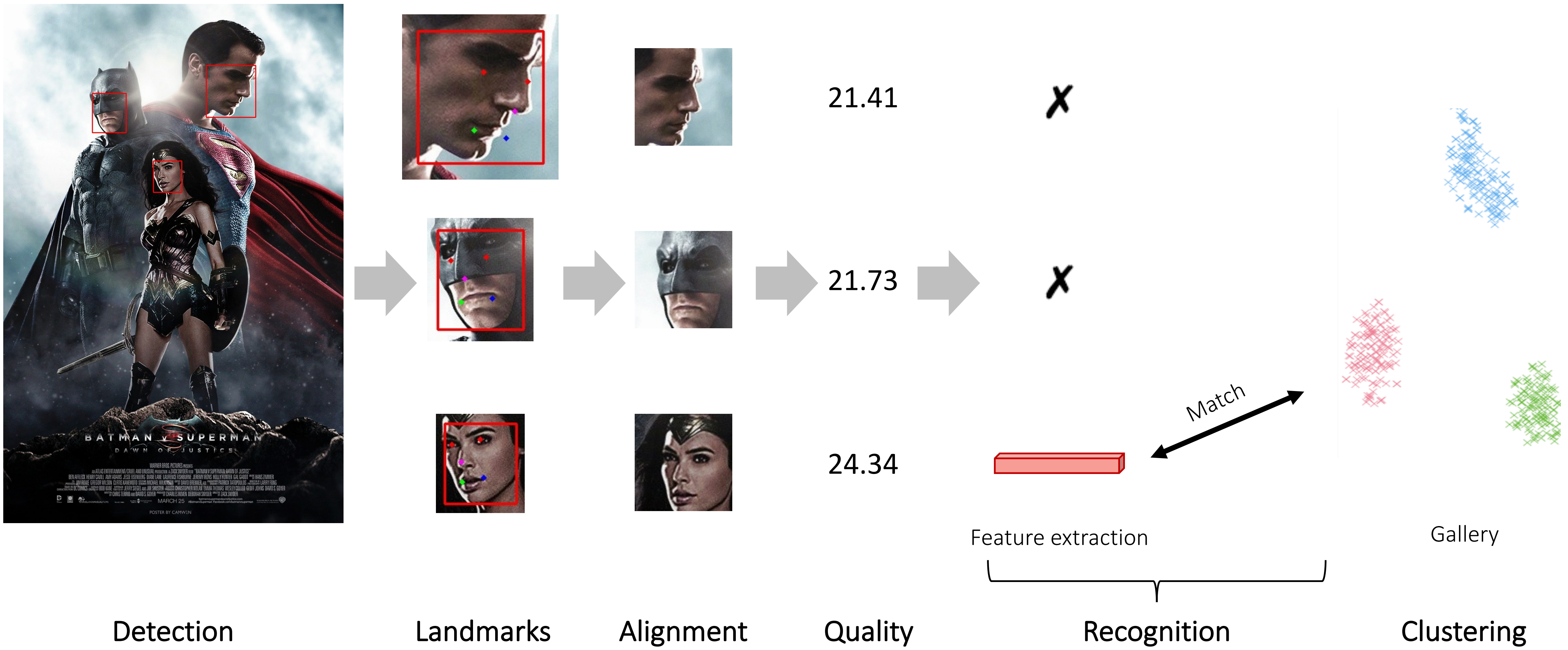
MagFace: A Universal Representation for Face Recognition and Quality Assessment

Qiang Meng, Shichao Zhao, Zhida Huang, Feng Zhou

Algorithm Research, Aibee Inc.

Introduction

A typical face pipeline:

