SphereFace: Deep Hypersphere Embedding for Face Recognition

Weiyang Liu[1] Yandong Wen[2] Zhiding Yu[2] Ming Li[3] Bhiksha Raj[2] Le Song[1]
[1]Georgia Institute of Technology [2]Carnegie Mellon University [3]Sun Yat-Sen University
wyliu@gatech.edu, {yandongw,yzhiding}@andrew.cmu.edu, lsong@cc.gatech.edu

Abstract

we propose the angular softmax (A-Softmax) loss that enables convolutional neural networks (CNNs) to learn angularly discriminative features.

1. Introduction

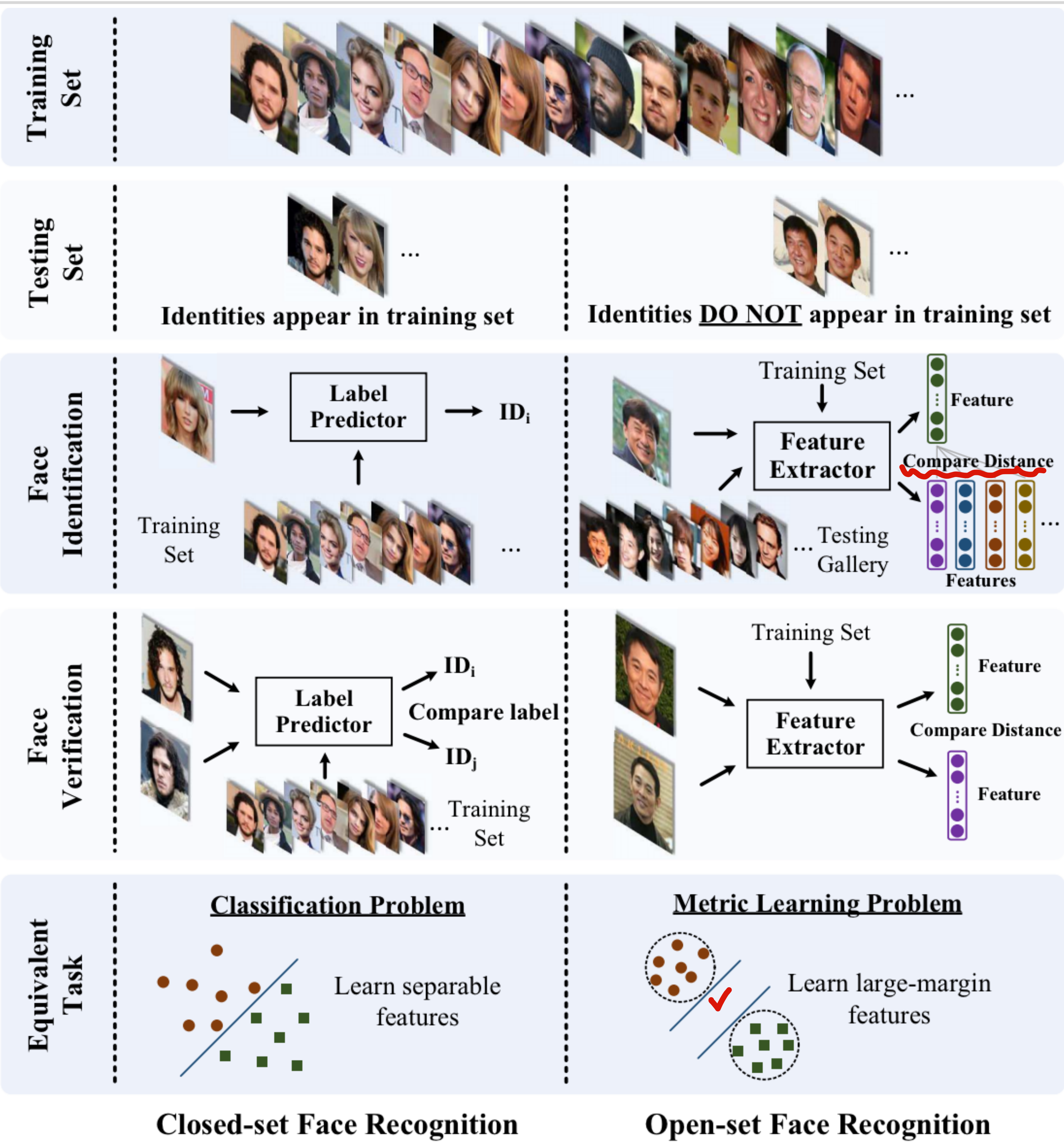
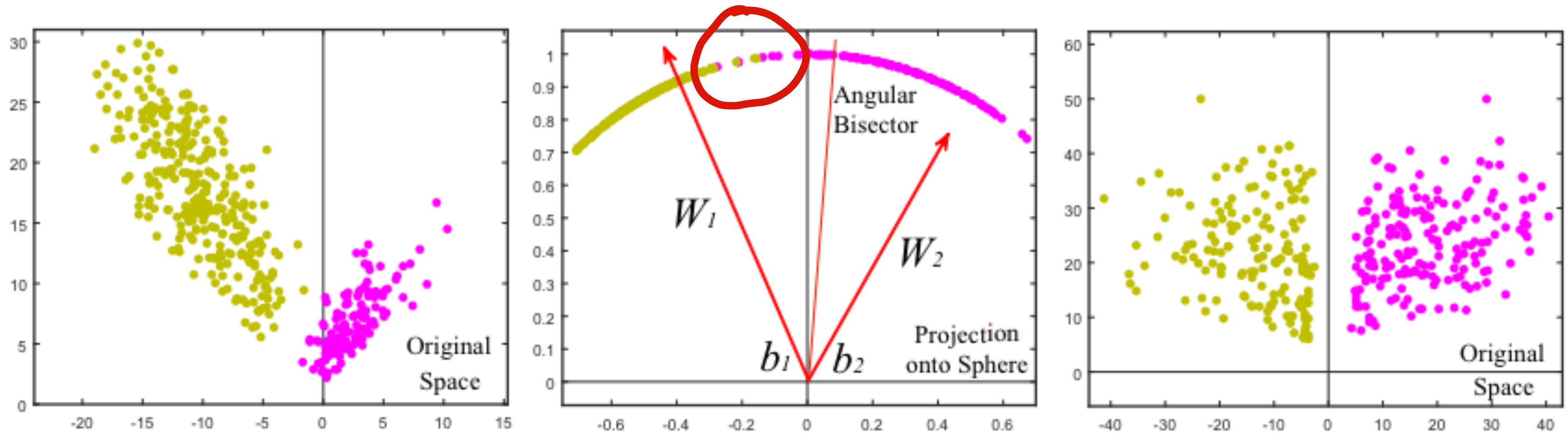
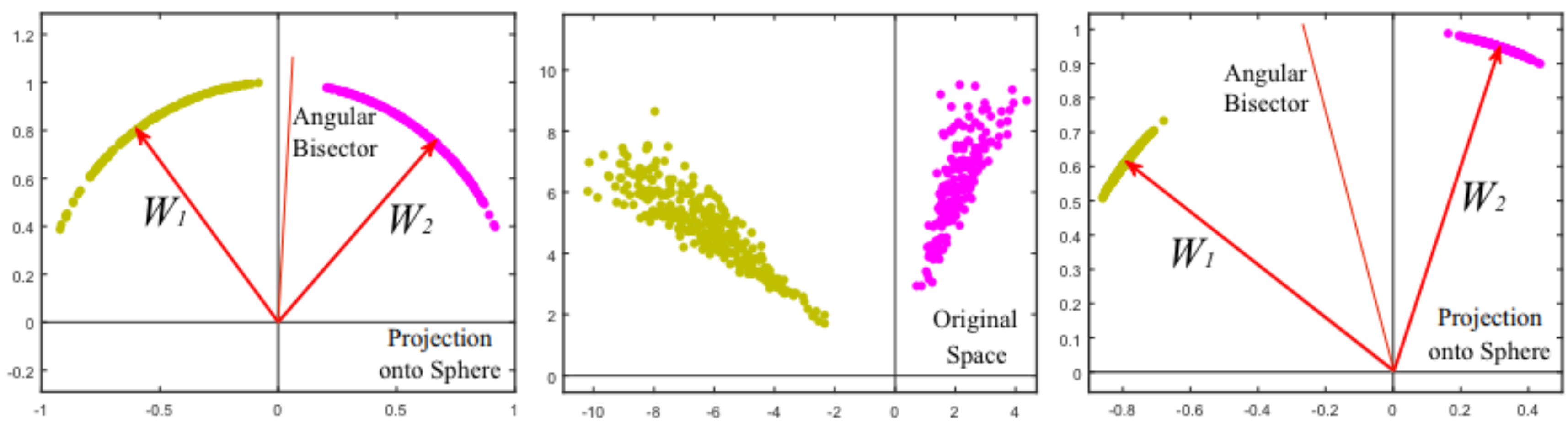


