青廉科技

使用分类卷积神经网络训练的通用人脸对比模型研究

作　　者 王天泽

摘要

在现代信息化社会中，人脸识别技术的重要性不断增加，带来了全新的生活方式，具有广泛的应用场景。身份识别又是人脸识别的重要课题，而人脸对比是实现身份识别的快捷方法。传统的人脸对比方案会产生大量数据，提高训练成本。使用卷积神经网络搭建分类网络，并在使用时去除全连接层对比向量相似度的方法，可以有效解决这个问题，为身份识别提供技术支持。

关键词：人脸识别；人脸对比；卷积神经网络
文献标识码：A 中图分类号：TP391.41

目录

一、 引言

（一）背景

在现代信息化社会中，人脸识别技术的重要性不断增加，带来了全新的生活方式，具有广泛的应用场景[1]。其中，身份识别是人脸识别的重要课题，在金融、安防等领域有广泛的应用场景[1]，而人脸对比是实现身份识别的快捷方法。人脸对比不需要对个人进行大量的数据采集，只需采集一张人脸图片，即可判断相似度，从而实现身份识别。本文研究并提出了一种人脸对比的实现方法。

（二）国内外现状

现有的人脸对比实现已经使用上了神经网络，极大增强了其识别准确率。Florian Schroff 等人曾提出了 FaceNet 网络，在 Labeled Faces in the Wild（LFW）数据集上，其准确率达到 99.63%[2]。然而，这种网络在训练过程中需要传入两张图片及其是否为同一人，这将会导致预训练时处理一个标准的数据集（多个人，每个人有多张图片）的时候，会产生$O(n^3)$数量级的数据，即是原来长度的 3 次方。这会产生大量的数据成本和训练成本。

本文提出了一种人脸对比实现方法。该方法将会先通过分类模型训练人脸对比网络。当需要进行人脸对比时，去除其最后的全连接层并通过计算两个向量的相似度，并与设定的自信度阈值进行比较，从而实现人脸对比。

（三）框架选取

本文将采用由百度飞桨研发的 PaddlePaddle 深度学习框架（下文简称 Paddle 框架）。该框架支持传统深度学习框架的基本功能，如模型组网、显卡训练等。此外，该框架提供了高级 API，可以更加方便快捷地完成训练脚本的编写。

二、数据集选取

（一）训练数据集

本文选择 VggFace2 数据集（https://aistudio.baidu.com/datasetdetail/107435）。这个数据集包含了 1333 个人的 440028 张人脸图片（如图 1、图 2 所示），总量和分类丰富，且覆盖较大范围的姿态、年龄和民族，适合本项目。

图 1 第 87 号人脸的两张照片

图 2 第 341 号人脸的两张照片

（二）测试数据集

本文采用 facecap 数据集（https://aistudio.baidu.com/datasetdetail/104646）。该数据集包括 500 个人的人脸各 100 张（如图 3 所示），其总量相对较小，但仍有较大规模，适合在节省性能的同时获得更高的可信度。

图 3 第 5 号人脸的两张照片

三、数据集处理

本文采用 Paddle 框架搭建数据集。在数据集初始化时，程序会顺序执行以下操作：

1. 获取数据集文件夹下的所有子文件夹
2. 对每个子文件夹，遍历每个图片，顺序进行如下操作：
    a) 使用 PIL 打开图片
    b) 重设图片大小为 64x64
    c) 将通道维度置于最前，以适应 Paddle 框架。即将图片维度从(64, 64, 3)转为(3, 64, 64)
    d) 将图片和标签存入对应列表

当外界需要获取指定索引的数据时，程序会按照对应索引，从列表中读取并返回数据。

四、模型训练

（一）模型选取

本文采用 ResNet（Residual Neural Network）作为分类模型。ResNet 是深度学习中一种非常有效的卷积神经网络（CNN）结构，通过引入残差块（Residual Block）解决了深度神经网络训练中的梯度消失和模型退化问题。ResNet 的提出使得训练深度达到几百甚至上千层的神经网络成为可能，提高了模型的准确性，并在多项计算机视觉任务中取得了显著的性能提升[3]。

