

基于 ArcFace 损失函数的人脸识别校园门禁

胡登峰 李丹

四川大学锦城学院 四川 成都 610065

【摘要】在当今的社会中,人们已经越来越关注和重视校园安全问题。其中校园门禁系统能很好保障校园安全。本文针对目前的校园门禁系统传统形式,提出基于 ArcFace 损失函数的人脸识别校园门禁系统,并介绍、分析、比较不同的人脸识别中的损失函数,在 CASIA-FaceV5 数据集上达到了 97.1% 准确率,并在 MXNet 深度学习框架下完成了使用摄像头的人脸识别校园门禁基本原型,验证了本文提出的基于 ArcFace 损失函数的人脸识别校园门禁的高准确率和有效性。

【关键词】 ArcFace; 损失函数; 人脸识别; 校园安全; 门禁系统; 特征提取

1 引言

由于现代科技不断的发展,人脸识别技术的改进与更新完善,再加之 5G 时代的到来与发展,网络水平和计算机硬件的提高,人脸面部识别的速度也越来越快,实现的难度与花费也降低了很多。

深度学习发展迅速,目前在计算机视觉方向上已经取得了惊人的成就。在文字识别、手写识别、和人脸识别等多种领域也得到了各界关注获得了许多进步。目前人脸识别已经用于了诸多场景,小到手机解锁,大到支付宝刷脸支付和高铁人脸验票。处处可见的人脸识别足以说明人脸识别是安全可靠的。但是人脸识别用于学校和教育领域的却十分稀少。

本文对比了目前比较常用的人脸识别算法,根据不同损失函数分类为: Softmax、SphereFace。以及 CosFace。还有 ArcFace(又名: InsightFace)。探究在多种不一样的网络结构,多种不同类型的损失函数在不同的数据集下识别的准确情况。并在亚洲人脸 CASIA-FaceV5 数据集上准确率达到 97.1%。最终设计并使用一种基于 ArcFace 的人脸识别校园门禁系统。

2 人脸识别过程及基本原理

2.1 人脸识别过程

目前的人脸识别系统的一般流程可以主要划分为 4 个部分,分别是: 1. 首先是确定人脸的区域(其中主要会使用到 MTCNN,多任务卷积神经网络) 2. 然后是人脸识别的图像进行预处理 3. 再对处理了完成了的人脸图像进行特征的提炼提取 4. 最后就可以进行人脸识别并与人脸数据库匹配。

确定人脸的区域首先是需要使用 MTCNN,通过使

用该网络来确定被识别的图片或是视频中的人脸区域。然后再进行取样和裁切,然后通过人脸图像预处理,我们可以把样本调整成同样大小方便输入通过把样本裁切、旋转、拉伸等操作,之后再提炼和提取样本中的人脸图像的特征,可以根据需求的不同使用各种不一样的网络结构与各种不同的损失函数等组合,最终就可以提取到不同特征,这是人脸识别算法的核心,识别效果好坏主要由此决定,也是人脸识别可以改进的关键,最后就可以进行人脸识别并与人脸数据库匹配了。本文主要对比了不同的网络结构以及损失函数的识别准确率。之后本文会介绍不同的网络结构还有不同的损失函数的差异及原理。

2.2 网络结构

现如今在人脸识别的领域中使用较多的网络结构有: AlexNet、VGG 网络、Inception 网络、ResNet 网络(又名: 残差网络)。

在这些网络中,跟其他网络相比较, AlexNet 是最早诞生的,在当时 GPU 处理速度还比较慢时就有了许多的优点优势,其一是 AlexNet 巧妙的结合了 Relu 激活函数,有效地改善了一般当网络深度达到较深时候所产生的一系列梯度弥散问题,而且 AlexNet 加速了训练的过程,尽管在很早之前 Relu 激活函数就存在了,但它其实是由于 AlexNet 然后才被广泛使用的。其二是 AlexNet 其中加入了 Dropout 层,而使这种网络结构能更好的避免过拟合。

VGG 网络结构优点是十分简洁明了,在网络中采用的卷积核全都是一样大小(3×3)。在最大池化层中也全都是统一的大小(2×2),其中多个小滤波器组成的卷积层比大滤波器组成的卷积层拥有更加优秀的效果,