Figure 1: Comparison of open-set and closed-set face recognition.



(a) Original Softmax Loss (b) Original Softmax Loss (c) Modified Softmax Loss



(d) Modified Softmax Loss (e) A-Softmax Loss (f) A-Softmax Loss

Yellow dots represent the first class face features, while purple dots represent the second class face feature

2. Related Work

Wi and bi are weights and bias of last fully connected layer corresponding to class i, respectively

softmax loss, multi-class classification problem

feature extracted from face data

$$p_1 = \frac{\exp(\boldsymbol{W}_1^T\boldsymbol{x} + b_1)}{\exp(\boldsymbol{W}_1^T\boldsymbol{x} + b_1) + \exp(\boldsymbol{W}_2^T\boldsymbol{x} + b_2)}$$

$$p_2 = \frac{\exp(\boldsymbol{W}_2^T\boldsymbol{x} + b_2)}{\exp(\boldsymbol{W}_1^T\boldsymbol{x} + b_1) + \exp(\boldsymbol{W}_2^T\boldsymbol{x} + b_2)}$$

class 1 if p1 >p2 and class 2 if p1 <p2.

decision boundary

$$(\boldsymbol{W}_1 - \boldsymbol{W}_2)\boldsymbol{x} + b_1 - b_2 = 0.$$

In CNNs, $\boldsymbol{f}$ is usually the output of a fully connected layer $\boldsymbol{W}$, so $f_j = \boldsymbol{W}_j^T \boldsymbol{x}_i + b_j$ and $f_{y_i} = \boldsymbol{W}_{y_i}^T \boldsymbol{x}_i + b_{y_i}$ where $\boldsymbol{x}_i$, $\boldsymbol{W}_j$, $\boldsymbol{W}_{y_i}$ are the i-th training sample, the j-th and y_i-th column of $\boldsymbol{W}$ respectively.

g.t

$$L = \frac{1}{N}\sum_i L_i = \frac{1}{N}\sum_i -\log\Big(\frac{e^{f_{y_i}}}{\sum_j e^{f_j}}\Big)$$

$$L_i = -\log\Big(\frac{e^{\boldsymbol{W}_{y_i}^T \boldsymbol{x}_i + b_{y_i}}}{\sum_j e^{\boldsymbol{W}_j^T \boldsymbol{x}_i + b_j}}\Big)$$

모든 classes

$$= -\log\Big(\frac{e^{\|\boldsymbol{W}_{y_i}\|\|\boldsymbol{x}_i\|\cos(\theta_{y_i,i}) + b_{y_i}}}{\sum_j e^{\|\boldsymbol{W}_j\|\|\boldsymbol{x}_i\|\cos(\theta_{j,i}) + b_j}}\Big)$$

