MagFace: A Universal Representation for Face Recognition and Quality Assessment

Qiang Meng, Shichao Zhao, Zhida Huang, Feng Zhou

Algorithm Research, Aibee Inc.

March 23, 2021

Outline

Introduction
- Motivations
- Goals

Methodology
- Proposed MagFace
- Theoretical Proofs
- Analysis on Feature Magnitude

Experiments
- Face Recognition
- Face Quality Assessment
- Face Cluster

Introduction — Motivations

Quality methods:

Face qualities are affected by

- describable image properties (luminances, distortions, occlusions...);
- describable face properties (poses, expressions, ages, races...);
- indescribable properties (completeness, spontaneousness, fidelity...).

Current quality methods based on artificially or human labelled values

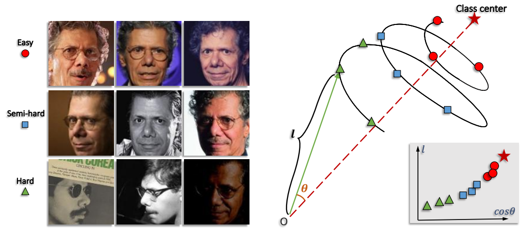
- lack a clear definition of quality.
- ignore the relationships between qualities and face recognition.

Face recognition methods:

- Features are normalized for better performances (*e.g.*, L2-softmax, NormFace, CosFace, ArcFace...);
- Easy and hard samples are treated equally.

In our work, quality is defined as "hardness for recognition" and revealed by feature magnitude.

Introduction — Goals

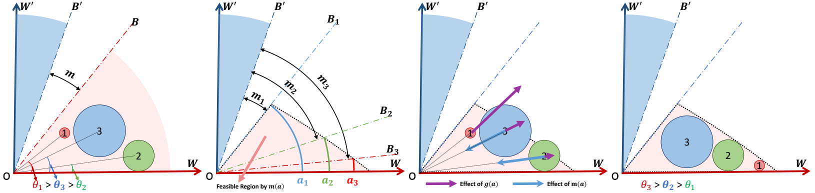


Our goal is to propose a category of losses which

- only requires recognition labels;
- pulls the easier samples closer to the class center;
- pushes the easier samples away from the origin o;
- connects face recognition, quality assessment and face cluster.

feature direction $\rightarrow$ *recognition*
feature magnitude $\rightarrow$ *quality*
feature distribution $\rightarrow$ *cluster*

Methodology — MagFace



(a) (b) (c) (d)

$$L_{Mag} = \frac{1}{N}\sum_{i=1}^{N} L_i, \quad \text{where}$$

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

a_i – feature manigutde for sample i.

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

$m(a_i)$ is an increasing convex function in $[l_a, u_a]$ and $m'(a_i) \in (0, K]$.

$g(a_i)$ – the regularizer.

$g(a_i)$ is a strictly convex function with $g'(u_a) = 0$.

λ_g – weight for $g(a_i)$.

λ_g is greater than or equal to $sK / -g'(l_a)$.

Methodology — Proofs

Property of Convergence

For $a_i \in [l_a, u_a]$, L_i is a strictly convex function which has a unique optima a_i^*.

Assume that f_i is top-k correctly classified and $m(a_i) \in [0, \pi/2]$. If the number of identities n is much larger than k (*i.e.*, $n \gg k$), the probability of $\theta_{y_i} + m(a_i) \in [0, \pi/2]$ approaches 1.

Strictly convexity:

$$\frac{\partial^2 L_i}{(\partial a_i)^2} > 0 \ .$$

Existence of optima:

$$\frac{\partial L_i}{\partial a_i}(u_a) > 0, \ \frac{\partial L_i}{\partial a_i}(l_a) < 0.$$

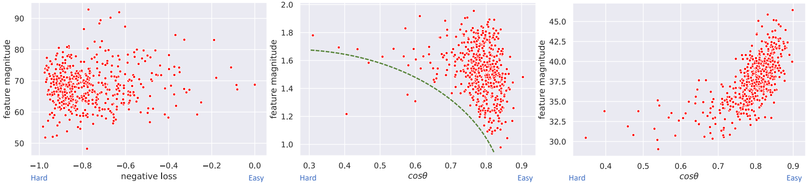
Property of Monotonicity

The optimal a_i^* is monotonically increasing as the cosine-distance to its class center decreases and the cos-distances to other classes increase.

