ArcFace: Additive Angular Margin Loss for Deep Face Recognition

**Author: Jiankang Deng, Jia Guo, Niannan Xue, Stefanos Zafeiriou**

**Conference: CVPR**

**Year: 2019**

*One of the main challenges in feature learning using DCNNs for largescale face recognition is the design of appropriate loss functions that enhance discriminative power. Recently, most research is to incorporate margins in well-established loss functions to maximize face class separability. Deng et al. proposes an Additive Angular Margin Loss (ArcFace) to obtain highly discriminative features for face recognition. The proposed ArcFace has a clear geometric interpretation due to the exact correspondence to the geodesic distance on the hypersphere. ArcFace consistently outperforms the state-of-the-art face recognition methods on over 10 face recognition benchmarks and can be easily implemented with negligible computational overhead.*

Recently, there are two main lines of research to train DCNNs for face recognition. One is that learn directly an embedding, such as the triplet loss. The other is that train a multi-class classiﬁer, such by using a softmax classiﬁer. Both of them can obtain excellent performance while they also have some drawbacks. For the triplet loss, there is a combinatorial explosion in the number of face triplets especially for large-scale datasets, leading to a signiﬁcant increase in the number of iteration steps and semi-hard sample mining is difficult for effective model training. For the softmax loss, the size of the linear transformation matrix increases linearly with the identities number and the learned features are not discriminative enough for the open-set face recognition problem.

To take the advantages and avoid the disadvantages of softmax loss, several variants have been proposed to enhance the discriminative power. Wen et al. pioneered the centre loss to obtain intra-class compactness while it is extremely difficult to update the actual centres as the number of face classes has dramatically increased. Liu et al. proposed Sphereface which introduced the important idea of angular margin. However, it required a series of approximations, which resulted in an unstable training of the network. Wang et al. proposed CosFace which directly adds cosine margin penalty to the target logit, which maximize the intra-class compactness and inter-class discrepancy in Cosine space but clearer geodesic interpretation can be obtained in angular space.

In order to tackle these problems, Additive Angular Margin Loss (ArcFace) is proposed to further improve the discriminative power of the face recognition model and to stabilize the training process. Compared with CosFace, ArcFace has a clear geometric interpretation due to the exact correspondence to the geodesic distance on the hypersphere in form of angular space. To make the derivation process more understandable, we first introduce the most widely used classiﬁcation loss function, softmax loss:

Where d is the embedding feature dimension, denotes the deep feature of the i-th sample, belonging to the -th class. denotes the j-th column of the weight and is the bias term. The batch size and the class number are N and n, respectively. For simplicity, fix the bias . Then, transform the logit as , where is the angle between the weight and the feature . Fix the individual weight by normalisation. Fix the embedding feature by normalisation and re-scale it to s. The normalisation step on features and weights makes the predictions only depend on the angle between the feature and the weight. The learned embedding features are thus distributed on a hypersphere with a radius of s. Then the loss function will be in the form of .

As the embedding features are distributed around each feature center on the hypersphere, an additive angular margin penalty *m* is added between and to simultaneously enhance the intra-class compactness and inter-class discrepancy. Then the loss function will be in the form of , which is used in ArcFace.

It is shown that ArcFace can achieve better performance compared to other state-of-the art architectures. ArcFace trained on MS1MV2 with ResNet100 beats the baselines by a significant margin on both LFW and YTF. The performance of ArcFace on datasets CPLFW and CALFW shows higher pose and age variations with same identities from LFW. Among all of the open-sourced face recognition models, it is evaluated as the top-ranked face recognition model. On MegaFace, ArcFace still clearly outperforms CosFace and achieves the best performance on both verification and identification. When VGG2 dataset is employed as the training data and the ResNet50 as the embedding network to train ArcFace, it can obviously boost the performance on both IJB-B and IJB-C (about 3 ∼ 5%). When combining all identities from MS1MV2 and Asian celebrities from DeepGlint, ArcFace achieves the best identification performance 84.840%. The MLP learned on the iQIYI-VID training set significantly boosts the MAP by 6.60%.

According to the design of Arcface, it directly optimizes the geodesic distance margin by virtue of the exact correspondence between the angle and arc in the normalized hypersphere. And it adds negligible computational complexity during training while it can achieve state-of-the-art performance on ten face recognition benchmarks including large-scale image and video datasets. Besides, ArcFace does not need to be combined with other loss functions in order to have stable performance, and can easily converge on any training datasets.

In view of the advantages of ArcFace, it can not only be used in face recognition but also in other classification tasks. In addition, ArcFace has achieve the-state-of-art performance in angular space. Therefore, future work should focus on other fields to better performance.

By

许俊杰 16319058

黄俊凯 16340082

陈亚楠 16340041