由于本文采用的训练数据集较大，且研究设备性能较低。为了减少训练时的时间成本，本文采用 ResNet-18（18 层的 ResNet 模型）进行分类任务，达到较好的性能与准确的平衡。

（二）优化器选取

在训练神经网络时，优化器的选择对模型的性能具有重要影响。优化器用于

更新和计算影响模型训练和模型参数的梯度，以最小化损失函数。本文采用 Adam（Adaptive Moment Estimation）作为优化器。

Adam 优化器将在梯度大时加速训练，梯度小时减速训练，而不是以固定的学习率训练。这将会大大加速训练过程，并且减少因训练集梯度问题而产生的训练问题，增加准确率[4]。

（三）损失函数选取

在机器学习和深度学习中，损失函数用于量化模型预测与真实标签之间的差异。选择一个合适的损失函数对于训练一个高效且准确的模型至关重要。本文采用交叉熵损失函数（Cross-Entropy Loss Function）进行训练。

在多分类问题中，通常有多个类别（假设为 C 个），模型的输出是一个 C 维的向量，每个元素表示样本属于对应类别的概率，其表达式如式(4.1)所示

$$L = -\sum_{i=1}^{C} y_i \log(p_i) \tag{4.1}$$

其中，C 表示样本个数，y_i是第 i 个样本的标签（0 或 1），p_i是模型输出的 C 维向量的第 i 维。

交叉熵损失函数常用于分类问题，衡量预测概率分布与真实概率分布之间的差异。

使用交叉熵损失函数的好处之一是它能够有效地处理类别不平衡的问题，因为它关心的是预测概率与真实概率之间的差异，而不是仅仅关注是否预测正确。此外，交叉熵损失函数在梯度下降过程中表现良好，能够提供稳定的梯度更新，有助于模型的快速收敛[5]。

（四）训练脚本编写

Paddle 框架提供的高级 API 极大简化了训练脚本的编写。因此，该脚本只需要设置优化器、损失函数和评价指标（准确率）并调用 model.fit 方法即可进行训练。另外，该脚本还通过回调函数，实现损失和准确率的记录，并写到一个 csv 文件中，以便获得损失和准确率的曲线。

