

Face Recognition method survey:5

AdaFace: Quality Adaptive Margin for Face Recognition [\[pdf\]](#)

前言：本文的方法通过用feature norms来近似图像质量，这里是以自适应边缘函数的形式来实现这一点。




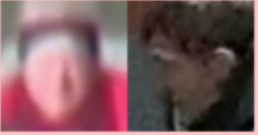
大量的实验表明，AdaFace在4个数据集(IJB-B、IJB-C、IJB-S和TinyFace)上提高了现有的(SOTA)的人脸识别性能

文章的主要贡献大概两点：


- 1) 提出一种loss: AdaFace，进一步对不同的难样本赋予不同的权重，挖掘难样本的同时，避免关注不可识别的image；
- 2) 发现：添加了angular margin，会因sample的不同而产生对signal(梯度)的缩放。

什么是图像质量？

图像质量是一个属性的组合，表明一个图像如何如实地捕获原始场景。影响图像质量的因素包括亮度、对比度、锐度、噪声、色彩一致性、分辨率、色调再现等。当人脸图像质量较低时，根据质量程度的不同，识别任务变得不可行。像下图最下面一行这样的低质量图像正越来越成为人脸识别数据集的重要组成部分，因为它们会在监控视频和无人机镜头中遇到。低质量的人脸图像的一个问题是，它们往往无法辨认。当图像退化过大时，相关的身份信息从图像中消失，导致图像无法识别。这些无法识别的图像对训练过程有害的，因为模型将试图利用图像中的其他视觉特征，如服装颜色或图像分辨率，进而会影响训练损失。如果这些图像在低质量图像的分布中占主导地位，那么该模型在测试期间很可能在低质量的数据集上表现不佳。

Recognizability Image Quality	Easy to Recognize	Hard to Recognize	Impossible to Recognize
High Quality			
Low Quality			

 : Images contain enough clues to identify the subject

 : Images **do not** have enough clues to identify the subject

相关工作

基于Margin的softmax损失函数被广泛应用于人脸识别训练中(FR)。在Softmax损失中加入了Margin，是因为加入Margin后模型可以学习到更好的类间表征和类内表征，特征也就更具有可判别性。典型的形式有：SphereFace、CosFace和ArcFace，他们都可以统一为：

$$\mathcal{L} = -\log \frac{\exp(f(\theta_{y_i}, m))}{\exp(f(\theta_{y_i}, m)) + \sum_{j \neq y_i} \exp(s \cos \theta_j)}$$

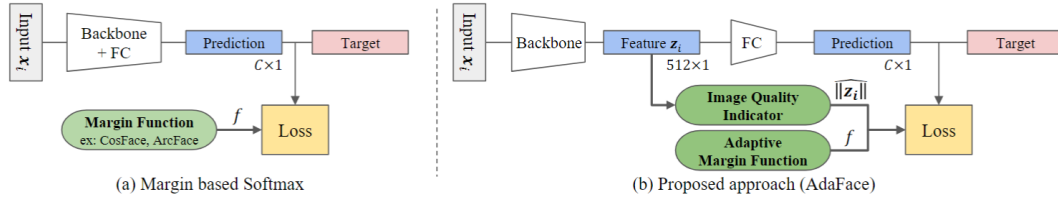
$$f(\theta_j, m)_{\text{SphereFace}} = \begin{cases} s \cos(m\theta_j) & j = y_i \\ s \cos \theta_j & j \neq y_i \end{cases},$$

$$f(\theta_j, m)_{\text{CosFace}} = \begin{cases} s(\cos \theta_j - m) & j = y_i \\ s \cos \theta_j & j \neq y_i \end{cases},$$

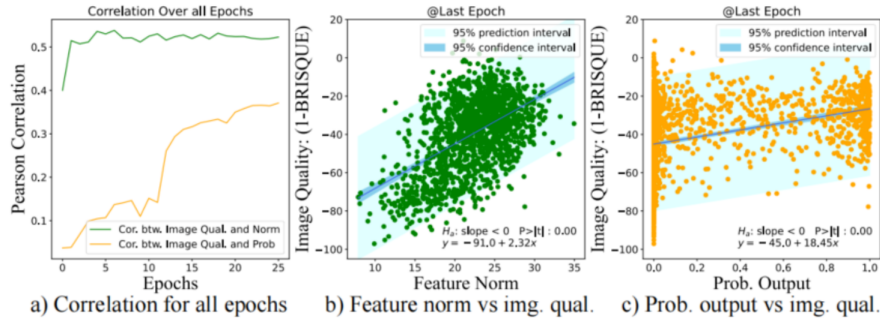
$$f(\theta_j, m)_{\text{ArcFace}} = \begin{cases} s \cos(\theta_j + m) & j = y_i \\ s \cos \theta_j & j \neq y_i \end{cases}.$$