证明了可以通过不停的加深网络层数来提高性能，但是其中的不足是 VGG 网络在使用时会占用更多的算力资源，而且由于更加多的参数被使用，所以会导致存储资源占用更多。

Inception 网络结构有许多的优点，首先就是卷积层采用了不同的大小，所以增加了网络的多样性，其二是末尾最后的全连接层被舍弃去除，改进方法是末尾使用全局均值池化层，其三是有着良好分类的性能下同时控制和减少了参数和计算量，较好的一点是残差思想在 Inception V4 ResNet 网络也有很好的体现。

在人脸识别领域中 ResNet 网络是最被广泛应用的网络，因为其残差的结构可以使网络再更深的同时不会出现梯度消失或者梯度爆炸现象，为探索和研究更深层次的网络提供了良好的保障和有力的方向。

由于不同的网络结构对精度和速度有着不同的效果，所以具体使用哪种网络结构需要看实际应用。本文我们选用了不同参数的 ResNet 网络进行对比实验。

2.3 损失函数

2.3.1 传统损失函数与人脸识别损失函数的不同

目前人脸识别的损失函数与普通的分类损失函数相比有一定的不同。其主要原因是因为传统的 Softmax 分类是一个闭集分类问题，也就是训练数据和测试数据中每一个不同的类别都是有自己对应的标签标记的，不包含未知的类别。而人脸识别领域主要是一个开集分类问题，除了已有类别还包含了不同未知的类别。举个例子：我们以传统的字符分类为例，如果我们需要识别一串字符中的 ABC 并进行分类，其中数据只含 A, B, C 三个的字母类别，测试我们也是只测试 A, B, C 三个的字母类别，不含其他的未知类别，即使有其他未知类别也会被错误的分类到 A, B, C 三类中的一种，这就是闭集分类问题。而开集分类问题就不是只包含 A, B, C 三个的字母类别的分类，其中可能还包含其他类别, D, E, F 或者是其他数字或者字符，但是由于其他未知类别是没有具体标签的，所以分类器需要把这些类别统一都拒绝并分类为其他未知类别。所以对于传统的 Softmax 分类中，我们目标是要在不同的类别间画出一个决策边界来对不同的类别进行区分和分类。但是在人脸识别领域不同，因为人脸识别领域大多是开集分类问题，并不是所有的类都是在我们已知的集合内的，所以我们需要用到 margin（余度）来更好的归纳和总结以实现完美的约束。所以综上：最大的类内距离的重要性 < 最大的类间距离的重要性。

2.3.2 Softmax 损失函数

Softmax Loss 其主要是运用了 Softmax 激活函数再加上交叉熵（Cross Entropy）函数来计算损失。其公式如

下（1）：

$$L_1 = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{W_{y_i}^T x_i + b_{y_i}}}{\sum_{j=1}^n e^{W_j^T x_i + b_j}}, \quad (1)$$

其中属于类的，第 i 个样本的深度特征为 $x_i \in \mathfrak{R}^d$ 。

在本文中，将特征的维数 d 大小设置为 512。其中的 $W_j \in \mathfrak{R}^d$ 中 j 表示第 j 列，而它的权重为 $W \in \mathfrak{R}^{d \times n}$ ，其中 $b_j \in \mathfrak{R}^n$ 是偏置项。N 则是表示批次大小，n 则为类别的数量。

在人脸识别的领域中 Softmax Loss 常常被使用到，但是由于其自身局限性，它在类间样本中没有增强其多样性，在类内样本中也没有增强其相似性，这就导致它在正样本中无法取得特别高的相似度，在负样本中无法取得特别低的相似度，最终导致分类效果不理想。

2.3.3 SphereFace 损失函数

SphereFace 损失函数。也是 A-Softmax。它是对前代 Softmax Loss 的更新优化而由来产生的，首先是将 Softmax Loss 把其中 $e^{W_{y_i}^T x_i + b_{y_i}}$ 变化为 $e^{\|W_{y_i}\| \|x_i\| \psi(\theta_{y_i})}$ 其中 $\psi(\theta_{y_i}) = \begin{cases} \cos(m\theta), & 0 \leq \theta \leq \frac{\pi}{m} \\ D(\theta), & \frac{\pi}{m} \leq \theta \leq \pi \end{cases}$ 再通过对权重归一化

$\|W_{y_i}\| = 1, b=0$ 使得公式变为如下（2）：

$$L_2 = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{\|x_i\| \cos(m\theta_{y_i})}}{e^{\|x_i\| \cos(m\theta_{y_i})} + \sum_{j=1, j \neq y_i}^k e^{\|x_i\| \cos(\theta_j)}}, \quad (2)$$