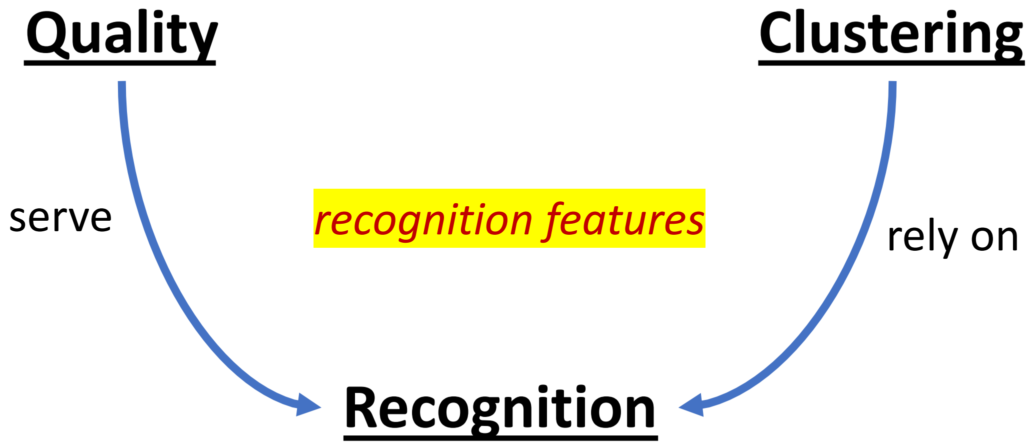


Introduction

Quality *filter out images whose recognition features are not robust.*

Recognition *recognize or verify faces by recognition features.*

Clustering *use the distribution of recognition features to cluster images.*



Goals

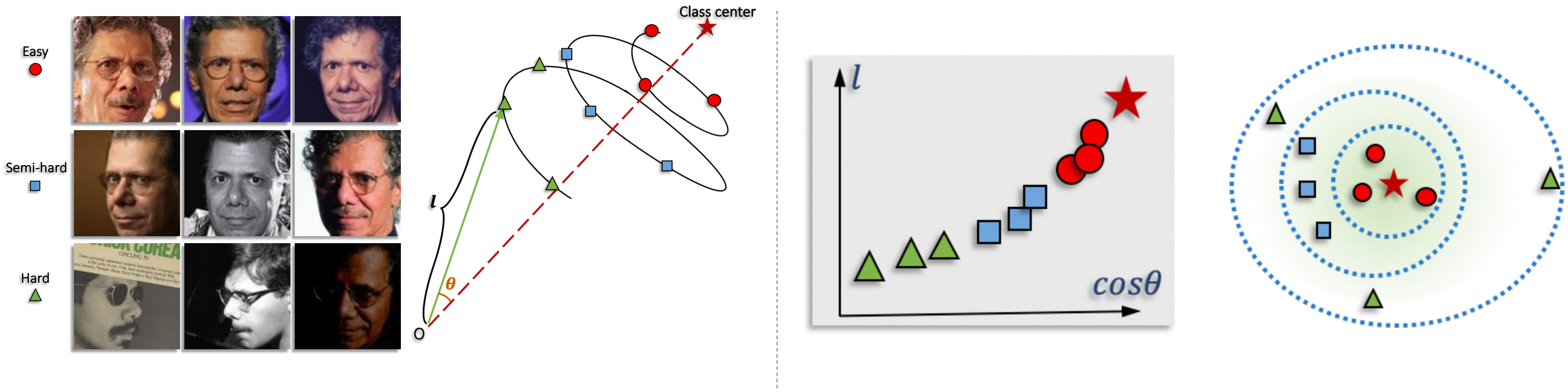
How can we connect all three tasks by recognition features?

MagFace!

A recognition loss which can

1. [quality] tell whether a face image can be recognized.
2. [recognition] fully utilize easy samples while prevent noisy samples from overwhelming the training.
3. [clustering] more suitable distributions.

Goals



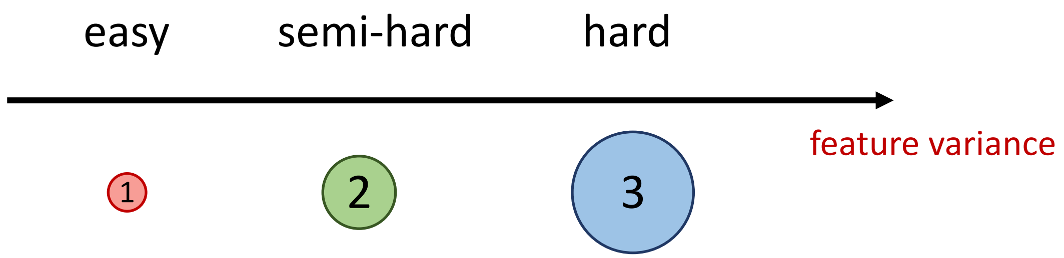
(a) (b)

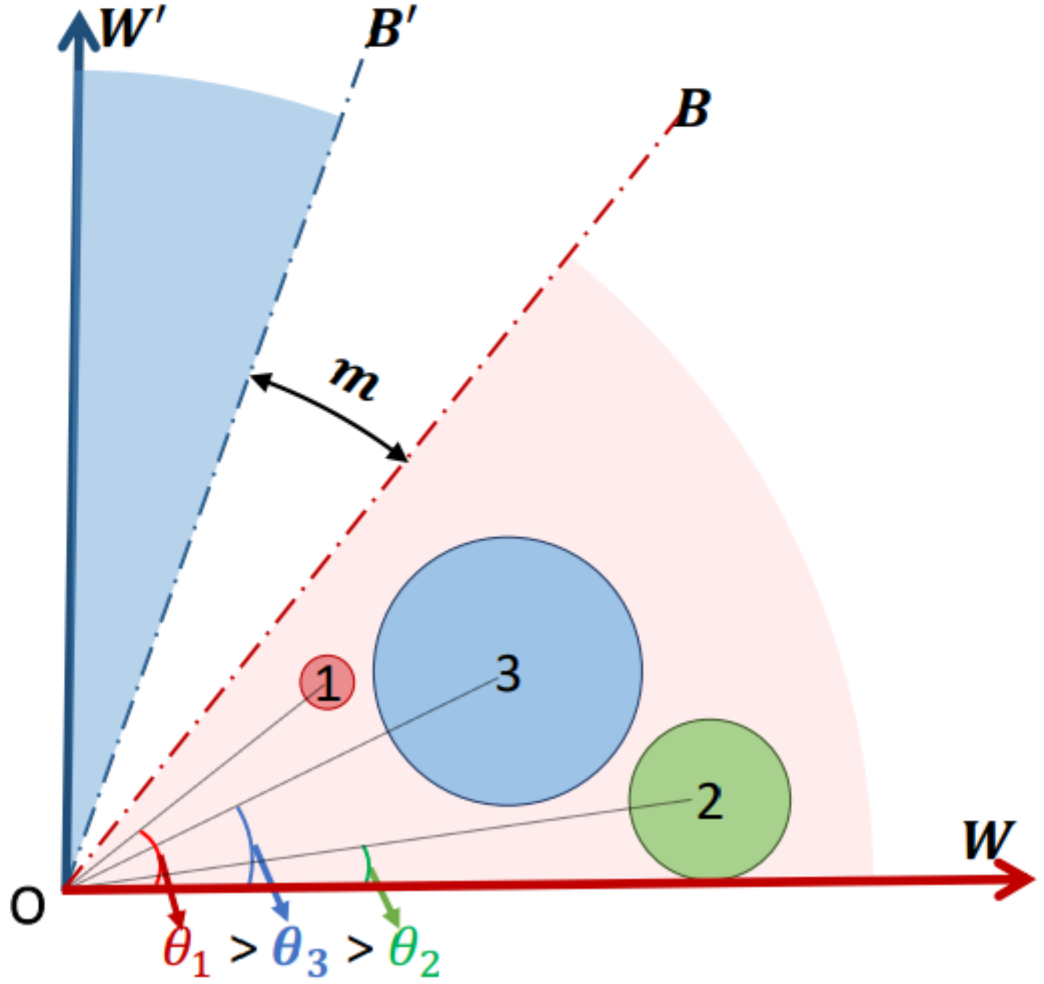
$$feature\ direction \rightarrow recognition$$
$$feature\ magnitude \rightarrow quality$$
$$feature\ distribution \rightarrow clustering$$