（五）训练结果

经过 20 个 epoch 的训练后，模型的准确率达到 98%，并仍在缓慢上升。损失尽管也逐渐趋于平稳，但仍然存在一定的波动，如图 4 和图 5 所示。

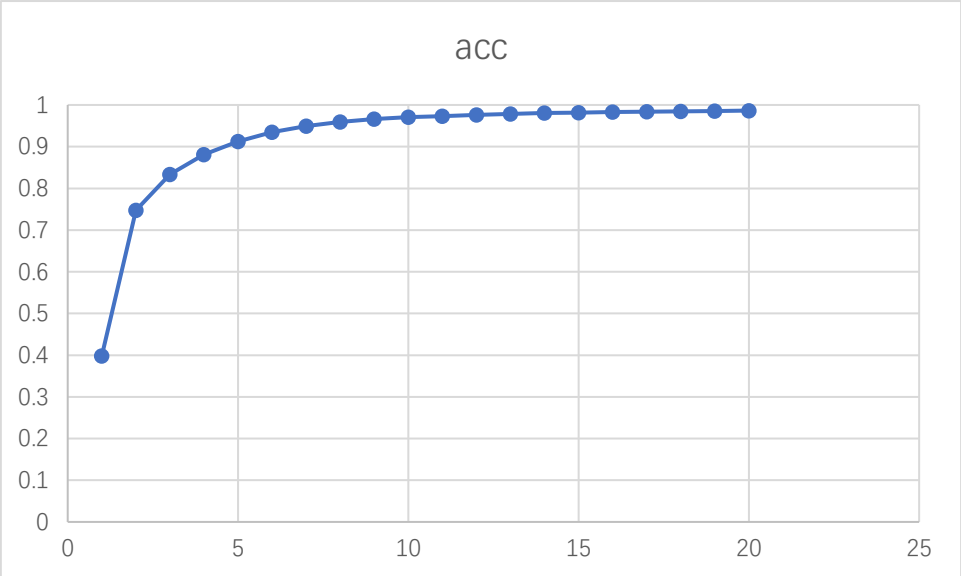

图 4 准确率随 epoch 的变化

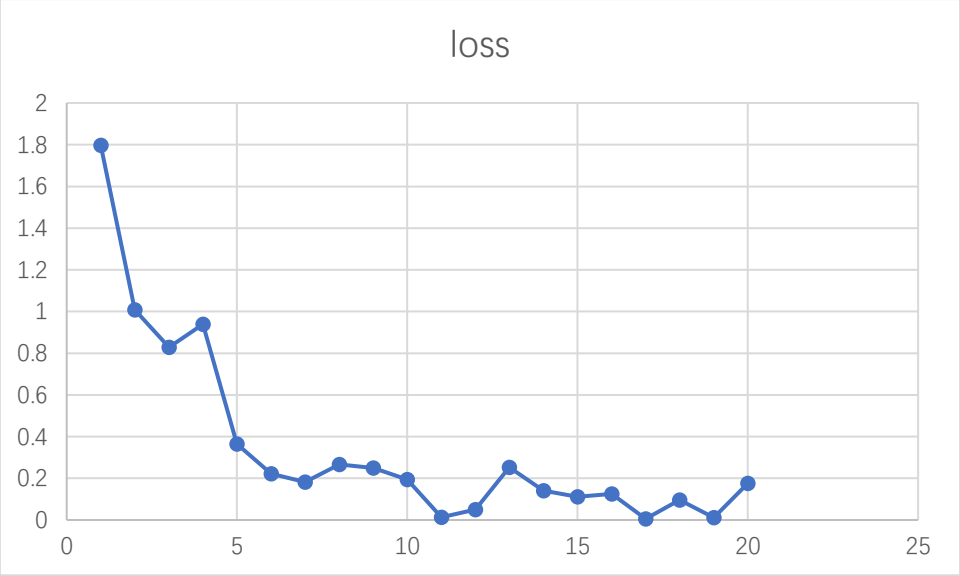

图 5 损失随 epoch 的变化

可以看到，模型仍有训练空间。但是由于实验机性能不足，无法继续训练。

五、相似度对比

在模型使用时，需要先去除模型的全连接层。当需要对比时，只需要将两张

人脸图片放入模型，然后对比输出后的向量的相似度即可。

模型将会输出一个 512 维的向量，如图 所示。该向量反映了该人脸的特征。

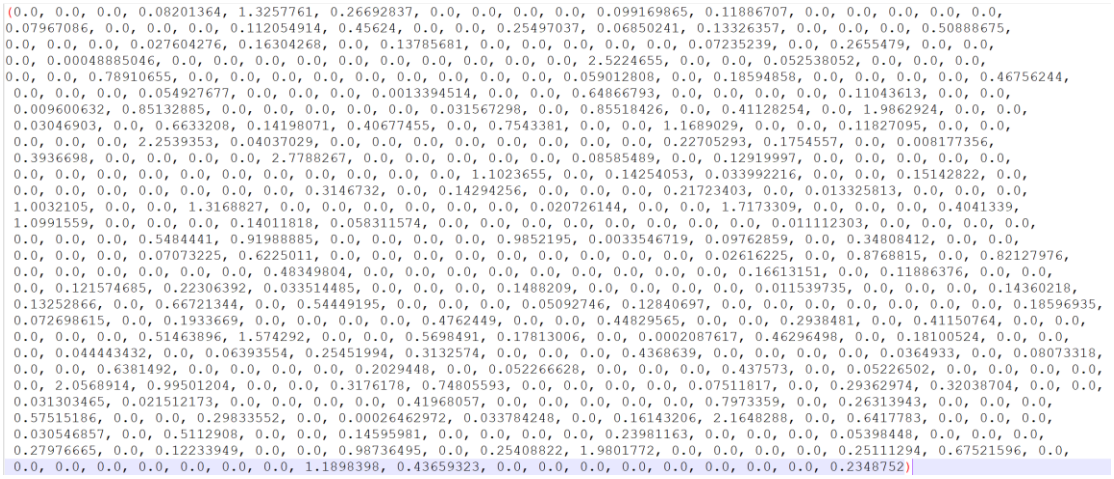

图 6 模型输出的特征向量示例

（一）相似度对比的方法

在对比相似度时，本文比较了两种方法。

1. 夹角余弦值判断

对于两个向量计算相似度时，可以向量夹角的余弦值进行判断，如式(5.1)所示[6]。

$$\cos\theta = \frac{\vec{A}\cdot\vec{B}}{|\vec{A}|\cdot|\vec{B}|} \tag{5.1}$$

由于本文所使用的模型最后会经过一个 ReLU 层，两个向量的夹角小于 180°。使用这种方法计算后的相似度，其值域为$[0,1]$，值域有上下界，比较适合比较相似度。