通过这样的转化变形，使得原本只有决策边界的结果引入了决策余度（decision margin），这样对于开集问题的分类就能拥有更好的效果，但是由于 Sphreface 的余度的 θ 角度会随着余度的减小而减小，而当 $\theta=0$ 时则会完全消失。也就是 θ 太小时，其效果就会大幅下滑。

2.3.4 CosFace 损失函数

CosFace 损失函数。也是 LMCL。它跟 SphereFace 损失函数大部分相同，只有很小的不同点。CosFace 损失函数是在余弦空间里面定义的决策余度，而不同于 SphereFace 损失函数是在角度空间定义的决策余度。这样就使决策余度和决策边界能更好的各类样本的情况，所以，不管权重的角度如何变化 CosFace 损失函数都拥有很好的适应性，其具体公式如下（3）：

$$L_3 = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{s(\cos(\theta_{y_i})-m)}}{e^{s(\cos(\theta_{y_i})-m)} + \sum_{j=1, j \neq y_i}^k e^{s \cos \theta_j}}, \quad (3)$$

其中的 s 是超球体半径，m 为余量。

2.3.5 ArcFace 损失函数

ArcFace 损失函数与 CosFace 损失函数之间有许多相同的相似之处，也有一些细小的差异。它们的目的是要把分类界限最大化。CosFace 损失函数把分类的界限最大化的过程是余弦空间之中的，如果画出其分类边界图像其坐标应该是 $\cos \theta$ ，而 ArcFace 是在角度空间中

直接使分类的界限最大化如果画出其分类边界图像其坐标应该是 θ 。所以 ArcFace 损失函数对于角度细微的改变更加的敏感，角度变化对其影响更加直接。其具体的公式如下 (4)：

$$L_4 = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{s(\cos(\theta_{y_i+m}))}}{e^{s(\cos(\theta_{y_i+m}))} + \sum_{j=1, j \neq y_i}^k e^{s \cos \theta_j}}, \quad (4)$$

ArcFace 损失函数使用的是加法角度余度。所以这就使特征分布更为紧密，对应的决策边界也更明显，一个弧长即可表示一个类别。如图 1 所示。

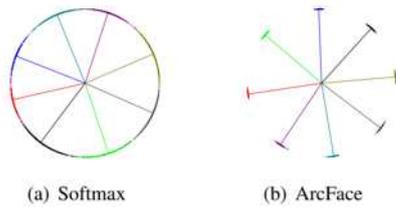


图 1 比较 Softmax 和 ArcFace

Fig.1 Compare ArcFace with Softmax

2.3.6 比较不同的损失函数

表 1 不同损失函数的决策边界

Tab.1 Decision Boundary of different Loss Functions

损失函数	决策边界
Softmax	$(W_1 - W_2)x + b_1 - b_2 = 0$
W-Norm Softmax	$\ x\ (\cos \theta_1 - \cos \theta_2) = 0$
SphereFace	$\ x\ (\cos m \theta_1 - \cos \theta_2) = 0$
F-Norm SphereFace	$s(\cos m \theta_1 - \cos \theta_2) = 0$
CosFace	$s(\cos \theta_1 - m - \cos \theta_2) = 0$
ArcFace	$s(\cos(\theta_1 + m) - \cos \theta_2) = 0$

不一样的损失函数对应也有不一样的决策边界，其描绘出来的图像也不同，具体如图 2 所示。

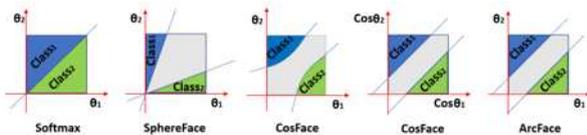


图 2 不同的损失函数的不同决策边界

Fig.2 Different Decision Boundaries of different Loss Functions

2.4 实验介绍及分析

2.4.1 实验设置

因为我们主要用于国内高校门禁识别，所以我们主要采用亚洲人脸 CASIA-FaceV5 数据集，其中共包含 2500 张亚洲人脸图片 (共 500 个身份)。然后我们对数据集使用 MTCNN 对数据进行人脸检测和关键点检测再通过旋转、拉伸等调整裁切为 112×112 的标准面部裁切图。对于网络结构，我们选择使用目前最为广泛使用的 ResNet50。通过比较不一样的损失函数对预测结果的影响。我们选择将 128 设置为批量大小值。把 0.0005 设置为权重衰减。并且我们将 0.1 设置为学习率。并且使用多分部的策略，步长分别设置为 1800, 3000, 当

到达指定步长时，学习率则除以 10，训练过程在迭代次数到达 5000 时完成。使用深度学习框架 MXNet 在 GeForce GTX1060 的 GPU 上训练，采用 python 语言。