$\|W\| = 1$

$b = 0$

$$L_{\text{modified}} = \frac{1}{N}\sum_i -\log\Big(\frac{e^{\|\boldsymbol{x}_i\|\cos(\theta_{y_i,i})}}{\sum_j e^{\|\boldsymbol{x}_i\|\cos(\theta_{j,i})}}\Big)$$

$$L_{\text{ang}} = \frac{1}{N}\sum_i -\log\Big(\frac{e^{\|\boldsymbol{x}_i\|\cos(m\theta_{y_i,i})}}{e^{\|\boldsymbol{x}_i\|\cos(m\theta_{y_i,i})} + \sum_{j \neq y_i} e^{\|\boldsymbol{x}_i\|\cos(\theta_{j,i})}}\Big) \tag{6}$$

y_i (gt)는 제외한 다른 feature와 margin을 줌

Loss Function	Decision Boundary
Softmax Loss	$(\boldsymbol{W}_1 - \boldsymbol{W}_2)\boldsymbol{x} + b_1 - b_2 = 0$
Modified Softmax Loss	$\|\boldsymbol{x}\|(\cos\theta_1 - \cos\theta_2) = 0$
A-Softmax Loss	$\|\boldsymbol{x}\|(\cos m\theta_1 - \cos\theta_2) = 0$ for class 1 $\|\boldsymbol{x}\|(\cos\theta_1 - \cos m\theta_2) = 0$ for class 2

Class 1 이려면 $\theta_1 < \theta_2$

decision boundaries can greatly affect the feature distribution

Assume a learned feature x from class 1 is given and θi

is the angle between x and Wi, it is known that the modified

softmax loss requires cos(θ1)>cos(θ2) to correctly classify x.

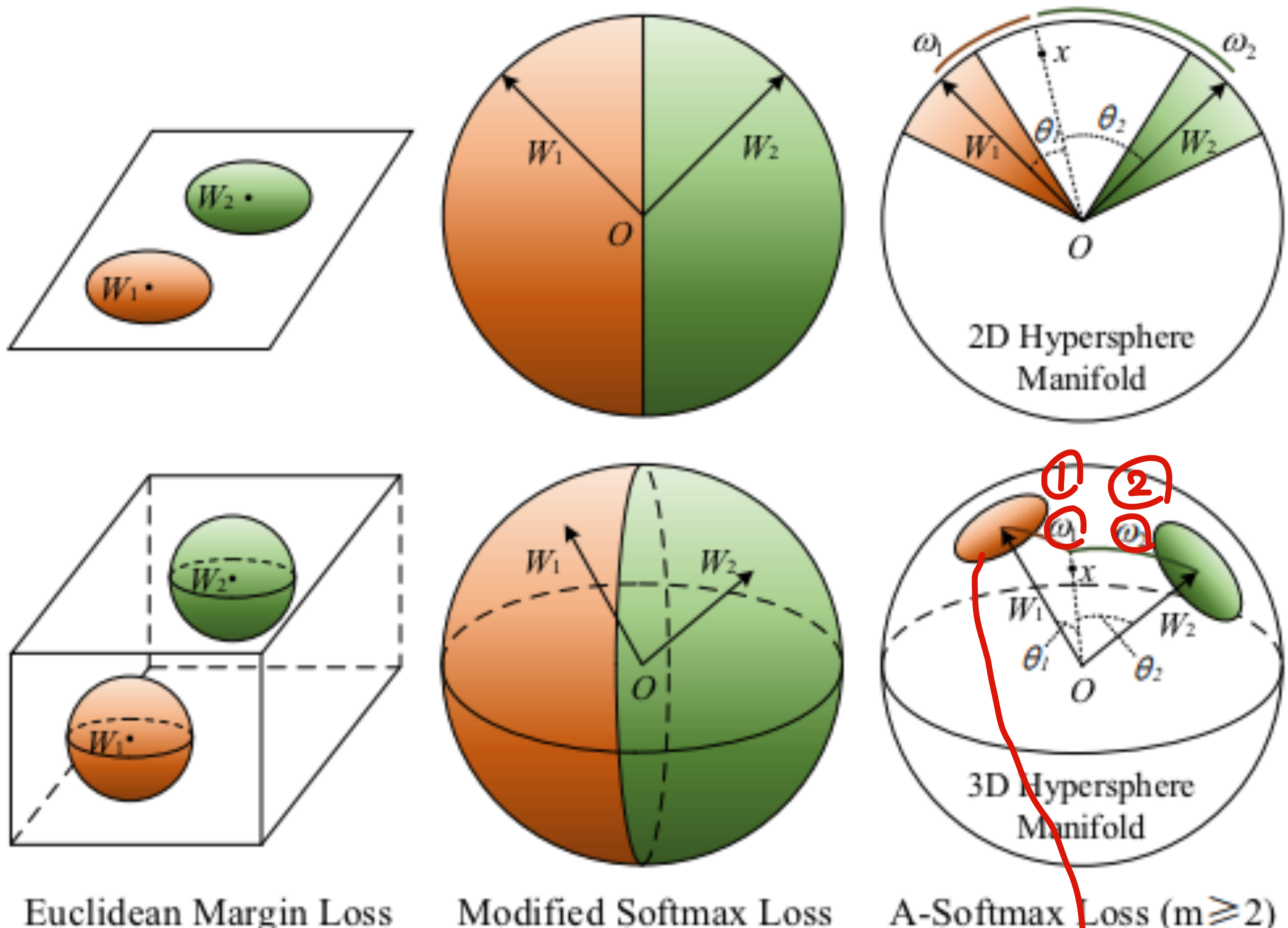
The decision boundary for class 1 is cos(mθ1)= cos(θ2). = $\cos(\theta_3)$ = $\cos(\theta_4)$ = ··