Methodology — MagFace

ArcFace :

$$L_i = -\log \frac{e^{s\cos(\theta_{y_i}+m)}}{e^{s\cos(\theta_{y_i}+m)} + \sum_{j\neq y_i} e^{s\cos\theta_j}}$$





Methodology — MagFace

ArcFace :

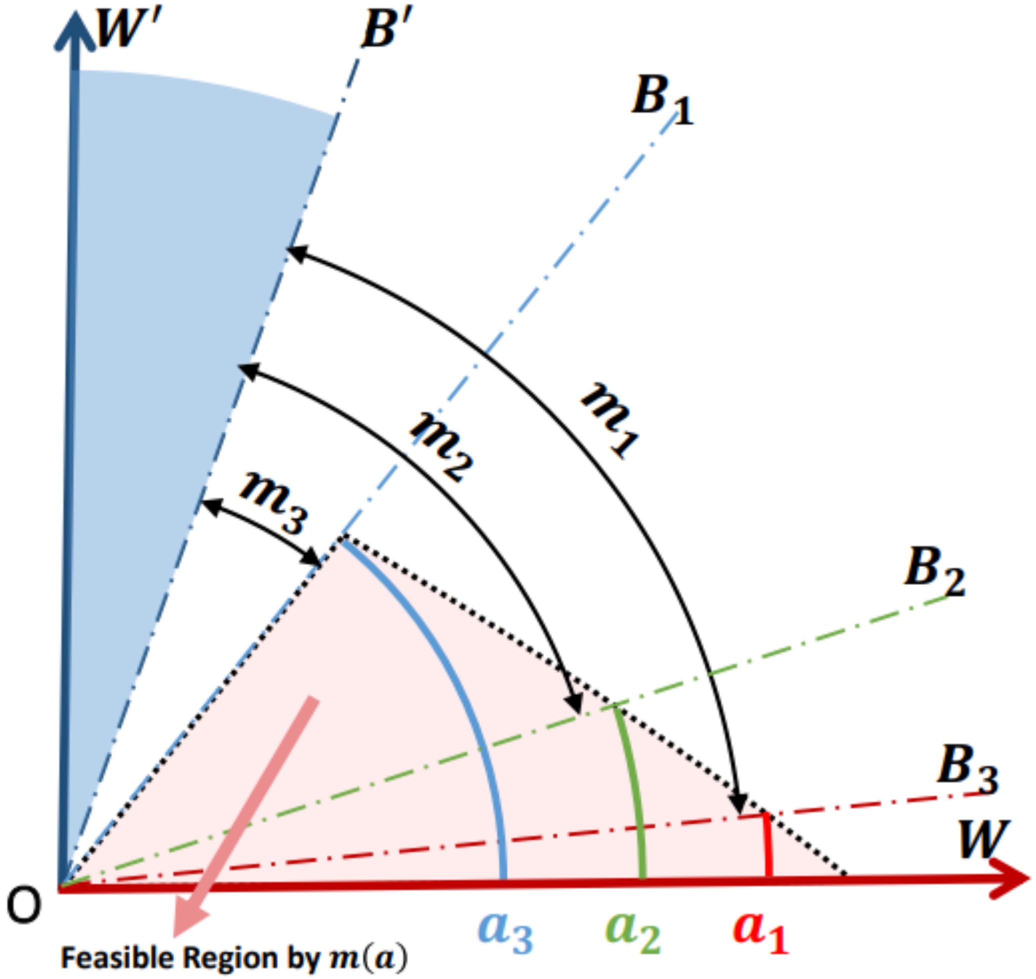
$$L_i = -\log \frac{e^{s\cos(\theta_{y_i}+m)}}{e^{s\cos(\theta_{y_i}+m)} + \sum_{j\neq y_i} e^{s\cos\theta_j}}$$

MagFace :

$$L_i = -\log \frac{e^{s\cos(\theta_{y_i}+m(a_i))}}{e^{s\cos(\theta_{y_i}+m(a_i))} + \sum_{j\neq y_i} e^{s\cos\theta_j}} + \lambda_g \cdot g(a_i)$$

$m(a_i)$ – the magnitude-aware angular margin.

$g(a_i)$ – the regularizer.



