一种增强人脸识别模型训练稳定性的损失函数

周书田，颜 信，谢镇汕*

(电子科技大学格拉斯哥学院 成都 611731)

【摘要】随着卷积神经网路的快速发展，深度学习在人脸识别领域进行了大量的应用。近几年，人脸识别准确率快速提高，主要归功于新颖损失函数的提出。在目前最大的人脸评测数据集 MegaFace 上，最顶尖的模型已经实现了 97.91% 的 1∶*N* 查找性能，但是训练过程中收敛稳定性问题没有得到解决。该文提出一种新型的损失函数 LineFace，其 logit 曲线在余弦空间中呈线性，使训练中梯度收敛更加稳定。大量实验表明，该损失函数可以实现良好的模型性能收敛性与识别性能。

关 键 词 卷积神经网络； 人脸识别； 损失函数； 模型收敛性

中图分类号 TP391.4 **文献标志码** A doi:10.12178/1001-0548.2020226

New Loss Function to Enhance the Training Stability of Face Recognition Model

ZHOU Shu-tian, YAN Xin, and XIE Zhen-shan*

(Glasgow Collge, University of Electronic Science and Technology of China Chengdu 611731)

Abstract With the rapid development of convolutional neural networks, deep learning has been widely used in the field of face recognition. In recent years, the accuracy of face recognition has increased rapidly, mainly due to the proposition of novel loss functions. On the current largest face test set MegaFace, the top model has achieved 97.91% 1∶*N* search performance, but the problem of convergence stability during training has not been properly solved. Thus a new type of loss function, Lineface, is proposed in this paper. Its logic curve is linear in the cosine space, which makes the gradient convergence better and more stable during training. A large number of experiments show that good model performance and convergence can be achieved.

Key words convolutional neural network; face recognition; loss function; model convergence

深度学习在人脸识别领域取得了巨大的成就，已经广泛应用在金融认证、门禁控制等领域，创造了巨大的经济与社会价值。在目前最大的公开人脸评测数据集 MegaFace[1] 上，顶尖的方法已经实现了 97.91% 的 1:*N* Top-1 查准率[2]，并且已经超过人眼识别的水平。典型的人脸识别流程包含人脸检测[3]、人脸剪裁[4]、人脸特征提取[5]、人脸特征比对与查找[6]4 个流程。人脸识别模型优化的本质为：在一个特征空间中，属于同一个人的图像特征尽可能聚拢，不属于同一个人的图像特征尽可能发散，模型对于特征建模能力的强弱直接影响人脸识别模型的识别准确率。目前，人脸识别系统的巨大进步主要来自于网络架构[7] 的不断革新，以及损失函数的设计。设计优良的损失函数，可以最大化实现类内特征的聚合与类间特征的离散。最近，大量的损失函数，如 Triplet Loss[8]，使负样本对的距离大于正样本对的距离，大大提升了网络对于特异性特征向量的建模能力。

人脸识别模型依托于大规模的训练，训练图片数量高达数百万张，训练过程经常持续数周。但现有人脸识别训练模型常常会带来收敛过程中的震荡，使训练成本加大。本文对现有的人脸识别损失函数 ArcFace[2] 进行改进，将其 logit 曲线变换成为一条直线，使收敛过程更加稳定，本文的方法命名为 LineFace。

1 模型与方法

1.1 预置知识

人脸识别问题是一个典型的分类问题。解决传统分类问题的方法为使用交叉熵-softmax 损失函数

收稿日期：2020－05－18；修回日期：2020－07－06
作者简介：周书田 (1998－)，男，主要从事计算机视觉方面的研究.
通信作者：谢镇汕，E-mail：ivanxie1022@gmail.com

进行分类损失监督，然后将损失进行反向传递，来更新识别网络参数。对于一张人脸照片，首先通过骨干网络提取其特征向量，将其表示为$\boldsymbol{x}_i \in R^d$，并且属于类别y_i，对于类i的分类概率为：

$$P_i = \frac{e^{\boldsymbol{W}_{y_i}^{\mathrm{T}}\boldsymbol{x}_i + b_{y_i}}}{\sum_{j=1}^{n} e^{\boldsymbol{W}_j^{\mathrm{T}}\boldsymbol{x}_i + b_j}} \tag{1}$$