the decision boundary for class 2 is cos(mθ2)=cos(θ1) = $\cos(\theta_3)$ = $\cos(\theta_4)$ = ...

?

such decision boundaries will produce an angular margin of $\frac{m-1}{m+1}\theta_2^1$ where θ_2^1 is the angle between $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$.

3.3. Hypersphere Interpretation of A-Softmax Loss



Class 1 이려면, $\cos(m\theta_1) - \cos\theta_2 > 0$

$m\theta_1 < \theta_2$ ← $m=2$ 일 때

$\theta_1 < \frac{\theta_2}{2}$

$1\ \theta_1 < 1 \cdot \frac{\theta_2}{2}$

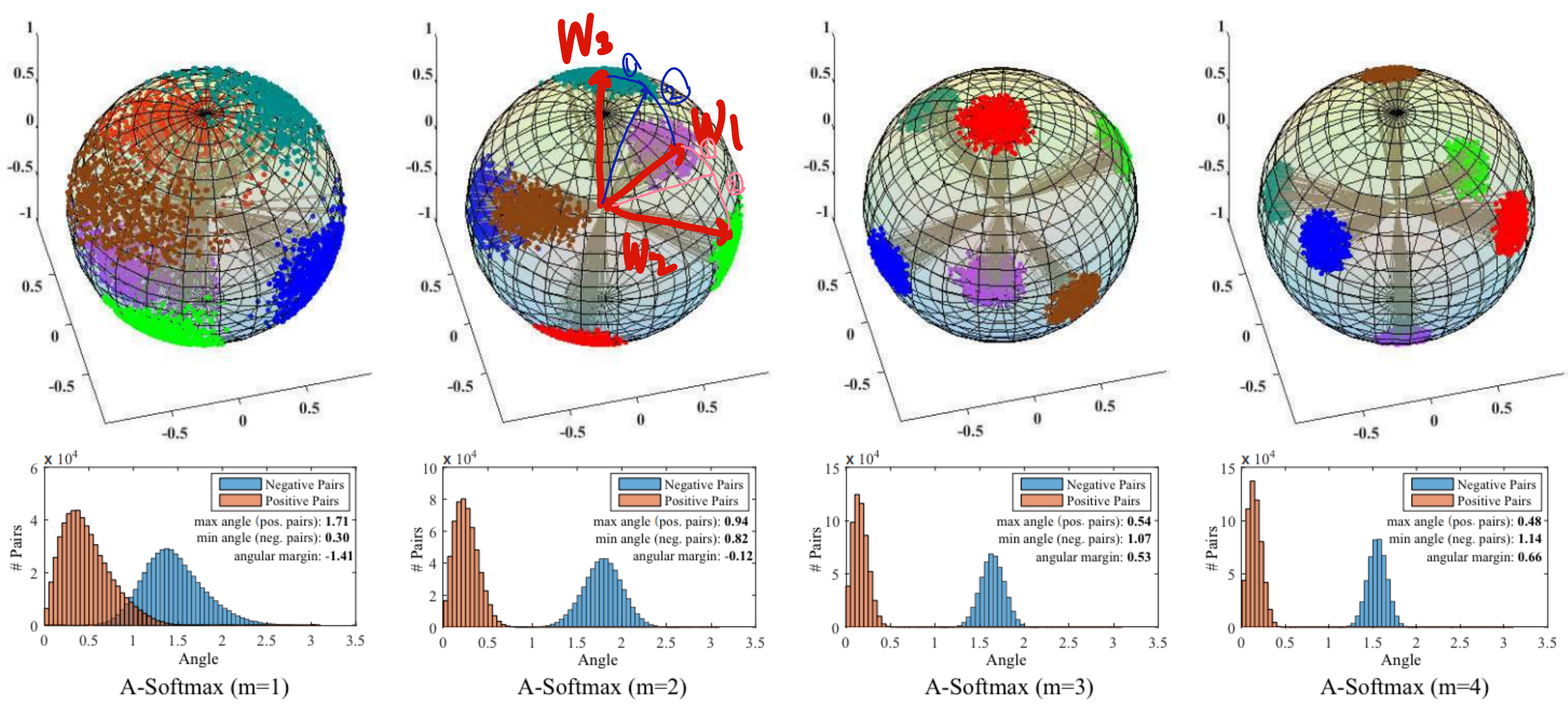
반지름

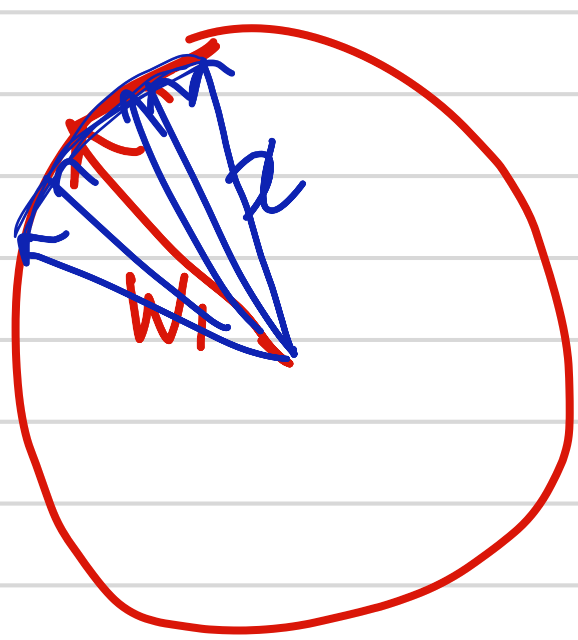
$\therefore w_1 < \frac{w_2}{2}$

이 부분에 찍히게끔 training 됨

4.2. Exploratory Experiments

Dataset	Original	m=1	m=2	m=3	m=4
LFW	97.88	97.90	98.40	99.25	**99.42**
YTF	93.1	93.2	93.8	94.4	**95.0**





Method	Models	Data	LFW	YTF
DeepFace [30]	3	4M*	97.35	91.4
FaceNet [22]	1	200M*	**99.65**	95.1
Deep FR [20]	1	2.6M	98.95	**97.3**
DeepID2+ [27]	1	300K*	98.70	N/A
DeepID2+ [27]	25	300K*	99.47	93.2
Baidu [15]	1	1.3M*	99.13	N/A
Center Face [34]	1	0.7M*	99.28	94.9
Yi et al. [37]	1	WebFace	97.73	92.2
Ding et al. [2]	1	WebFace	98.43	N/A
Liu et al. [16]	1	WebFace	98.71	N/A
Softmax Loss	1	WebFace	97.88	93.1
Softmax+Contrastive [26]	1	WebFace	98.78	93.5
Triplet Loss [22]	1	WebFace	98.70	93.4
L-Softmax Loss [16]	1	WebFace	99.10	94.0
Softmax+Center Loss [34]	1	WebFace	99.05	94.4
SphereFace	1	WebFace	**99.42**	**95.0**

0.49M