-
Model A	0.7M	1	97.37%	91.1%
Model B	0.7M	1	99.10%	93.8%
Model C (Proposed)	**0.7M**	**1**	**99.28%**	**94.9%**

- **MegaFace**
  — Our model is trained on 490K face images, termed as model C-.

Method	Protocol	Identification Acc. (Set 1)	Verification Acc. (Set 1)
NTechLAB - facenx_large	large	73.300%	85.081%
Google - FaceNet v8		70.496%	86.473%
Beijing Faceall Co. - FaceAll_Norm_1600		64.803%	67.118%
Beijing Faceall Co. - FaceAll_1600		63.977%	63.960%
Barebones_FR - cnn	small	59.363%	59.036%
NTechLAB - facenx_small		58.218%	66.366%
3DiVi Company – tdvm6		33.705%	36.927%
model A-	small	41.863%	41.297%
Model B-		57.175%	69.897%
Model C- (Proposed)		**65.234%**	**76.516%**