Methodology — MagFace

ArcFace :

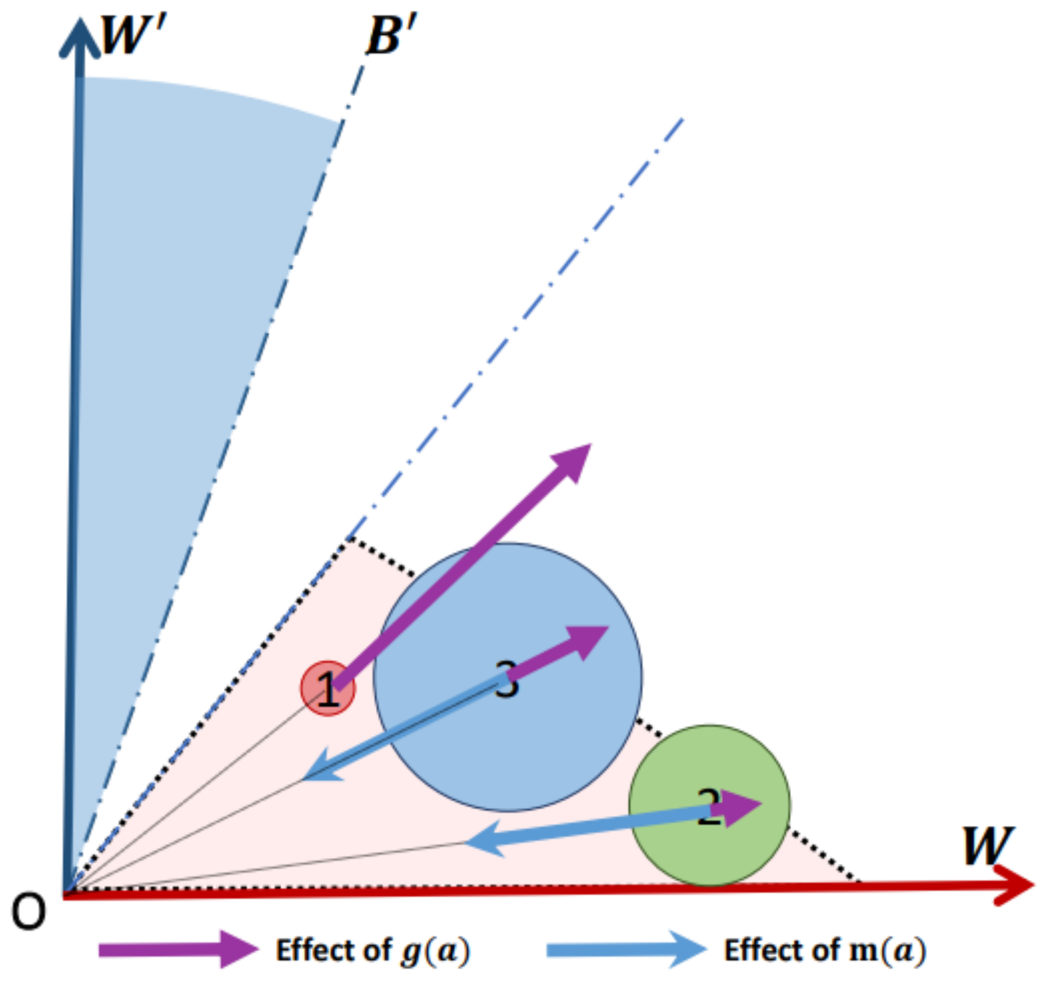
$$L_i = -\log \frac{e^{s\cos(\theta_{y_i}+m)}}{e^{s\cos(\theta_{y_i}+m)} + \sum_{j\neq y_i} e^{s\cos\theta_j}}$$

MagFace :

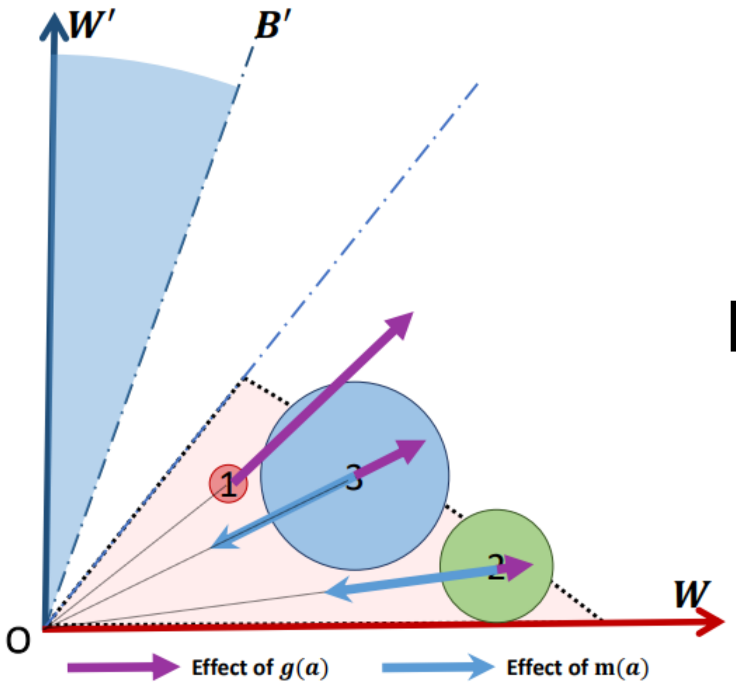
$$L_i = -\log \frac{e^{s\cos(\theta_{y_i}+m(a_i))}}{e^{s\cos(\theta_{y_i}+m(a_i))} + \sum_{j\neq y_i} e^{s\cos\theta_j}} + \lambda_g \cdot g(a_i)$$

$m(a_i)$ – the magnitude-aware angular margin.

