



改进的轻量化人脸识别算法

屈东东, 贺利乐, 何林

引用本文:

屈东东,贺利乐,何林. 改进的轻量化人脸识别算法[J]. *智能系统学报*, 2023, 18(3): 544–551.

QU Dongdong, HE Lile, HE Lin. Improved lightweight face recognition algorithm[J]. *CAAI Transactions on Intelligent Systems*, 2023, 18(3): 544–551.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202111051>

您可能感兴趣的其他文章

facenet皮尔森判别网络的人脸识别方法

Face recognition method based on facenet Pearson discrimination network

智能系统学报. 2022, 17(1): 107–115 <https://dx.doi.org/10.11992/tis.202104008>

基于改进残差学习的东巴象形文字识别

Dongba pictographs recognition based on improved residual learning

智能系统学报. 2022, 17(1): 79–87 <https://dx.doi.org/10.11992/tis.202112009>

神经网络多层特征信息融合的人脸识别方法

Face recognition method based on neural network multi-layer feature information fusion

智能系统学报. 2021, 16(2): 279–285 <https://dx.doi.org/10.11992/tis.201907053>

基于增强AlexNet的音乐流派识别研究

Music genre recognition research based on enhanced AlexNet

智能系统学报. 2020, 15(4): 750–757 <https://dx.doi.org/10.11992/tis.201909032>

人脸识别背后的数据清理问题研究

Research on the data cleansing problem for face recognition technology

智能系统学报. 2017, 12(5): 616–623 <https://dx.doi.org/10.11992/tis.201706025>

一种特征加权融合人脸识别方法

Face recognition by weighted fusion of facial features

智能系统学报. 2015, 10(6): 912–920 <https://dx.doi.org/10.11992/tis.201509025>

DOI: 10.11992/tis.202111051

网络出版地址: <https://kns.cnki.net/kcms/detail/23.1538.tp.20230330.1818.002.html>

改进的轻量化人脸识别算法

屈东东¹, 贺利乐¹, 何林²

(1. 西安建筑科技大学 机电工程学院, 陕西 西安 710055; 2. 西安建筑科技大学 理学院, 陕西 西安 710055)

摘要: 嵌入式平台计算资源有限, 无法实时运行计算量和参数量巨大的深度学习模型。基于 mobilenet v2 提出一种改进的轻量化人脸识别算法 L-mobilenet v2, 首先对原有网络结构进行优化, 然后以三元损失函数为主, 将传统分类任务中的 softmax 损失改为 Am-softmax 作为辅助损失函数, 使用 10575 个人的 49 万张图片进行联合训练。相比于改进前的模型及训练方法, 新模型在 LFW 测试集和自制数据集的识别准确率达到 98.56% 和 95%, 将模型参数量缩减 72.3% 的同时将识别准确率提高了 1.56% 和 7.1%, 在嵌入式平台 Jetson nano 上的平均识别帧率提升了 36.3%。该模型可以在计算资源受限的移动端实时运行。

关键词: 嵌入式平台; 深度学习; 人脸识别; 轻量化网络; mobilenet v2 模型; softmax 损失; Am-softmax 损失; Jetson nano 平台

中图分类号: TP391.4 **文献标志码:** A **文章编号:** 1673-4785(2023)03-0544-08

中文引用格式: 屈东东, 贺利乐, 何林. 改进的轻量化人脸识别算法 [J]. 智能系统学报, 2023, 18(3): 544-551.

英文引用格式: QU Dongdong, HE Lile, HE Lin. Improved lightweight face recognition algorithm[J]. CAAI transactions on intelligent systems, 2023, 18(3): 544-551.

Improved lightweight face recognition algorithm

QU Dongdong¹, HE Lile¹, HE Lin²

(1. School of Mechanical and Electrical Engineering, Xi'an University of Architecture and Technology, Xi'an 710055, China; 2. School of Science, Xi'an University of Architecture and Technology, Xi'an 710055, China)

Abstract: Due to the limited computing resources, the embedded platform can not run the deep learning model with huge amounts of calculation and parameters in real time. An improved lightweight face recognition algorithm L-mobilenet v2 is proposed based on the mobilenet v2. The algorithm first optimizes the original network structure, then uses the triplet loss function as the main loss function to change the softmax loss in traditional classification task to Am-softmax, which is used as the auxiliary loss function, and uses 490 thousand images of 10575 people for joint training. Compared with the previous model and training method, the recognition accuracy of the new model on LFW test data set and self-made data set has reached 98.56% and 95%, respectively, which increased the recognition accuracy by 1.56 % and 7.1% while reducing the number of model parameters by 72.3%. And at the same time, the rate of frame recognition on average on the embedded platform Jetson nano has increased by 36.3%. The model can run in real time on mobile terminals with limited computing resources.

