Today’ work

Make a theoretical Summary of loss functions by reading papers.

# ***Softmax with*** ***multi-classification***

Early deep face recognition method often use the framework of “CNN + Softmax”, training CNN with a difficult task such as “super multi-classification”, forcing a relatively compact, highly discriminative deep facial feature neural network.

Reference:

Sun Y, Wang X, Tang X. **Deep learning face representation from predicting 10,000 classes** [C]// CVPR, 2014

(20W training set, training 4 layers of CNN + 160-dimensional feature map + 10000 class Softmax, multi-scale multi-patch feature, Joint Bayesian classifier, reaching 97.45% on LFW)

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

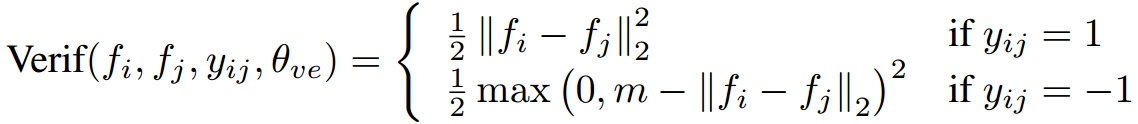
(4.4M training set, training 6 layers CNN + 4096 feature mapping + 4030 class Softmax, combined with technologies such as 3D Aligement, model ensembel, etc., reaching 97.35% on LFW)

Disadvantage:

1. multi-classification is very difficult
2. Softmax guarantees that the category is separable but does not require intra-class compact and inter-class separation. This is very unsuitable for face recognition tasks. (It is possible that the eigenvector distances of the same kind are larger than the eigenvectors of different classes.)

# ***Softmax + Contrastive Loss***

Contrastive Loss



The Euclidean distance of similar features is as small as possible, and the Euclidean distance distance of different types of features is greater than margin. Contrastive Loss also constrains intra-class compactness and inter-class separation.

Reference:

* Sun Y, Chen Y, Wang X, et al. **Deep learning face representation by joint identification-verification** [C]// NIPS, 2014.
* Sun Y, Wang X, Tang X. **Deeply learned face representations are sparse, selective, and robust** [C]// CVPR, 2015.

25 patches trained 25 CNNs, feature joint PAC dimension reduction, training Joint Bayesian classification, DeepID2 of Softmax + Contrast Loss reached 99.15% on LFW, and DeepID2+ of multi-layer + supervision information reached 99.47, if using larger deep CNN DeepID3, it reaches 99.53%

Disadvantage:

The DeepID series is an early deep learning face recognition method, but the code is not open source, and the deep feature is multi-patch combination, training the classifier is also cumbersome and impractical.

# ***Largin Angular margin***

1.Large-Margin Softmax Loss

2.SphereFace is an improvement of L-Softmax, it normalize the weight W, so that the training is more concentrated on optimizing the depth feature map and the feature vector angle, reducing the sample size imbalance problem.

Both L-Softmax and SphereFace use multiplicative margins to make the different categories more separate, with feature similarities using cos distance. It should be noted that these two loss direct trainings are difficult to converge, and the annealing optimization strategy is used in actual training.

Reference:

Liu W, Wen Y, Yu Z, et al. **Large-Margin Softmax Loss for Convolutional Neural Networks** [C]// ICML, 2016.( Large-Margin Softmax Loss)

Liu W, Wen Y, Yu Z, et al. **SphereFace: Deep Hypersphere Embedding for Face Recognition** [C]// CVPR. 2017.(SphereFace)

I recommend we try this way.

Github:

<https://github.com/wy1iu/LargeMargin_Softmax_Loss>

<https://github.com/wy1iu/sphereface>

# ***Center Loss***

Learn a center for each category and pull all feature vectors for each category toward the corresponding category center for use with Softmax

Reference:

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

Disadvantage:

Center Loss needs to maintain a category center for each category. When the number of categories is large (>10000), this memory consumption is very impressive, and the memory requirements of the GPU are high.

Github:

<https://github.com/ydwen/caffe-face>

# ***others***

**Center Invariant Loss**: Softmax + Center loss

Except for Center loss, each class is pulled toward the category center, and the category center of each class is additionally constrained to a fixed radius hypersphere. This radius is the modulus mean of the centers of all categories, and the feature area caused by the category imbalance is mitigated. difference.

Reference:

Wu Y, Liu H, Li J, et al. **Deep Face Recognition with Center Invariant Loss** [C]// ACM MM workshop, 2017.

**Range loss:**

Reference:

Zhang X, Fang Z, Wen Y, et al. Range loss for deep face recognition with long-tail [C]// ICCV, 2017.

**Ring loss:**Softmax+ SphereFace

Reference:

Yutong Zheng, Dipan K. Pal and Marios Savvides. Ring loss: Convex Feature Normalization for Face Recognition [C]// CVPR, 2018.

Tomorrow’ plan

Implementation of SphereFace