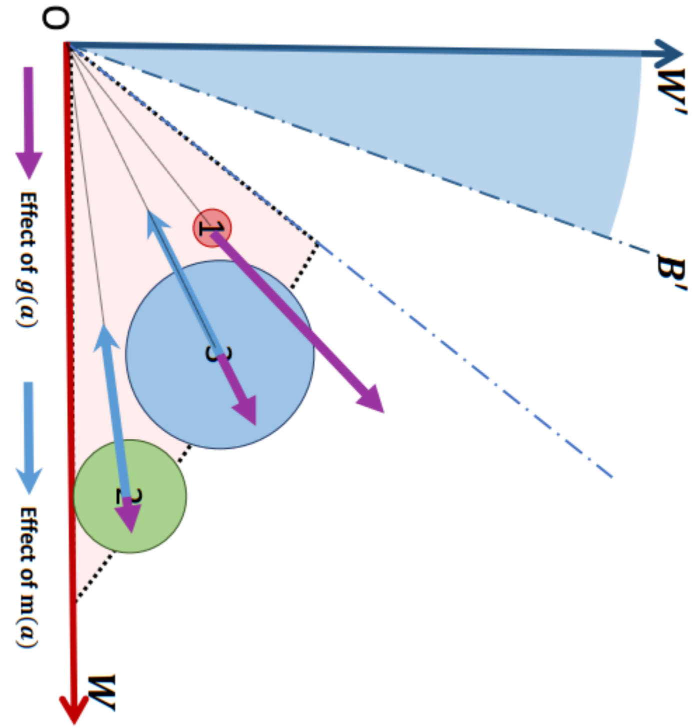
$g(a_i)$ – the regularizer.



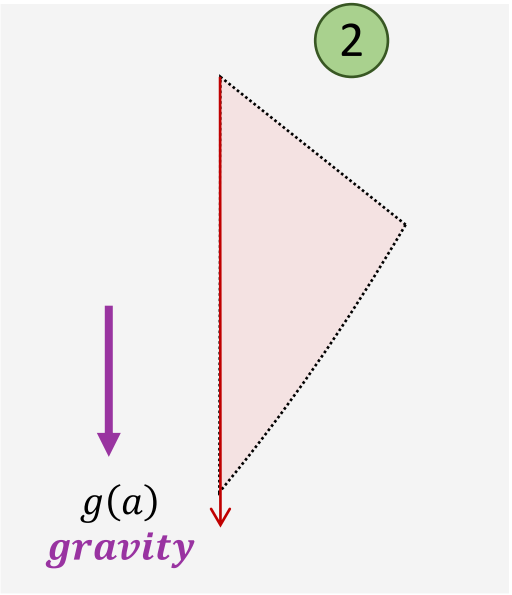
Methodology — MagFace



①

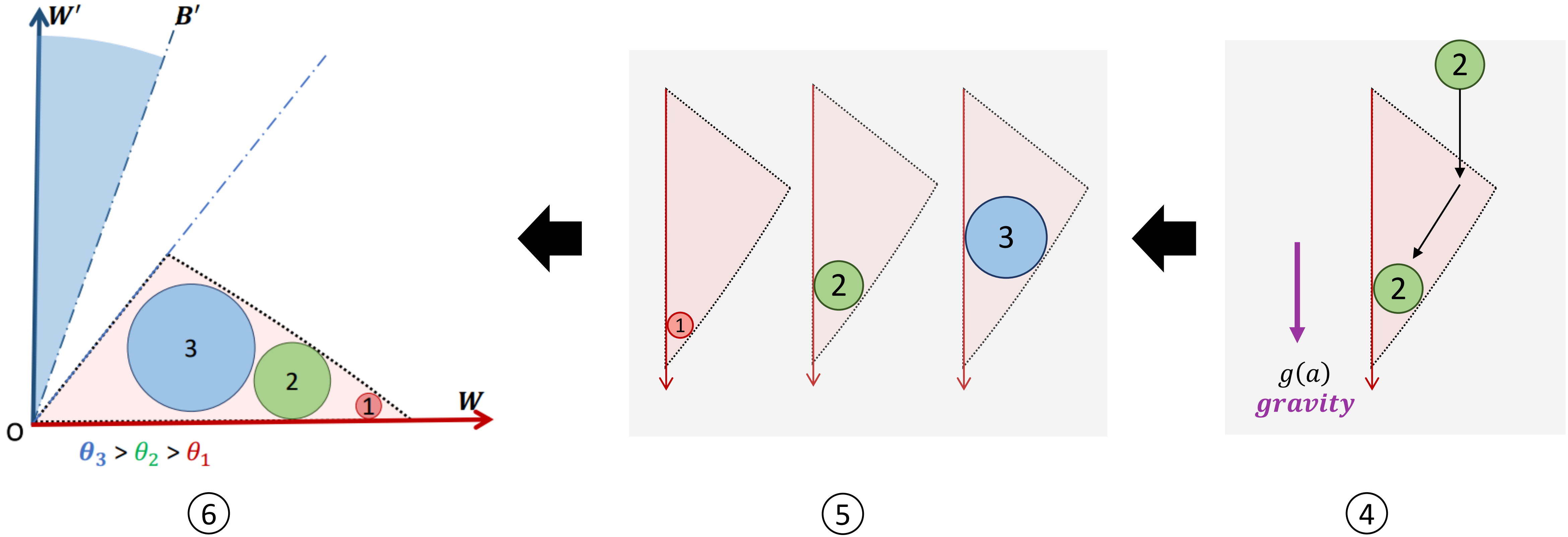


②



③

Methodology — MagFace



Methodology — Proofs

$$L_i = -\log \frac{e^{s\cos(\theta_{y_i}+m(a_i))}}{e^{s\cos(\theta_{y_i}+m(a_i))} + \sum_{j\neq y_i} e^{s\cos\theta_j}} + \lambda_g \cdot g(a_i)$$

In MagFace, $m(a_i), g(a_i), \lambda_g$ are required to have the following constraints:

1. $m(a_i)$ is an increasing convex function in $[l_a, u_a]$ and $m'(a_i) \in (0, K]$, where K is a upper bound;
2. $g(a_i)$ is a strictly convex function with $g'(u_a) = 0$;
3. $\lambda_g \geq \frac{sK}{-g'(l_a)}$.