Keywords: embedded platform; deep learning; face recognition; lightweight network; mobilenet v2 model; softmax loss; Am-softmax loss; Jetson nano

人脸识别技术是现代社会应用最广泛的技术之一。基于特征空间的主元分析法 (principal component analysis, PCA) 将人脸映射到低维空间

后存在较多的非零元, 李东博等^[1]提出了一种重加权稀疏主成分分析算法对此进行了优化。徐竞泽等^[2]结合主元分析法 PCA、线性判别分析法 (linear discriminant analysis, LDA) 和支持向量机 (support vector machine, SVM) 提出一种新的人脸识别融合算法。这些传统的人脸识别算法多是提

收稿日期: 2021-11-29. 网络出版日期: 2023-03-31.

基金项目: 陕西省教育厅专项科研项目 (21JK0732).

通信作者: 何林. E-mail: helin716@163.com.

©《智能系统学报》编辑部版权所有

取人脸较为浅层的特征, 深度学习的兴起改变了这一现状。Liu 等^[3] 基于 ResNet64 主干网络提出带乘性角度间隔的损失函数 SphereFace, 训练后在测试集 LFW (labeled faces in the wild) 准确率为 99.42%。Wang 等^[4] 将特征和权重归一化, 将 SphereFace 的乘性角度间隔改变为加性余弦间隔 CosFace, 将在 LFW 上的准确率提升为 99.73%。2019 年, Deng 等^[5] 提出一种类似 CosFace 的加性角度损失函数 ArcFace, 在 LFW 的识别准确率达到 99.83%。以上损失函数都是基于 softmax 做出的系列改进, 目的在于学习到更加精细化的特征空间, 这些损失函数在引入类边际的同时也对类边际进行了固定。2020 年, Sun 等^[6] 提出 Circleloss 损失将类边际进行拆分, 在 LFW 的识别准确率达到 99.73%。2021 年, 才华等^[7] 基于 Circleloss 进一步对边际问题进行优化, 提出自适应圆边际的 AdaCMloss, 增强了模型泛化能力, 在 LFW 的识别准确率达到 99.79%。

基于深度学习的人脸识别算法准确率不断提高, 模型也愈加复杂, 参数量也愈加庞大。为了解决这些问题, 各种轻量化的网络相继提出, Howard^[8] 等的 mobilenet 系列网络、Zhang^[9] 等的 ShuffleNet、Tan 等^[10] 的 EfficientNet, 这些轻量化网络的提出对人脸识别在计算资源有限的嵌入式设备上部署提供了可能。本文在 mobilenet v2^[11] 的基础上提出一种极其轻量化的人脸识别模型 L-mobilen-

et v2, 在显著减少模型参数数量的同时提高人脸识别的准确率, 可以方便地在移动端进行部署。

1 改进的 mobilenet v2 人脸识别算法

1.1 改进的 mobilenet v2 人脸识别原理

改进的 mobilenet v2 人脸识别原理包括人脸检测与对齐、人脸识别两部分。目前人脸检测常用的检测方法包括 Fast R-CNN^[12]、SSD^[13]、RetinaFace^[14] 和级联卷积网络。级联网络是传统方法和卷积网络的典型代表, 本模型的人脸检测与对齐部分采用多任务级联卷积网络 MTCNN^[15] 来实现。MTCNN 包含粗略 (P-Net)、半精细 (R-Net)、精细 (O-Net) 3 种不同精度的网络结构。首先对输入图像进行不同尺度的缩放, 将缩放后的金字塔图像送入三层级联网络依次处理, 回归出人脸位置 (bounding box)、人脸关键点 (facial landmark)、人脸分类置信度 (classification)。然后对输入图像使用人脸位置的相关参数截取人脸, 最后使用人脸关键点对齐人脸。参考谷歌 Facenet^[16] 人脸识别, 利用相同人脸照片的高内聚性、不同人脸照片的低耦合性, 将改进后的网络模型作为人脸特征提取器, 把级联网络已经对齐好的人脸送入 L-mobilenet v2 模型提取人脸特征, 得到一个长度为 128 维的特征向量。通过比较高维空间中特征向量的欧氏距离, 从而判定两张人脸是否为同一个人, 本文人脸识别原理如下图 1 所示。

