

Jiankang Deng, Jia Guo, Niannan Xue

Imperial College London, Insight Face

Abstract

- 이 논문에서는 기존의 loss function을 보다 조금 더 discriminative 한 feature를 뽑을 수 있는 방법에 대하여 제안하였고, SOTA를 달성하였다.
- 개인적으로 SOTA를 달성한 부분 외에 특별한 contribution은 없는 것 같다.

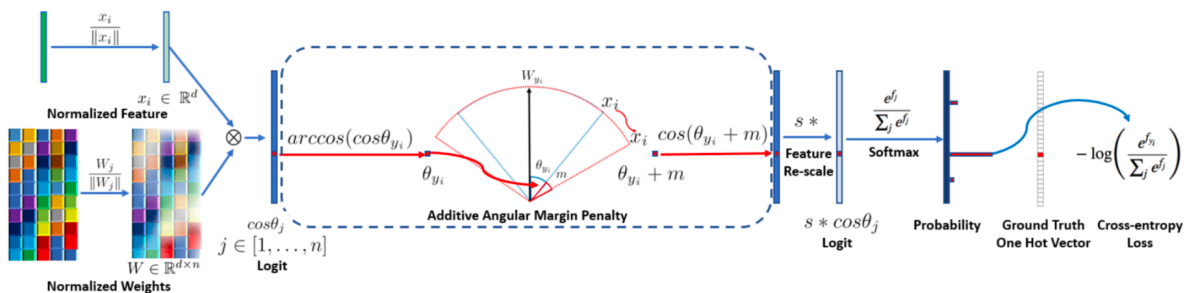
Introduction

- 얼굴인식을 위한 DCNN 의 2가지 주요 흐름.
 - Softmax
 - 학습데이터와 테스트데이터의 Class label 이 같은 경우일 때만 효율적으로 동작한다.
 - Learn directly an embedding (ex: FaceNet - Triplet Loss)
 - triplet 을 만드는 부분에서 연산량이 매우 많이 증가한다.
 - semi-hard sample mining 을 구성하는 것이 어렵다.

Proposed Approach

ArcFace

- 전체 구조도



- 기존 많이 사용되던 Softmax 의 수식은 다음과 같다.

$$L_1 = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{W_{y_i}^T x_i + b_{y_i}}}{\sum_{j=1}^n e^{W_j^T x_i + b_j}}$$

$x_i \in R^d$ 은 y_i 클래스에 속해있는 i 번째 샘플의 deep feature를 나타낸다. embedding dimension d 는 512로 설정하였다. W_j 는 weight 의 j 번째 column 을 나타내고, b 는 bias 를 나타낸다. N, n 은 각각 batch size, 클래스의 갯수를 나타낸다.

- Softmax loss function 은 inter-class 간의 차이를 늘려주고, intra-class 끼리 차이를 줄여주는 부분에 대해서는 최적화가 잘 이루어지지 못한다. 따라서 이러한 부분을 해결하기위해 다음과 같은 과정을 거친다.

1. b_j 를 0 으로 normalization 을 수행한다.
2. $W_j^T x_i = \|W_j\| \|x_i\| \cos \theta_j$ 로 변경한다.
3. $\|W_j\| = 1$ 로 normalization 작업을 수행한다.
4. $\|x_i\| = 1$ 로 L_2 Normalization 을 수행한 이후에, $\|x_i\| = s$ 로 re-scaling 해준다. (SphereFace 와 유사)

$$L_2 = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{s \cos \theta_{y_i}}}{e^{s \cos \theta_{y_i}} + \sum_{j=1, j \neq y_i}^n e^{s \cos \theta_j}}$$

5. x_i 와 W_{y_i} 사이에 angular margin penalty term 을 추가해준다.

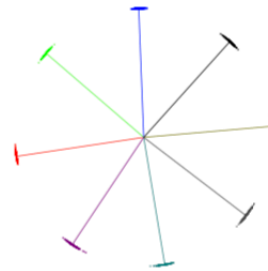
(normalized hypersphere 상에서 geodesic distance margin penalty 를 추가한 형태이기 때문에 이름을 ArcFace 라고 지었다.)

$$L_3 = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{s(\cos(\theta_{y_i} + m))}}{e^{s(\cos(\theta_{y_i} + m))} + \sum_{j=1, j \neq y_i}^n e^{s \cos \theta_j}}$$

- 아래 그림은 8개의 클래스를 갖는 face image 를 Softmax, ArcFace 를 사용하여 Feature embedding 한 결과를 보여준다. (점들은 샘플을 나타내고, 직선은 클래스 중점 방향을 나타낸다)