Methodology — Proofs

$$L_i = -\log \frac{e^{s\cos(\theta_{y_i}+m(a_i))}}{e^{s\cos(\theta_{y_i}+m(a_i))} + \sum_{j\neq y_i} e^{s\cos\theta_j}} + \lambda_g \cdot g(a_i)$$

Property of Convergence. *For* $a_i \in [l_a, u_a]$, L_i *is a strictly convex function which has a unique optimal solution* a_i^*.

Property of Monotonicity. *The* optimal a_i^* is monotonically increasing as hardness of recognition decreases.

Methodology — Proofs

$$L_i = -\log \frac{e^{s\cos(\theta_{y_i}+m(a_i))}}{e^{s\cos(\theta_{y_i}+m(a_i))} + \sum_{j\neq y_i} e^{s\cos\theta_j}} + \lambda_g \cdot g(a_i)$$

Property of Convergence. *For* $a_i \in [l_a, u_a]$, L_i *is a strictly convex function which has a unique optimal solution* a_i^*.

$$\frac{\partial^2 L_i}{(\partial a_i)^2} > 0 \qquad \frac{\partial L_i}{\partial a_i}(u_a) > 0 \qquad \frac{\partial L_i}{\partial a_i}(l_a) \leq 0$$

Property of Monotonicity. *The* optimal a_i^* is monotonically increasing as hardness of recognition decreases.

Methodology — Proofs

$$L_i = -\log \frac{e^{s\cos(\boxed{\theta_{y_i}}+m(a_i))}}{e^{s\cos(\theta_{y_i}+m(a_i))} + \boxed{\sum_{j\neq y_i} e^{s\cos\theta_j}}} + \lambda_g \cdot g(a_i)$$

B

Property of Convergence. *For $a_i \in [l_a, u_a]$, L_i is a strictly convex function which has a unique optimal solution a_i^*.*

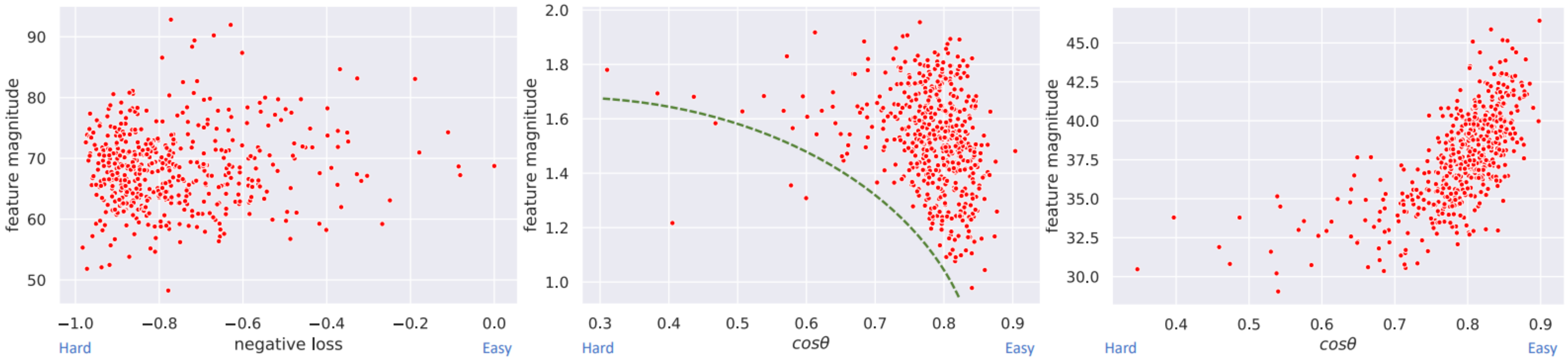
$$\frac{\partial^2 L_i}{(\partial a_i)^2} > 0 \qquad \frac{\partial L_i}{\partial a_i}(u_a) > 0 \qquad \frac{\partial L_i}{\partial a_i}(l_a) \leq 0$$

Property of Monotonicity. *The* optimal a_i^* is monotonically increasing as hardness of recognition decreases.

$$\theta_{y_i}^1 < \theta_{y_i}^2 \implies a_{i,1}^* > a_{i,2}^*. \qquad B_1 < B_2 \implies a_{i,1}^* > a_{i,2}^*.$$

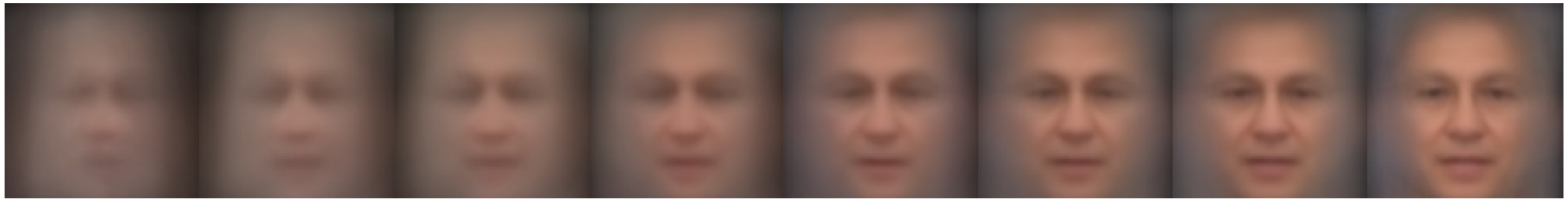
Experiments — Visualizations

Feature distributions:



(a) Softmax (b) ArcFace (c) MagFace

Mean faces on IJB-C:



(a) mean: 22.84 range: $(-\infty, 24)$ # of faces: 3692

(b) mean: 25.13 range: [24, 26) # of faces: 9955

(c) mean: 27.03 range: [26, 28) # of faces: 15459

(d) mean: 29.03 range: [28, 30) # of faces: 17565

(e) mean: 31.01 range: [30, 32) # of faces: 20627

(f) mean: 32.99 range: [32, 34) # of faces: 19743

(g) mean: 34.80 range: [34, 36) # of faces: 11238

(h) mean: 36.55 range: $[36, \infty)$ # of faces: 1721

Experiments — Visualizations

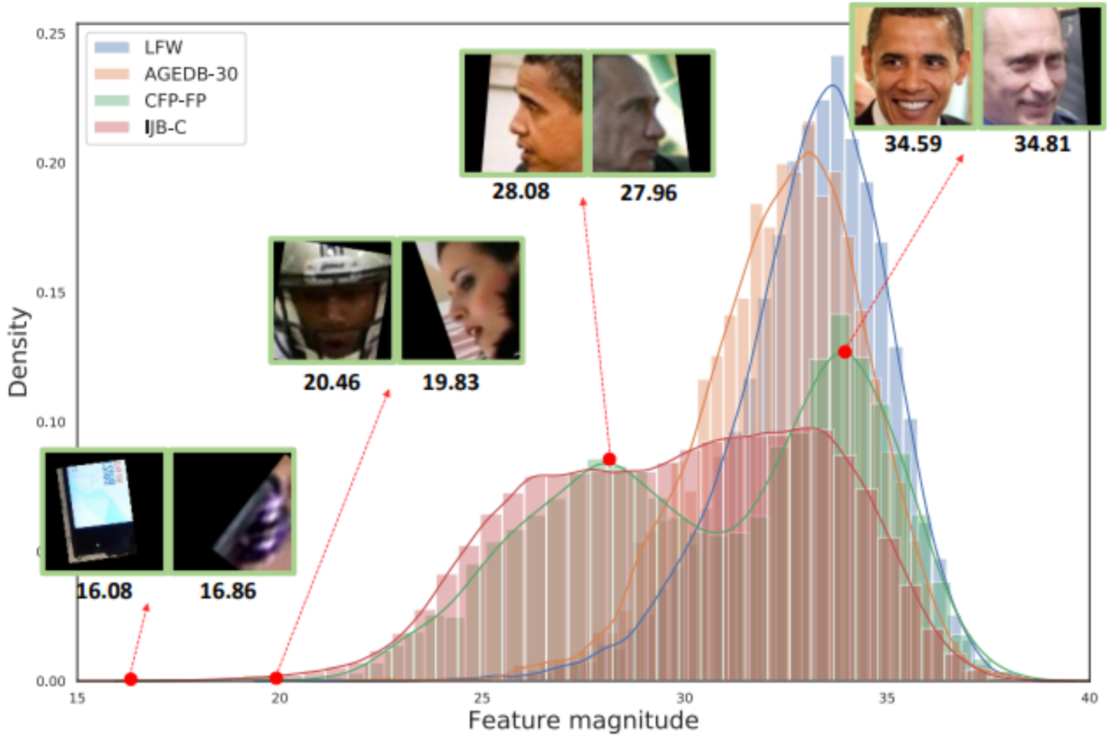


Figure: Distributions of magnitudes.

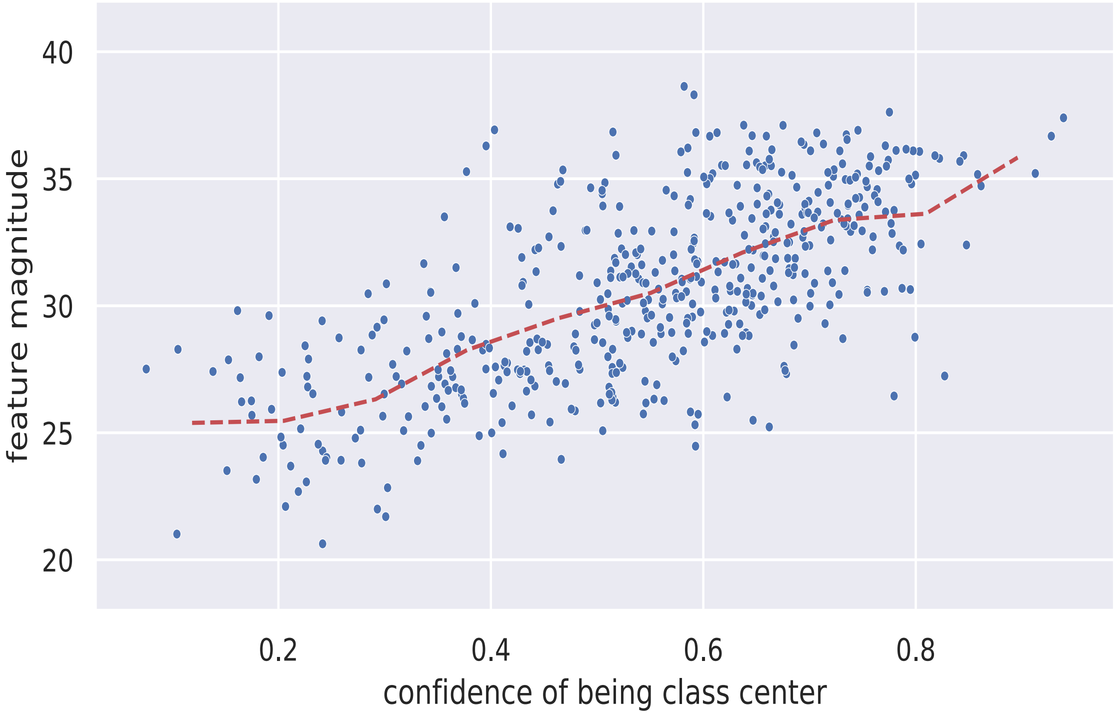


Figure : magnitudes v.s. confidences of being cluster centers.