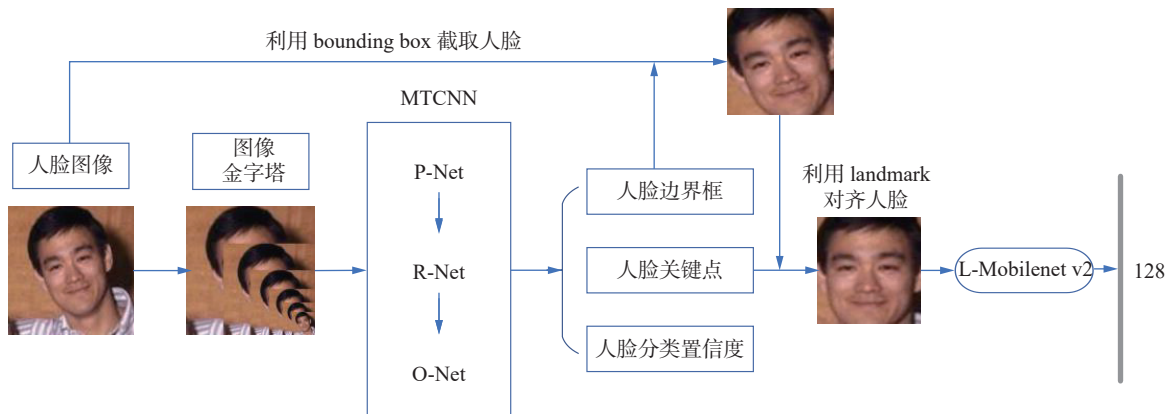


图 1 改进的 mobilenet v2 人脸识别原理

Fig. 1 Face recognition principle base on the improved mobilenet v2

1.2 改进的 mobilenet v2 人脸识别网络

改进的 mobilenet v2 继承了 mobilenet 的深度可分离卷积和 mobilenet v2 的倒置残差结构。深度可分离卷积是普通卷积进一步优化结果, 图 2 是利用单个卷积核进行普通卷积和深度可分离卷积的具体运算流程。对于一幅宽高为 $D_w \times D_h$ 、通道数为 M 的输入图像, 利用 M 个

$D_k \times D_k$ 卷积核 (D_k 一般为奇数) 进行卷积运算。在每个通道上, 普通卷积的卷积核分别与对应像素相乘再相加得到 X_i , 然后 M 个通道的 X_i 求和附加上偏置项 b , 得到输出特征图上一个像素; 深度可分离卷积的卷积核分别与对应像素相乘再相加得到 Y_i , 然后使用 $1 \times 1 \times M$ 的点卷积分别与深度卷积的输出 Y_i 对应相乘得到 Z_i ,