结合交叉熵损失函数变为：

$$L = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{\boldsymbol{W}_{y_i}^{\mathrm{T}}\boldsymbol{x}_i + b_{y_i}}}{\sum_{j=1}^{n} e^{\boldsymbol{W}_j^{\mathrm{T}}\boldsymbol{x}_i + b_j}} \tag{2}$$

式中，N为训练集批样本当中训练图片样本的数量；$\boldsymbol{W}_j$为全连接层的参数，$\boldsymbol{W}_j^{\mathrm{T}}$表示其转置；$b_j$表示偏置项，在绝大多数的条件下，偏置项为0。受文献[9]的启发，本文将参数和特征都进行归一化，经过归一化后的损失函数为：

$$L = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{s\cos\theta_{y_i}}}{e^{s\cos\theta_{y_i}} + \sum_{j=1,j\neq y_i}^{n} e^{s\cos\theta_j}} \tag{3}$$

式中，s表示归一尺度；$\cos\theta_{y_i}$表示归一化后特征向量与类向量的夹角余弦值。经过归一化后，该表达式的几何学意义变得明确，并且分类层的矩阵乘法变成了特征向量相似度的计算。归一化后的softmax损失函数，可以使训练的特征初步具有可分性，但其特征的特异性，即最小化类内距离与最大化类间距离，并没有被显示的约束加强。根据式(3)，当损失函数被优化时，正样本的logit仅被要求比负样本的logit大，但没有要求大很多，这使得在分类问题中，样本具备可分性。但在开集测试中，测试样本与训练样本在类别上并无交集，因此仅可分的特征限制了人脸识别模型的能力。近来，大量的损失函数被提出[2, 8-11]，其出发点均为使类内样本更加聚合，类间样本更加分散。Margin类的损失函数[2, 8-11]通过在正logit项上添加Margin约束，即使正logit大于负logit一个阈值，也保证了类内的聚类与类间的分散。Margin类的损失函数可以被统一表达为：

$$L = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{s(\cos(m_1\theta_{y_i}+m_3)-m_2)}}{e^{s(\cos(m_1\theta_{y_i}+m_3)-m_2)} + \sum_{j=1,j\neq y_i}^{n} e^{s\cos\theta_j}}$$

式中，m_1、m_2、m_3为添加的不同的Margin。如图1所示，所有的Margin类损失函数本质上都是改变logit曲线的形状。但是对于所有损失函数的不同的角度值，其梯度变化剧烈，这带来训练过程中收敛的困难。当夹角过小与过大时(分别处于训练的初始与结束阶段)，梯度缓慢，收敛缓慢，训练中途，梯度变大，造成训练中的参数更新不稳定。

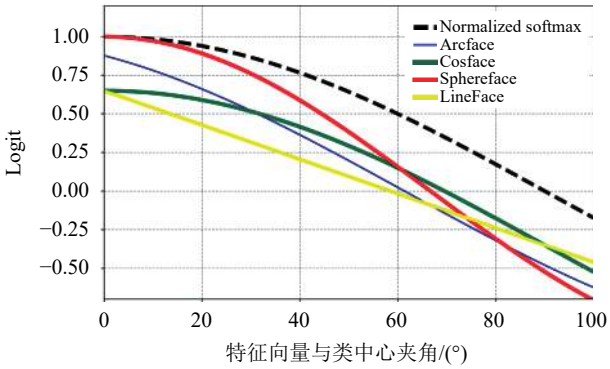

图1 不同损失函数logit曲线的对比

1.2 方 法

为了实现稳定的训练过程与极具特异性的特征表达，将logit曲线变更为一条直线，在不同的收敛过程与阶段，其梯度始终为定值，避免了梯度的跳变，这带来了稳定的收敛。该损失函数命名为LineFace，其表达式为：

$$L = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{s\cos\left(1-\frac{2}{\pi}\theta_{y_i}-m\right)}}{e^{s\cos\theta_{y_i}} + \sum_{j=1,j\neq y_i}^{n} e^{s\cos\theta_j}} \tag{4}$$

式中，

$$\theta_i = \arccos(\boldsymbol{W}_i^{\mathrm{T}}\boldsymbol{x}_i / s) \in [0,\pi] \tag{5}$$

LineFace在区间内单调递减，并且梯度恒定，拥有更好的可解释性。而对于ArcFace[2]和CosFace[10]，其在区间内，并不单调递减，并且梯度变化剧烈，使得训练不稳定。大量的实验表明，本文的损失函数可以大幅提高收敛过程的稳定性质，并带来识别性能的提高。LineFace在训练过程中改变监督函数，不会在推理阶段引入新的计算消耗，因此在提升训练稳定的同时，保持了与之前工作相同的计算复杂性。