Adaface

为了解决不可识别图像引起的问题，作者提出基于特征范数来适应Margin函数。



在使用基于Margin的Softmax Loss训练的模型中，特征范数表现出与图像质量相关的趋势。



作为特征范数，是一个模型依赖的量，使用batch统计和对其进行归一化。具体来说：

$$\widehat{\|z_i\|} = \left[\frac{\|z_i\| - \mu_z}{\sigma_z/h} \right]_{-1}^1$$

针对不同质量的图像，作者设计了一个Margin函数，主要想实现：如果图像质量高，强调困难样本，如果图像质量低，不强调困难样本。

$$f(\theta_j, m)_{\text{AdaFace}} = \begin{cases} s \cos(\theta_j + g_{\text{angle}}) - g_{\text{add}} & j = y_i \\ s \cos \theta_j & j \neq y_i \end{cases},$$

$$g_{\text{angle}} = -m \cdot \widehat{\|z_i\|}, \quad g_{\text{add}} = m \cdot \widehat{\|z_i\|} + m.$$

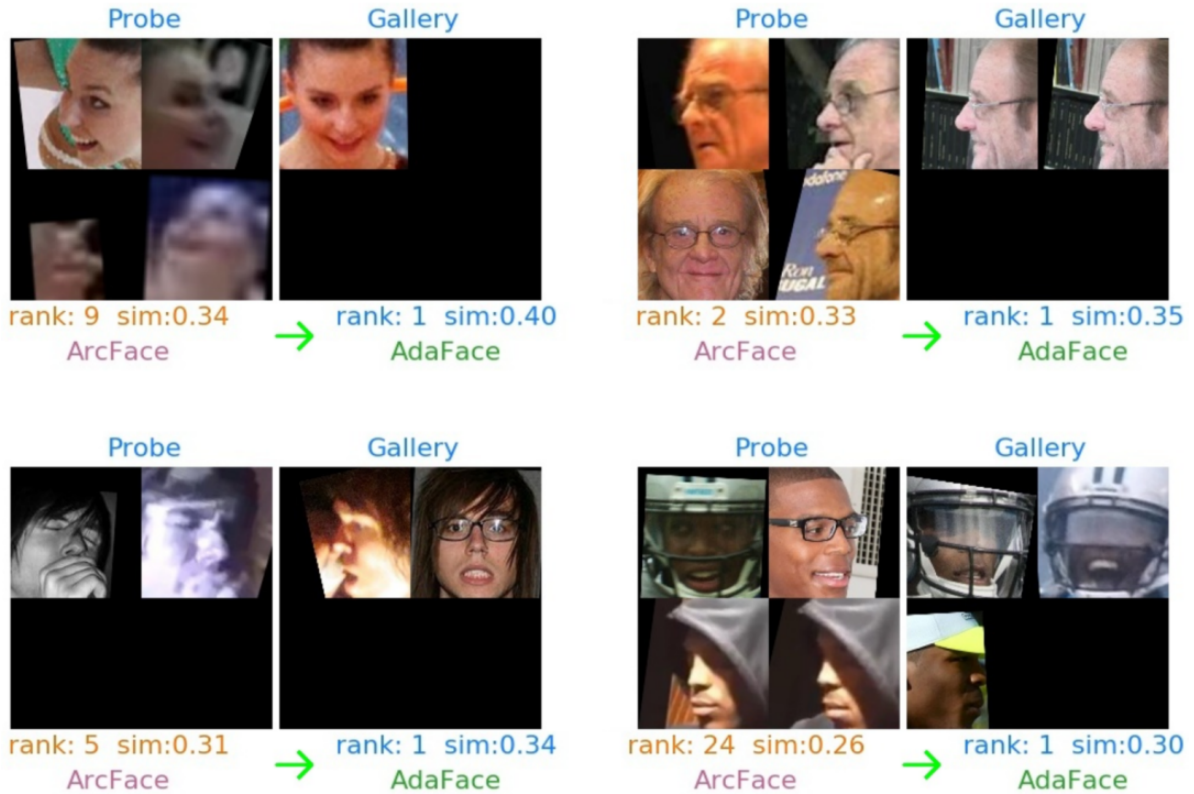
Method	Venue	Train Data	High Quality						Mixed Quality	
			LFW [14]	CFP-FP [31]	CPLFW [47]	AgeDB [29]	CALFW [48]	AVG	IJB-B [41]	IJB-C [26]
CosFace ($m = 0.35$) [39]	CVPR18	MS1MV2	99.81	98.12	92.28	98.11	95.76	96.82	94.80	96.37
ArcFace ($m = 0.50$) [7]	CVPR19	MS1MV2	99.83	98.27	92.08	98.28	95.45	96.78	94.25	96.03
AFRN [18]	ICCV19	MS1MV2	99.85	95.56	93.48	95.35	96.30	96.11	88.50	93.00
MV-Softmax [40]	AAAI20	MS1MV2	99.80	98.28	92.83	97.95	96.10	96.99	93.60	95.20
CurricularFace [16]	CVPR20	MS1MV2	99.80	98.37	93.13	98.32	96.20	97.16	94.80	96.10
URL [34]	CVPR20	MS1MV2	99.78	98.64	-	-	-	-	-	96.60
BroadFace [20]	ECCV20	MS1MV2	99.85	98.63	93.17	98.38	96.20	97.25	94.97	96.38
MagFace [27]	CVPR21	MS1MV2	99.83	98.46	92.87	98.17	96.15	97.10	94.51	95.97
SCF-ArcFace [21]	CVPR21	MS1MV2	99.82	98.40	93.16	98.30	96.12	97.16	94.74	96.09
DAM-CurricularFace [23]	ICCV21	MS1MV2	-	-	-	-	-	-	95.12	96.20
AdaFace ($m = 0.4$)	CVPR22	MS1MV2	99.82	98.49	93.53	98.05	96.08	97.19	95.67	96.89
VPL-ArcFace [8]	CVPR21	MS1MV3	99.83	99.11	93.45	98.60	96.12	97.42	95.56	96.76
AdaFace ($m = 0.4$)	CVPR22	MS1MV3	99.83	99.03	93.93	98.17	96.02	97.40	95.84	97.09
ArcFace* [7]	CVPR19	WebFace4M	99.83	99.19	94.35	97.95	96.00	97.46	95.75	97.16
AdaFace ($m = 0.4$)	CVPR22	WebFace4M	99.80	99.17	94.63	97.90	96.05	97.51	96.03	97.39