With fixed f_i and $W_j, j \in \{1, \cdots, n\}, j \neq y_i$, the optimal feature magnitude a_i^* is monotonically decreasing if the cosine-distance to its class center W_{y_i} increases.

With other things fixed, the optimal feature magnitude a_i^* is monotonically decreasing with a decreasing inter-class distance B.

Methodology — Analysis on Feature Magnitude



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

(a). Softmax suffers from poor recognition performances and reveals a relative week relationship.
(b). Easy-recognized samples have large variations of magnitudes. A lower bound exists in ArcFace.
(c). In MagFace, feature magnitudes and difficulties for recognition are positively related.

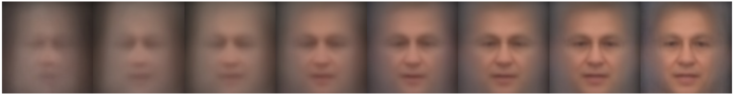
Experiments — Face Recognition

Training dataset: MS1MV2.
BackBone: ResNet100.

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**

Table: Verification accuracy (%) on difficult benchmarks. "*" indicates the result quoted from the original paper.

Experiments — Face Quality Assessment



(a) mean: 22.84 range: (-∞, 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, ∞) # of faces: 1721

Figure: Visualization of the mean faces of 100k images sampled from the IJB-C dataset.

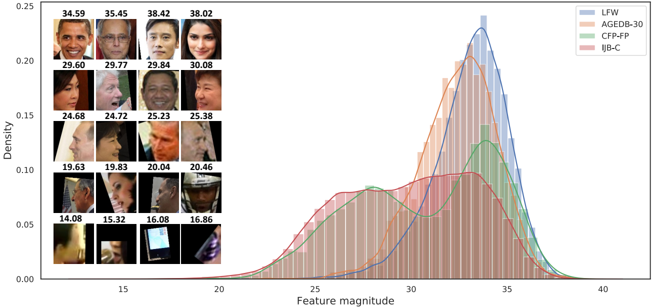
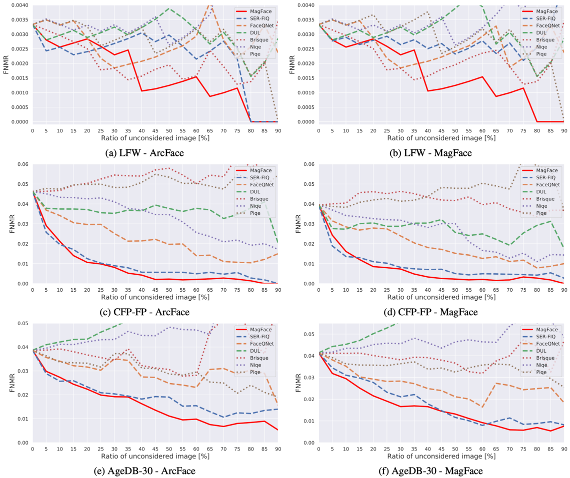


Figure: Distributions of magnitudes on different datasets.

Experiments — Face Quality Assessment



(a) LFW - ArcFace
(b) LFW - MagFace
(c) CFP-FP - ArcFace
(d) CFP-FP - MagFace
(e) AgeDB-30 - ArcFace
(f) AgeDB-30 - MagFace

Figure: Face verification performance for the predicted face quality values with two evaluation models (ArcFace and MagFace).

Experiments — Face Cluster

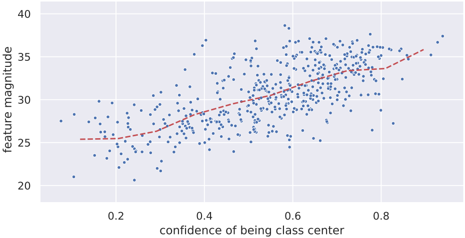


Figure: Visualization of MagFace magnitudes of 500 samples from IJB-B-1845 w.r.t. their confidences of being class centers.

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**

Table: F-score (%) and NMI (%) on clustering benchmarks.

Thank You