2. 欧氏距离判断

欧氏距离是指两个向量所表示的点之间的距离，如式(5.2)所示。

$$D = \sqrt{\sum_{i=1}^{n}(a_i - b_i)^2} \tag{5.2}$$

在欧氏距离的计算中，向量的长度和方向都被考虑进去，而不会向夹角余弦值只会考虑方向。但是，欧氏距离的值域是$[0,+\infty)$，会造成难以判断相似度最

大值，从而导致相似度判断不科学。

（二）方法选取

本文采用 P-R 曲线对每种方法其进行测试。

P 指精准率，其定义为被预测为正样本的实例中，真正为正样本的比例，如式(5.3)所示。

$$P = \frac{TP}{TP + FP} \tag{5.3}$$

R 指召回率，其定义为真正为正样本的实例中，被预测为正样本的比例，如式(5.4)所示。

$$R = \frac{TP}{TP + FN} \tag{5.4}$$

其中，TP 指真正例，即预测结果与实际均为正例的数量；FP 指假正例，即预测结果为正例，而实际为反例的数量；FN 指假反例，即预测结果是反例，而实际为正例的数量。

P-R 曲线需要测试不同阈值下的精准率和召回率，并以精准率为纵坐标，以召回率为横坐标。根据以上定义不难有如下结论：

- P-R 曲线越靠近右上角的(1, 1)点，模型的效果越好。
- 一个良好的模型的 P-R 曲线应该呈现递减趋势。
- 比较两个模型 A、B 的 P-R 曲线 a、b 时，如曲线 a 完全包住了曲线 b，则模型 A 由于模型 B.