M 个通道的 Z_i 相加得到输出特征图上的一个像素。

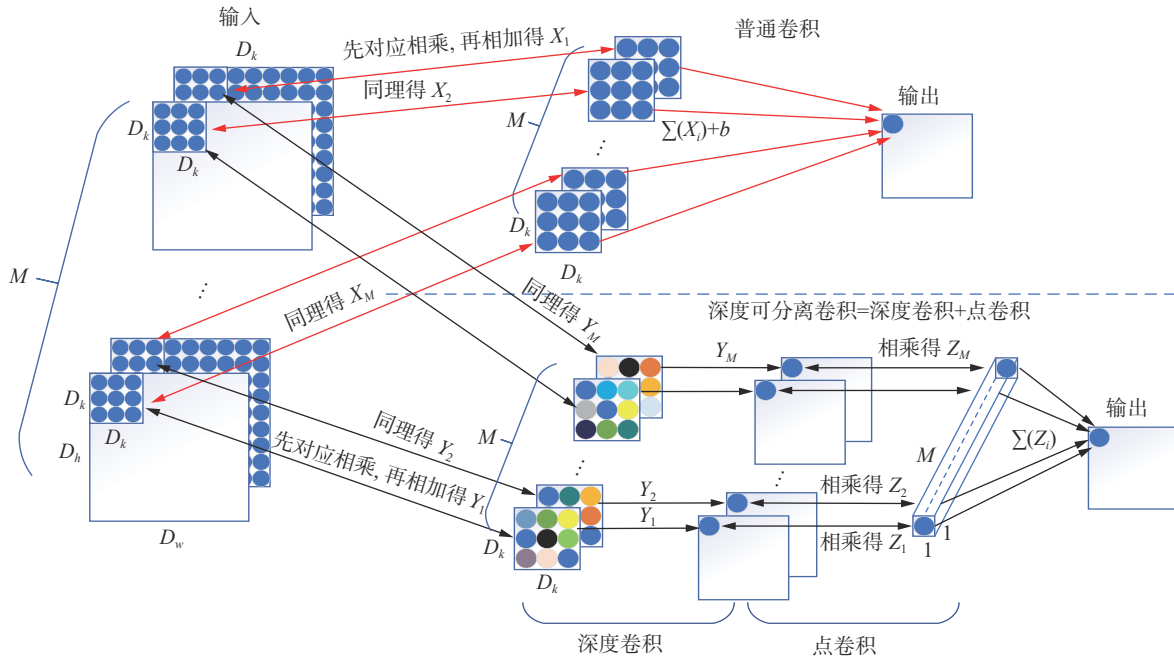


图 2 普通卷积和深度可分离卷积运算过程

Fig. 2 Operation process of ordinary convolution and depthwise separable convolution

对于 N 通道的输出特征图, 普通卷积参数量 $D_k \times D_k \times M \times N$, 计算量 $D_k \times D_k \times M \times N \times D_w \times D_h$; 深度可分离卷积参数量 $D_k \times D_k \times M + M \times N$, 计算量 $D_k \times D_k \times M \times D_w \times D_h + M \times N \times D_w \times D_h$ 。

$$\frac{D_k \times D_k \times M + M \times N}{D_k \times D_k \times M \times N} = \frac{1}{N} + \frac{1}{D_k^2} \quad (1)$$

$$\frac{D_k \times D_k \times M \times D_w \times D_h + M \times N \times D_w \times D_h}{D_k \times D_k \times M \times N \times D_w \times D_h} = \frac{1}{N} + \frac{1}{D_k^2} \quad (2)$$

通过 (1)、(2) 两式比较可以发现, 深度可分离卷积的参数量和前向传播的计算量都约等于普通卷积的 $\frac{1}{N} + \frac{1}{D_k^2}$ 倍, 而输出通道 N 一般较大, 卷积核尺寸 $D_k \times D_k$ 一般最小取 3×3 , 故深度可分离卷积比传统卷积的参数量和计算量都至少减少为传统卷积的 $\frac{1}{9}$ 。

倒置残差是残差结构 Residuals^[17] 的一种反设计。残差结构是在深度可分离卷积前使用点卷积对输入特征图进行维度上的“压缩”, 再使用深度可分离卷积中的点卷积进行维度上的“扩张”。倒置残差结构则先使用点卷积进行维度“扩张”, 再使用深度可分离卷积中的点卷积进行维度上的“压缩”还原。图 3 是本文改进后的 2 种倒置残差结构块。 t 是维度扩展因子, 步长 $s=1$ 时使用跳连接, $s=2$ 时不使用跳连接。

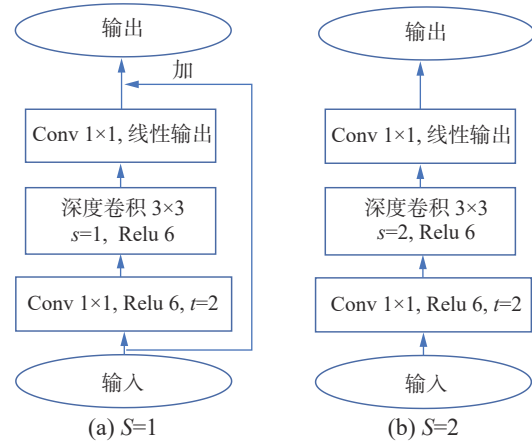


图 3 改进的 bottleneck 结构块

Fig. 3 Improved bottleneck structure blocks

本文网络结构如表 1 所示, 网络输入大小为 $160 \times 160 \times 3$ 的三通道图像, 与 mobilenet v2 主要区别如下: 1) 原模型输入层选择 32 个卷积核进行普通卷积, 本文选用 64 个 3×3 的卷积核, 经过步长为 2 的普通卷积后直接进行深度可分离卷积; 2) 在 bottleneck 结构块维度扩展因子上, 原模型的维度扩展因子 t 全部等于 6, 本文模型全部设置为 2; 3) 在 bottleneck 结构块数量选择上, 原模型结构块按照降采样设置的不同分为 7 种, 每种 bottleneck 结构块的初始步长 s 为 1-2-2-2-1-2-1, 重复次数 n 分别为 1-2-3-4-3-3-1。本文模型将 bottleneck 结构块降为 5 种, 初始步长 s 分别为 2-2-1-2-1, 重复次数 n 为 2-1-3-1-2; 4) 对于模