(a) A performance comparison of recent methods on high and mixed quality datasets.

Method	Train Data	Low Quality (IJB-S [17] and TinyFace [6])										
		Surveillance-to-Single [17]			Surveillance-to-Booking [17]			Surveillance-to-Surveillance [17]			TinyFace [6]	
		Rank-1	Rank-5	1%	Rank-1	Rank-5	1%	Rank-1	Rank-5	1%	Rank-1	Rank-5
PFE [33]	MS1MV2 [7]	50.16	58.33	31.88	53.60	61.75	35.99	9.20	20.82	0.84	-	-
ArcFace [7]	MS1MV2 [7]	57.35	64.42	41.85	57.36	64.95	41.23	-	-	-	-	-
URL [34]	MS1MV2 [7]	59.79	65.78	41.06	61.98	67.12	42.73	-	-	-	63.89	68.67
CurricularFace* [16]	MS1MV2 [7]	62.43	68.68	47.68	63.81	69.74	47.57	19.54	32.80	2.53	63.68	67.65
AdaFace ($m = 0.4$)	MS1MV2 [7]	65.26	70.53	51.66	66.27	71.61	50.87	23.74	37.47	2.50	68.21	71.54
AdaFace ($m = 0.4$)	MS1MV3 [9]	67.12	72.67	53.67	67.83	72.88	52.03	26.23	40.60	3.28	67.81	70.98
ArcFace* [7]	WebFace4M [49]	69.26	74.31	57.06	70.31	75.15	56.89	32.13	46.67	5.32	71.11	74.38
AdaFace ($m = 0.4$)	WebFace4M [49]	70.42	75.29	58.27	70.93	76.11	58.02	35.05	48.22	4.96	72.02	74.52

(b) A performance comparison of recent methods on low quality datasets.

Table 3. Comparison on benchmark datasets, with the ResNet100 backbone. For high quality and mixed quality datasets, 1:1 verification accuracy and TAR@FAR=0.01% are reported respectively. For IJB-S, open-set TPIR@FPIR=1% and closed-set rank retrieval (Rank-1 and Rank-5) are reported. Rank retrieval is also used for TinyFace. [KEYS: **Best**, **Second best**, *=our evaluation of the released model]



总结

实际上近几年所有的科学发展都逐渐在走向自适应，比如我之前在做的通信智能反射面的工作，就是根据入射信号自适应调整反射信号的相角，进而改变无线信道的通信增益，而且这种自适应的调整是最优化的，人工智能发展到后面也需要讨论针对所有样本做到自适应、端对端，就像目前图像去噪只能做到针对特定种类噪声的去除，并没有实现对所有噪声的大一统，所以说白了图像恢复目前只是做到了比传统方法要好，但是没做到更高级的真正的自适应。本文最大的创新就是基于样本本身的质量去构成损失函数、分配权重，一般如何去量化这种本身评判标准就有点模糊的特征是很难的，作者引入特征范数来代表图像质量避免强调难以识别的图像，专注于困难但可识别的样本。这项工作解决了训练数据中存在的无法识别的图像。然而，噪声标签也是大规模人脸训练数据集的突出特征之一。AdaFace损失函数

对贴错标签的样品没有特殊处理。由于自适应损失赋予高质量的困难样本很大的重要性，高质量的错误标记图像可能会被错误地强调。未来可以同时适应不可识别性和标签噪声。