(a) Softmax

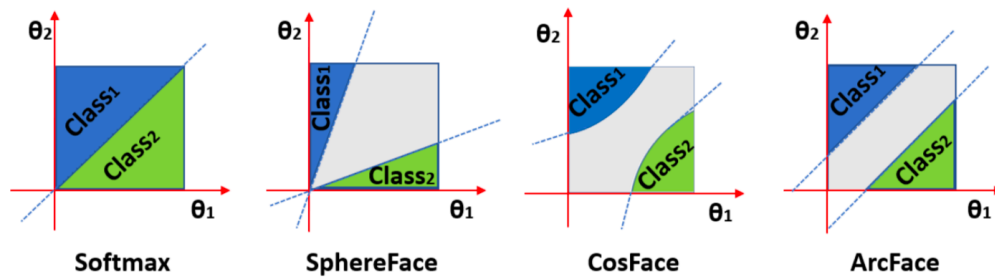


(b) ArcFace

Comparison with SphereFace and CosFace

- **Numerical Similarity**
 - SphereFace, ArcFace, CosFace 는 margin penalty 를 사용하여 target logit 을 제한함으로써 inter-class diversity 와 intra-class compactness 의 효과를 얻는다. (ex: multiplicative angular margin, additive angular margin, additive cosine margin)
- **Geometric Difference**

- SphereFace, ArcFace, CosFace 가 numerical similarity 를 갖고 있지만, geometric 한 특성은 서로 다르다. 아래 그림을 보면 각 방법이 서로 다른 decision boundary를 갖고있다는 것을 확인할 수 있다.



- SphereFace와, CosFace 는 nonlinear 한 angular margin 을 갖지만, ArcFace 는 linear 한 angular margin 을 갖는것을 알 수 있다. (non linear 한 특성을 없앴으로써, 트레이닝 초기에 발생할 수 있는 divergence 를 방지할 수 있다.)

Experiments

- 다양한 실험들을 굉장히 많이 수행하였다.
- 아래 표는 Face Dataset 에 대한 정보를 보여준다.

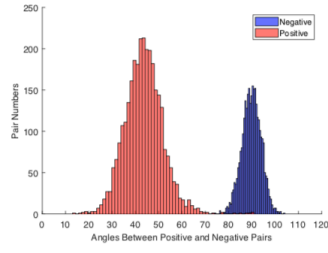
Datasets	#Identity	#Image/Video
CASIA [43]	10K	0.5M
VGGFace2 [6]	9.1K	3.3M
MS1MV2	85K	5.8M
MS1M-DeepGlint [2]	87K	3.9M
Asian-DeepGlint [2]	94 K	2.83M
LFW [13]	5,749	13,233
CFP-FP [30]	500	7,000
AgeDB-30 [22]	568	16,488
CPLFW [48]	5,749	11,652
CALFW [49]	5,749	12,174
YTF [40]	1,595	3,425
MegaFace [15]	530 (P)	1M (G)
IJB-B [39]	1,845	76.8K
IJB-C [21]	3,531	148.8K
Trillion-Pairs [2]	5,749 (P)	1.58M (G)
iQIYI-VID [20]	4,934	172,835

- Experimental Settings
 - 얼굴 내의 5개의 키포인트를 활용하여 112 x 112 사이즈의 normalized crop 이미지를 사용했다.
 - Backbone : ResNet50, ResNet100
 - 512-D embedding 을 만들때, BN-Dropout-FC-BN 구조를 사용하였다.

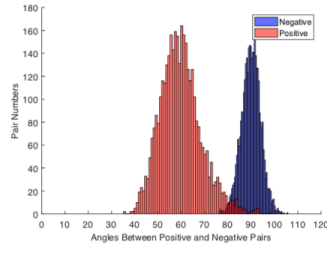
- batch-size: 512, momentum: 0.9, weight decay $5e-4$, s: 64, m: 0.5 를 사용하였으며, CASIA 데이터셋에서는 learning rate 를 0.1부터 시작하였다. (10, 20, 28K 마다 10씩 나눠주는 작업을 수행)
- Verification Results [CASIA-webface, ResNet50 사용]

Loss Functions	LFW	CFP-FP	AgeDB-30
ArcFace (0.4)	99.53	95.41	94.98
ArcFace (0.45)	99.46	95.47	94.93
ArcFace (0.5)	99.53	95.56	95.15
ArcFace (0.55)	99.41	95.32	95.05
SphereFace [18]	99.42	-	-
SphereFace (1.35)	99.11	94.38	91.70
CosFace [37]	99.33	-	-
CosFace (0.35)	99.51	95.44	94.56
CM1 (1, 0.3, 0.2)	99.48	95.12	94.38
CM2 (0.9, 0.4, 0.15)	99.50	95.24	94.86
Softmax	99.08	94.39	92.33
Norm-Softmax (NS)	98.56	89.79	88.72
NS+Intra	98.75	93.81	90.92
NS+Inter	98.68	90.67	89.50
NS+Intra+Inter	98.73	94.00	91.41
Triplet (0.35)	98.98	91.90	89.98
ArcFace+Intra	99.45	95.37	94.73
ArcFace+Inter	99.43	95.25	94.55
ArcFace+Intra+Inter	99.43	95.42	95.10
ArcFace+Triplet	99.50	95.51	94.40