2 实验设置

2.1 实验数据

近年来，大规模数据集的出现极大地推进了人脸识别模型性能的提升，本文采用了学术界广泛使用的MS-Celeb-1M[12]数据集作为训练数据集，也

是迄今为止开源的最大规模的人脸训练数据集。原始 MS-Celeb-1M[12] 数据的数据集被证明具有大量的数据噪声，因此与之前的工作相似，对 MS-Celeb-1M[12] 数据进行了清理，清理后的数据集包含来自 9 万人的 510 万张照片。对于测试过程，本文选取了常见的 LFW[13]、MegaFace[1]、YTF[14] 以及 IJB-A[15] 数据集进行了测试。

2.2　数据预处理与骨干模型

人脸识别的第一步为人脸检测与裁剪，本文使用 MTCNN[4] 检测人脸与关键点，人脸图像将会被仿射变换到 112*112，然后将像素值归一化到 [0, 1]，并且在最后添加了随机翻转以进行数据增广。

对于骨干模型，为了与之前的工作进行公平比较，采用了 ResNet-101 的网络结构，ResNet-101 是经典的深度学习骨干模型，其良好的泛化性已经在众多的任务中得到了验证。将特征嵌入层的维度设置为 512，与之前的工作保持一致。

2.3　训练设置

本文在 8 块 NVIDIA 1080TI GPUs 进行了实验，总训练步数为 11 万步，使用了步进学习率与权重衰减，初始学习率被设置为 0.1，并且每 4 万轮衰减 0.1 直到模型收敛，模型动量被设置为 0.9。

3　实验结果

3.1　在 LFW 上的结果

作为人脸识别领域的黄金标准，LFW[13] 测试集被广泛使用。根据报告，人眼在该测试集上的性能为 97.25%，它包含来自于 5 749 人的 13 233 张图片，所有的照片都在非受限场景下采集，包括极具变化的姿势与分辨率。

官方的测试流程包含了 6 000 图像对，其中 3 000 张为正样本对，3 000 张为负样本对。本文严格参照了官方的 10 折交叉验证的测试流程。

表 1　不同损失函数在 LFW 上的性能比较

方法	性能/%
DeepFace	97.35
FaceNet	99.65
DeepID	98.70
SoftMax	99.47
ArcFace	99.69
本文	99.63

如表 1 所示，本文的方法在 LFW 数据集上取得了有竞争力的结果，为 99.63%。LFW 的性能已经在近年趋于饱和，并且因为错标注的存在，LFW 数据集的理论上限为 99.85%。因此，本文在更严苛的数据集 YTF 与 MegaFace 上进行了测试。

3.2　在 YTF 上的结果

YTF[14] 数据集包含来自 1 595 人的 3 424 个视频，平均每个人 2.15 个视频，是现在广泛采用的视频人脸识别数据集。视频的长度从 48～6 070 帧不等，平均为 181.3 帧。视频人脸识别旨在测试模型在抖动模糊等极端场景下的建模能力。并且，在实际的验证场景中，更多的应用场景为视频数据。对于视频中的每个帧，都将其提取特征，并且使用平均池化来汇聚各个帧的信息。

表 2 是在 YTF 上识别的结果，可以看到，本文的损失函数实现了性能的提高，超过了目前最好的模型 ArcFace[2]，并且大幅度领先了对帧之间汇聚进行精心设计的方法 NAN[16]，实验结果证明了本文损失函数的有效性。

表 2　不同损失函数在 YTF 上的性能比较

方法	性能/%
DeepFace	91.40
FaceNet	95.12
DeepID	93.29
SoftMax	90.11
ArcFace	98.01
本文	98.21

3.3　在 MegaFace 上的结果

MegaFace[1] 被认为是目前最具有挑战性的人脸测试集，它由两个现存的数据集 Facescrub 和 FGNet 作为查询集，并且从互联网上收集了百万级别的干扰集。这是第一个在百万级别进行极限人脸识别测试的数据集。

表 3 是模型在 MegaFace 上识别的性能，从表中可以看到，本文在 MegaFace 上也取得了具有竞争力的结果，达到了 98.03%。因为 MegaFace 人脸测试集被证明包含大量的测试噪声，因此本文采用了与之前工作 ArcFace[2] 相同的数据清理策略。