据此，本文从测试集内选取 100 个正例、100 个反例，对夹角余弦值判断和欧氏距离判断分别绘制了 P-R 曲线。如图 7 所示。

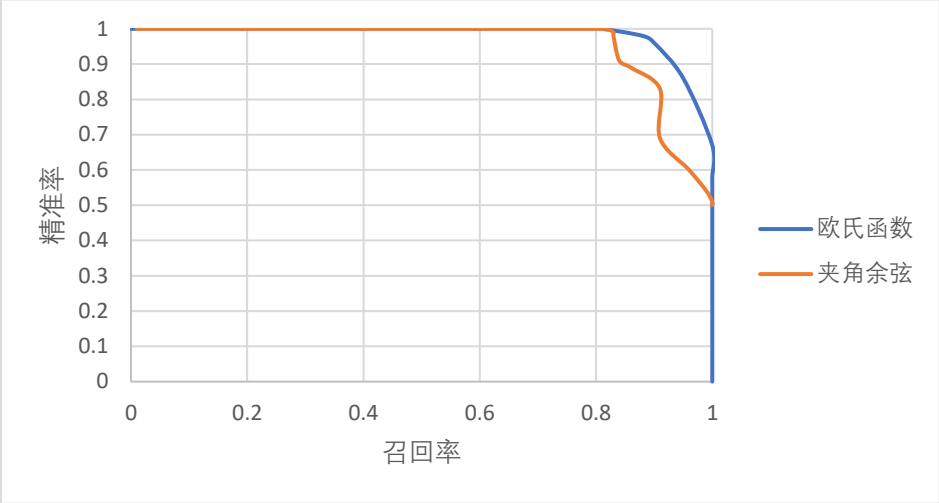

图 7 该模型的 P-R 曲线

可以看到，使用夹角余弦值和欧氏距离比较相似度的 P-R 曲线都比较靠近右上角的(1, 1)点，说明该模型的效果较好。而欧氏距离的曲线完全包住了夹角余弦的曲线，因此使用欧氏距离的效果更好。本文采用欧氏距离进行模型相似度比较。

六、自信度阈值选取

本文从测试集内选取 100 个正例、100 个反例，对不同自信度阈值测试其准确率。准确率定义为模型预测正确的次数与测试样例个数的比值。

本文通过对[0,20)的步长为 1 的自信度阈值测试其准确率，得出的结果如图 8 图 所示。

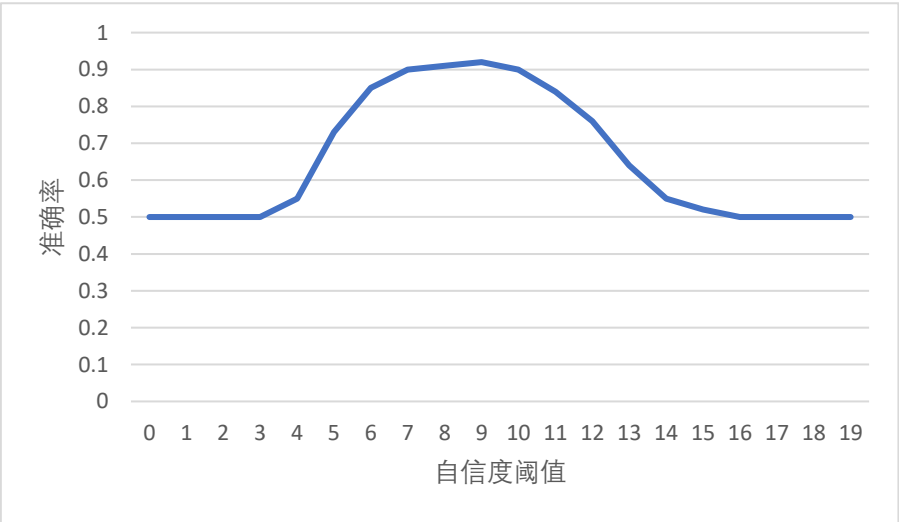

图 8 不同自信度阈值下的准确率

可以看到，图像呈现明显的中间高、两边低的趋势。当自信度阈值到 9 时，准确率到达最高点 92%。因此，本文选择 9 为自信度阈值。

七、模型推理能力测试

（一）模型准确度测试

本文测试集内选取 100 个正例、100 个反例，设定自信度阈值为 9，对该模型进行准确率测试。

本文最终测得模型准确率为 92%，与 FaceNet 仅差 7%左右。

另外，本文测得模型的精准率为 0.98，召回率为 0.85，F1 值为 0.91. 可以看出，该模型的准确度较高。

（二）模型速度测试

本文在 Redmi G 2022 笔记本上使用 GPU 进行模型推理，进行速度测试。该笔记本的 CPU 型号为 i7 12650H、内存为 DDR5 的 16GB 内存、显卡为 RTX3050Ti.