Experiments — Quantitative Results

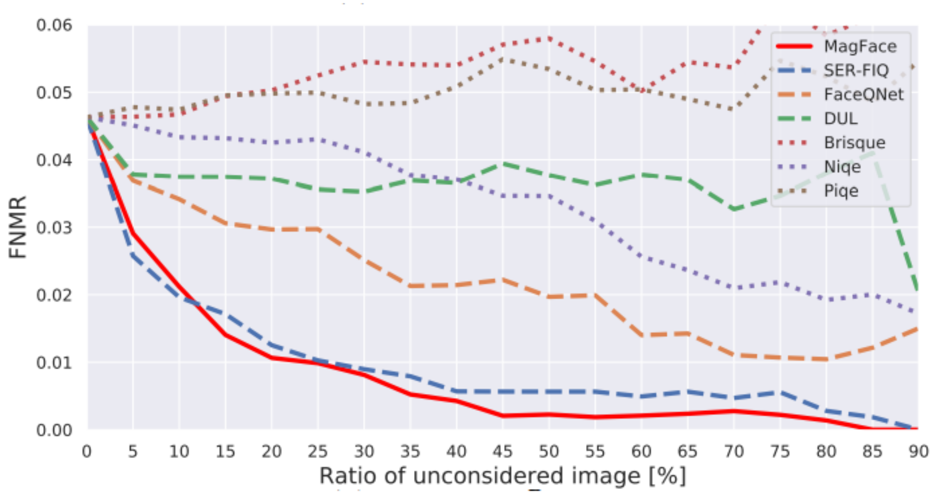
face recognition

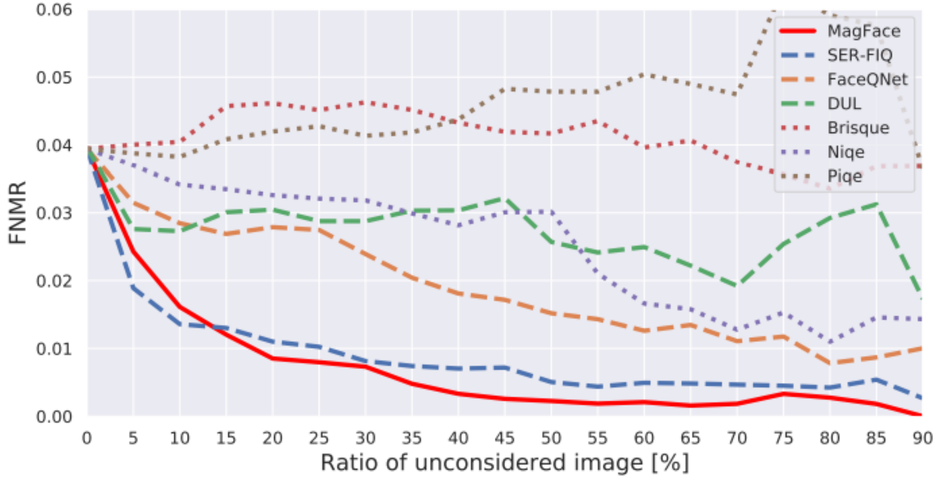
Method	IJB-B (TAR@FAR)			IJB-C (TAR@FAR)		
	1e-6	1e-5	1e-4	1e-6	1e-5	1e-4
VGGFace2*	-	67.10	80.00	-	74.70	84.00
CenterFace*	-	-	-	-	78.10	85.30
CircleLoss*	-	-	-	-	89.60	93.95
ArcFace*	-	-	94.20	-	-	95.60
Softmax	**46.73**	75.17	90.06	64.07	83.68	92.40
SV-AM-Softmax	29.81	69.25	84.79	63.45	80.30	88.34
SphereFace	39.40	73.58	89.19	68.86	83.33	91.77
CosFace	40.41	89.25	94.01	87.96	92.68	95.56
ArcFace	38.68	88.50	94.09	85.65	92.69	95.74
MagFace	40.91	89.88	94.33	89.26	93.67	95.81
MagFace+	42.32	**90.36**	**94.51**	**90.24**	**94.08**	**95.97**

face clustering

Method	Net	IJB-B-512		IJB-B-1024		IJB-B-1845	
		F	NMI	F	NMI	F	NMI
K-means	ArcFace	66.70	88.83	66.82	89.48	66.93	89.88
	MagFace	**66.75**	**88.86**	**67.33**	**89.62**	**67.06**	**89.96**
AHC	ArcFace	69.72	89.61	70.47	90.54	70.66	90.90
	MagFace	**70.24**	**89.99**	**70.68**	**90.67**	**70.98**	**91.06**
DBSCAN	ArcFace	72.72	90.42	72.50	91.15	73.89	91.96
	MagFace	**73.13**	**90.61**	**72.68**	**91.30**	**74.26**	**92.13**
L-GCN	ArcFace	84.92	93.72	83.50	93.78	80.35	92.30
	MagFace	**85.27**	**93.83**	**83.79**	**94.10**	**81.58**	**92.79**

quality assessments





(d) CFP-FP - MagFace

Summary

MagFace

1. a category of losses
2. only requires class labels
3. has rigorous theoretical guarantees
4. magnitudes can represents qualities
5. benefit recognition by balancing easy/semi-hard/hard samples
6. benefit clustering with
   - more reasonable distributions
   - the ability to predict cluster centers

Thank you!

https://github.com/IrvingMeng/MagFace

MagFace

https://irvingmeng.github.io/

Homepage