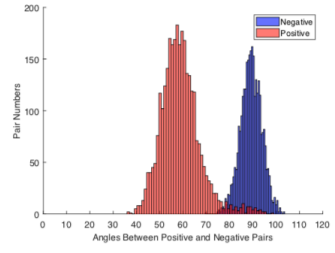
- ArcFace 의 괄호안의 숫자는 margin 값을 나타낸다.
- CM은 SphereFace, ArcFace, CosFace 를 섞은것을 의미한다.
- Angle distribution of both positive and negative pairs on LFW, CFP-FP, AgeDB-30, YTF, CPLFW, CALFW (Red: positive pairs, Blue: negative pairs) [MS1MV2, ResNet100, ArcFace 사용]



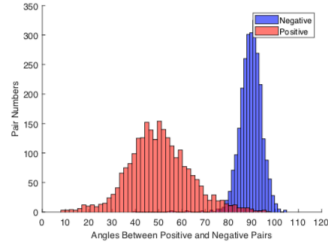
(a) LFW (99.83%)



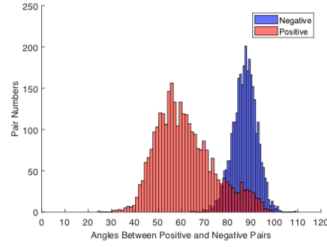
(b) CFP-FP (98.37%)



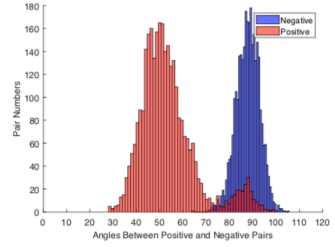
(c) AgeDB (98.15%)



(d) YTF (98.02%)

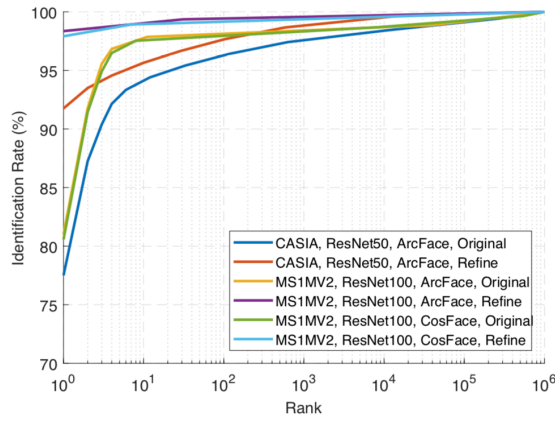


(e) CPLFW (92.08%)

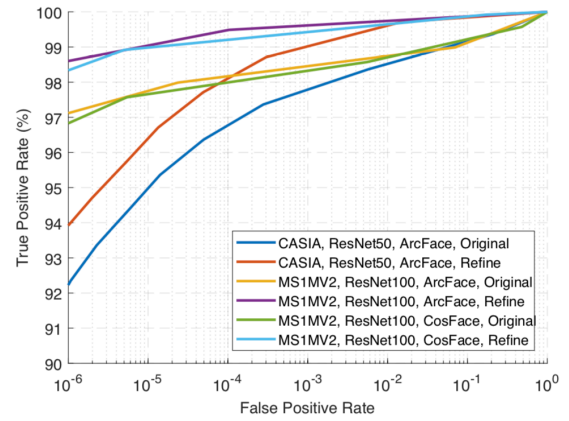


(f) CALFW (95.45%)

- CMC & ROC curves of different models



(a) CMC



(b) ROC

- 1:1 verification TAR (@FAR=1e-4) on the IJB-B and IJB-C dataset.

Method	IJB-B	IJB-C
ResNet50 [6]	0.784	0.825
SENet50 [6]	0.800	0.840
ResNet50+SENet50 [6]	0.800	0.841
MN-v [42]	0.818	0.852
MN-vc [42]	0.831	0.862
ResNet50+DCN(Kpts) [41]	0.850	0.867
ResNet50+DCN(Divs) [41]	0.841	0.880
SENet50+DCN(Kpts) [41]	0.846	0.874
SENet50+DCN(Divs) [41]	0.849	0.885
VGG2, R50, ArcFace	0.898	0.921
MS1MV2, R100, ArcFace	0.942	0.956

Conclusions

- Additive Angular Margin Loss 를 통해서 Face recognition 에서 SOTA를 달성하였다.