1. 启动速度测试

本文对该模型的启动速度进行 5 次测试，结果如图 6 所示。

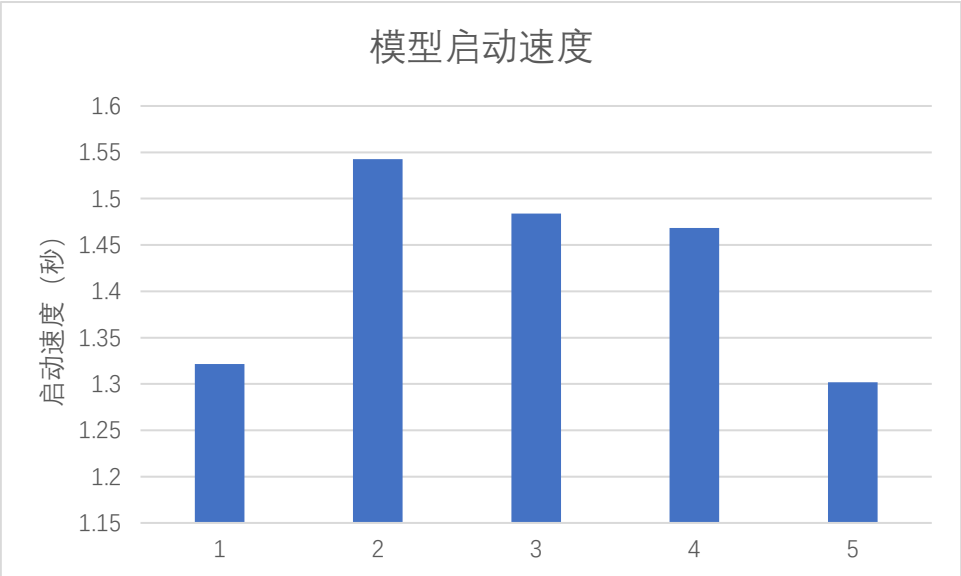

图 6 模型启动速度测试

该模型的启动速度平均值为 1.42 秒，可见该模型的启动较快。

2. 推理速度测试

本文对该模型的推理速度进行 100 次测试，并由计算机自动求出平均值，为 0.014 秒。可见该模型的推理速度较快。

八、结论与展望

本文提出了一种基于卷积神经网络和输出向量的人脸对比实现方法，并验证了其有效性，证明其拥有较高的准确度和训练性能。该方法和模型为人脸对比提供了一种方案。

具体来说，在训练时，该方法将会训练一个基于卷积神经网络的人脸分类模型。在使用时，只需要去除模型的全连接层，获得两个人脸的向量，并使用向量进行相似度对比即可实现人脸对比。

该模型较高的准确率可以胜任身份识别的任务，能够应用于金融、安防等多种场景。

参考文献

[1]赵晓霞. 人脸识别数字化[N]. 人民日报海外版,2010-06-24(004).

[2]Schroff F ,Kalenichenko D ,Philbin J .FaceNet: A Unified Embedding for Face Recognition and Clustering.[J].CoRR,2015,abs/1503.03832

[3]孙毅,吴斯曼,方伟,等.基于 ResNet 的安全监控目标检测[J/OL].集成技术 :1-10[2024-06-06].http://kns.cnki.net/kcms/detail/44.1691.t.20240527.1512.004.html.

[4]张波,肖杰.深度学习模型训练的优化器实验设计[J].电子制作,2024,32(02):114-117.DOI:10.16589/j.cnki.cn11-3571/tn.2024.02.023.

[5]黄辉城,李建新.基于深度视觉的智能卸垛机快递箱边界检测系统[J].光学技术,2024,50(02):220-227.DOI:10.13741/j.cnki.11-1879/o4.2024.02.018.

[6]李艺, 董玉琦, 等. 信息技术[M]. 北京: 教育科学出版社, 2020.