2.4.2 实验数据集介绍

CASIA-FaceV5(CASIA Face Image Database Version 5.0) 数据集，这些亚洲人脸图像是由中科院的自动化研究所整理收集的，使用罗技 USB 相机捕获的图片。CASIA-FaceV5 的志愿者包括研究生，工人，服务生等。所有面部图像均为 16 位彩色 BMP 文件，图像为 640×480 的分辨率。典型的类内变化包括照明，姿势，表情，眼镜，成像距离等。CASIA-FaceV5 数据集人脸图像示意图如图 3 所示。



图 3 CASIA-FaceV5 数据集人脸图像示意图

Fig.3 CASIA-FaceV5 dataset face image sketch map

2.4.3 实验结果

受限于机器设备原因，经过多次训练，在基于 ArcFace 损失函数的情况下我们最高得出准确率为 97.1%。其他损失函数所得出的最高准确率，如表 2 所示。

表 2 不同损失函数的实验结果对比

Tab.2 Comparison of experimental results with different Loss Functions

损失函数	准确率
Softmax	92.6%
N-Softmax	92.5%
SphereFace	97.0%
CosFace	93.9%
ArcFace	97.1%

使用多种不同的损失函数并画出其准确率随迭代次数的变化。如图 4 所示。

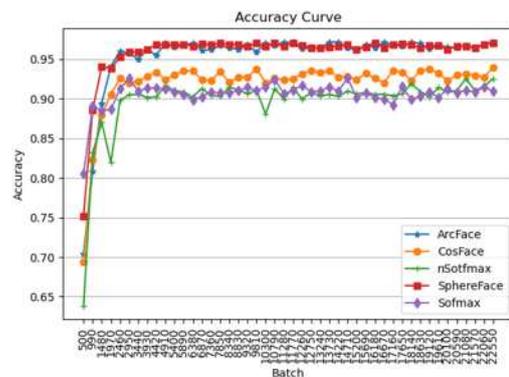


图 4 不同的损失函数准确率随迭代次数的变化图

Fig.4 The change chart of the accuracy of different Loss Functions with

the number of iterations

使用多种不同的损失函数并画出其损失值随迭代次数的变化。如图 5 所示。

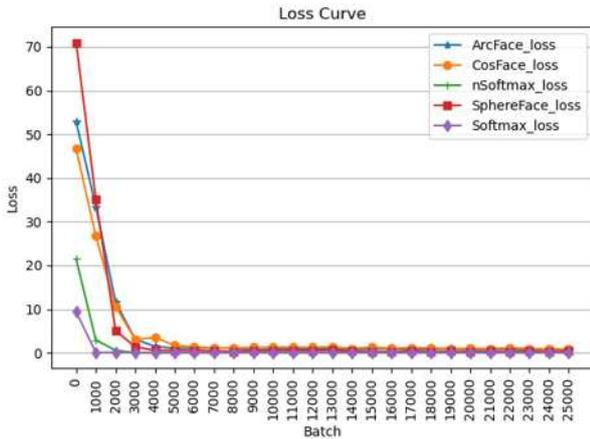


图 5 不同的损失函数损失值随迭代次数的变化图

Fig.5 The change chart of the loss of different Loss Functions with the number of iterations

2.4.4 实验结果分析

最终得出实验结果 ArcFace 损失函数的准确率最高，达到 97.1%，使用 Softmax 与 N-Softmax 还有 CosFace 得出的准确率比较低，整体上大致符合我们的预期结果。Softmax 与 N-Softmax 由于在类间样本中没有增强其多样性，在类内样本中也没有增强其相似性。所以决策边界没有划分的很开，最终导致准确率不如 ArcFace。

由于设备性能受限，最终得出最高准确率仅为 97.1%。虽然在要求较高的特定情况下还有提升空间，但是用于日常生活以及安全级别不是非常高的人脸识别的校园门禁上是能拥有良好的表现的。

3 校园门禁现状

随着中国的经济飞速的发展，社会越来越开放，校园安全也越来越受到重视，目前社会上的一些违法诈骗等事件也逐渐在进入校园。如何保障校园内安全是一个愈来愈重要和备受关注的事。目前传统的校园出入口一般设有门禁系统，并且要求刷卡出入，但是这样只识别校园卡，但不检测人员的方式，也存在着很大的安全隐患和缺陷。虽然目前大部分校园门口装有监控摄像，但是大多只是用来记录和录像的方式，因为采用人脸识别设备要求并不高，所以我们可以采用使用人脸识别来代替普通的监控作用，并且使用人脸识别也可以支持录像功能还能更好的区分不同人物。而使用人脸识别模式来替代传统的刷卡开门，可以更有效，更方便，也更安全地管理学校，避免了校园卡被盗刷的情况，也可以