表 3　不同损失函数在 MegaFace 上的性能比较

方法	性能/%
Tripplet	64.79
Center Loss	65.49
FaceNet	70.49
SoftMax	54.85
ArcFace	97.91
本文	98.03

3.4 模型收敛性的验证

通过对损失函数 logit 曲线的改变，将损失函数变成了余弦空间中的一条直线，从而提供了稳定的梯度，本文对训练过程进行了测试，如表 4 所示，本文的模型收敛更快，在模型训练早期的性能大幅领先目前的方法。表 4 中，在训练的迭代次数为 30 000 时，本文方法领先主流方法 6%～13%，证明了本文方法易于收敛。

表 4 不同方法与不同训练迭代数量在 LFW 数据集上的性能

方法	迭代数量		
	30 000	60 000	90 000
CosFace/%	78.20	90.12	98.01
ArcFace/%	85.22	93.82	99.41
本文/%	91.93	97.20	99.49

4 结 束 语

本文对目前现有的人脸损失函数进行了回顾，并且注意到模型收敛中梯度变化剧烈的问题。通过对 logit 曲线进行改变，将其变为一条直线，从而提供了在网络训练的各个阶段都稳定的梯度。大量的实验证明，本文的方法增强了模型的收敛性，并且在各个数据集上都实现了稳定的性能。

参 考 文 献

[1] KEMELMACHER-SHLIZERMAN I, SEITZ S M, MILLER D, et al. The MegaFace benchmark: 1 million faces for recognition at scale[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, Nevada, USA: IEEE, 2016: 4873-4882.

[2] DENG J, GUO J, XUE N, et al. Arcface: Additive angular margin loss for deep face recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA: IEEE, 2019: 4690-4699.

[3] ZHANG K, ZHANG Z, LI Z, et al. Joint face detection and alignment using multitask cascaded convolutional networks[J]. IEEE Signal Processing Letters, 2016, 23(10): 1499-1503.

[4] ZHOU E, CAO Z, SUN J. Gridface: Face rectification via learning local homography transformations[C]//Proceedings of the European Conference on Computer Vision (ECCV). Munich, Germany: ECAI, 2018: 3-19.

[5] SUN Y, WANG X, TANG X. Deep learning face representation from predicting 10 000 classes[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Columbus, Ohio, USA: IEEE, 2014: 1891-1898.

[6] CHEN W, CHEN J, ZOU F, et al. Vector and line quantization for billion-scale similarity search on GPUs[J]. Future Generation Computer System, 2019, 99: 295-307.

[7] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, Nevada, USA: IEEE, 2016: 770-778.

[8] SCHROFF F, KALENICHENKO D, PHILBIN J. Facenet: A unified embedding for face recognition and clustering[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston, Massachusetts, USA: IEEE, 2015: 815-823.

[9] WANG F, XIANG X, CHENG J, et al. Normface: L2 hypersphere embedding for face verification[C]// Proceedings of the 25th ACM International Conference on Multimedia. Mountain View, CA, USA: ACM, 2017: 1041-1049.

[10] WANG H, WANG Y, ZHOU Z, et al. Cosface: Large margin cosine loss for deep face recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, Utah, USA: IEEE, 2018: 5265-5274.

[11] LIU W, WEN Y, YU Z, et al. Sphereface: Deep hypersphere embedding for face recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, Hawaii, USA: IEEE, 2017: 212-220.

[12] GUO Y, ZHANG L, HU Y, et al. Ms-celeb-1m: A dataset and benchmark for large-scale face recognition[C]// European Conference on Computer Vision. Cham: Springer, 2016: 87-102.

[13] LEARNED-MILLER E, HUANG G B, ROYCHOWDHURY A, et al. Labeled faces in the wild: A survey[M]//Advances in Face Detection and Facial Image Analysis. Cham: Springer, 2016: 189-248.

[14] WOLF L, HASSNER T, MAOZ I. Face recognition in unconstrained videos with matched background similarity[C]//CVPR 2011. Colorado, USA: IEEE, 2011: 529-534.

[15] KLARE B F, KLEIN B, TABORSKY E, et al. Pushing the frontiers of unconstrained face detection and recognition: Iarpa janus benchmark a[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston, Massachusetts, USA: IEEE, 2015: 1931-1939.

[16] YANG J, REN P, ZHANG D, et al. Neural aggregation network for video face recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, Hawaii, USA: IEEE, 2017: 4362-4371.

编 辑 漆 蓉