记录晚归逃课等多种用途。

校园门禁系统，首先需要建立并注册一个校园人脸数据库，把数据库建立保存在服务器上并且训练好模型方便我们直接识别使用，然后当我们需要出入校园时，摄像头通过拍摄传送给服务器实时检测识别人脸，与建立的校园人脸数据库中的已经存储的或标记人脸进行对比，如若识别成功，则发送信号给门禁锁打开门禁。大致流程图，如图 6 所示。人脸校园门禁系统原型效果图如图 7 所示。

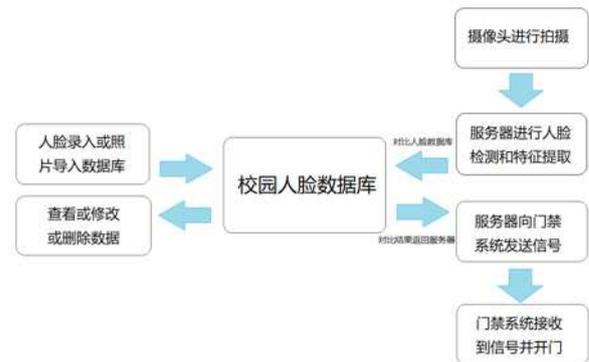


图 6 人脸校园门禁系统流程图

Fig.6 Campus face recognition access control system flow chart

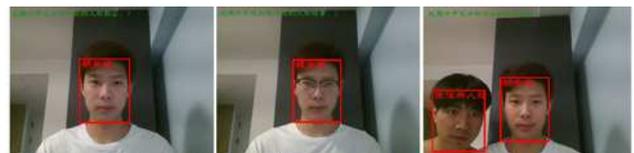


图 7 人脸校园门禁系统原型效果图

Fig.7 Prototype of campus face recognition access control system

4 结论

本文根据基于 ArcFace 损失函数的人脸识别算法设计与实现了基于 ArcFace 损失函数的人脸识别校园门禁，并且对比了多种不同损失函数，以验证识别效果高于其他损失函数的人脸识别算法。该系统可以通过摄像头，服务器，以及校园人脸数据库自动化对出入校园人员进行身份识别，判断是否许可其出入校园。有效地改善了传统意义上的校园门禁的便利性和安全性。

后续研究工作可以考虑进一步优化和改进网络结构使识别速度和准确率更高，以及加入活体检测功能和分析光照和遮挡物的影响因素。伴随着时代的进步与人脸识别技术的发展和更新换代，其准确率和速度不断地提高，人脸识别的技术也会愈加贴近到我们的生活学习与日常中，让我们的生活更加便利。

【参考文献】

- [1] F. Wang, W. Liu, H. Liu, and J. Cheng. Additive margin softmax for face verification. *IEEE Signal Processing Letters*, 2018.
- [2] Liu W Y, Wen Y D, Yu Z D, et al. SphereFace: deep hypersphere embedding for face recognition[C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21–26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 6738–6746.
- [3] Wang H, Wang Y, Zhou Z, et al. Cosface: large margin cosine loss for deep face recognition[C]// The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June18–23, 2018, Salt Lake City, UT, USA. New York: IEEE, 2018: 5265–5274.
- [4] Deng J, Guo J, Xue N, Zafeiriou S. ArcFace: additive angular margin loss for deep face recognition[C]// The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June15–20, 2019, Long Beach, CA, USA. New York: IEEE, 2019: 4690–4699.
- [5] Zhang K P, Zhang Z P, Li Z F, et al. Joint face detection and alignment using multitask cascaded convolutional networks[J]. *IEEE Signal Processing Letters*, 2016, 23(10): 1499–1503.
- [6] 顾岩. 基于深度学习的人脸识别技术及在油田作业区的应用研究 [D]. 成都: 电子科技大学, 2019.
- [7] A. Krizhevsky, I. Sutskever, G. E. Hinton. Imagenet classification with deep convolutional neural networks[C]. *Advances in neural information processing systems*, Lake Tahoe, USA, 2012, 1097–1105.
- [8] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint:1409.1556*, 2014.
- [9] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *CVPR*, 2015, 1.
- [10] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *CVPR*, 2016.