型倒数第 2 层 512 个通道的 10×10 特征图, 显然每张特征图边缘部分和中心区域所包含信息的重要程度不同, 原模型对整张特征图全部使用平均池化是不合适的, 本文使用全局深度卷积 (512 个 10×10 的卷积核分别对应 512 个通道, 卷积核和对应像素相乘再求和) 来代替平均池化, 根据不同位置的像素点赋予不同的权重, 从而达到加权平均的效果, 以此提升网络的性能。

表 1 L-mobilenet v2 网络结构
Table 1 L-mobilenet v2 network structure

| 输入 | op | t | c | n | s |
|-------------------|----------------------|-----|-----|-----|-----|
| $160^2 \times 3$ | conv2d, 3×3 | — | 64 | 1 | 2 |
| $80^2 \times 64$ | dw, 3×3 | — | 64 | 1 | 1 |
| $80^2 \times 64$ | bottleneck | 2 | 64 | 2 | 2 |
| $40^2 \times 64$ | bottleneck | 2 | 128 | 1 | 2 |
| $20^2 \times 128$ | bottleneck | 2 | 128 | 3 | 1 |
| $20^2 \times 128$ | bottleneck | 2 | 128 | 1 | 2 |
| $10^2 \times 128$ | bottleneck | 2 | 128 | 2 | 1 |
| $10^2 \times 128$ | conv2d, 1×1 | — | 512 | 1 | 1 |
| $10^2 \times 512$ | gdw, 10×10 | — | 512 | 1 | 1 |
| $1^2 \times 512$ | fc | — | 128 | 1 | 1 |

表 1 中, op 表示某网络层或者结构块的运算方式; t 表示 bottleneck 结构块维度扩展因子; c 表示卷积核个数; n 表示某层或者结构块重复次数; s 表示步长 (在 bottleneck 中仅表示第一个结构块的步长, 其余步长全为 1); gdw 表示全局深度卷积, 即选择 512 个 10×10 的卷积核与原图进行对应相乘再求和, 相当于对原图不同位置赋予不同权重, 达到加权平均的效果; fc 表示全连接层。

1.3 改进的联合监督信号训练模型

损失函数在模型训练中起着监督和指导作用, 影响着模型最终的识别精度。目前人脸损失函数要么在样本欧氏裕度上进行监督, 如三元损失 Triplet loss、Centerloss^[18] 等, 要么在余弦和角度间隔上施加约束, 如 SphereFace、CosFace、Am-softmax^[19] 等。本文提出一种改进的联合训练方法, 以三元损失作为主损失函数, 以 Am-softmax 作为辅助损失来训练网络模型。

三元损失主要用来训练类似于人脸这样差异性较小的样本。1 个三元组在训练集上选择 3 个样本, 固定样本、正样本、负样本, 固定样本和正样本为同一个人的最不相似的两张照片, 负样本为训练集中和此人最相似的非同一人照片。三元

损失的目的就是通过学习训练数据集中最有难度的此类样本, 从而使得相同样本类内距离小于不同样本的类间距离, 构建的约束条件为

$$\|f(\mathbf{x}_i^a) - f(\mathbf{x}_i^p)\|_2^2 + \alpha < \|f(\mathbf{x}_i^a) - f(\mathbf{x}_i^n)\|_2^2 \quad (3)$$

$$L_{TRI} = \sum_{i=1}^N [\|f(\mathbf{x}_i^a) - f(\mathbf{x}_i^p)\|_2^2 - \|f(\mathbf{x}_i^a) - f(\mathbf{x}_i^n)\|_2^2 + \alpha]_+$$

将式 (3) 写成待优化的损失函数形式:

$$\forall (f(\mathbf{x}_i^a), f(\mathbf{x}_i^p), f(\mathbf{x}_i^n)) \in \tau \quad (4)$$

式中: τ 表示所有可能的三元组合; N 表示批次的大小; $f(\mathbf{x}_i^a), f(\mathbf{x}_i^p), f(\mathbf{x}_i^n)$ 分别表示第 i 个三元组中固定样本、正样本、负样本的模型输出; $\|\cdot\|_2^2$ 下标的数字 2 分别表示平方和 2 范数; α 相当于附加余量, 决定类间距的最小值, 本文取 0.2; + 号表示 $[\]$ 内的值为正时取为损失值, 为负时损失为零。

传统 softmax 损失在分类问题中取得了较好的效果, 然而在超多分类问题中, 随着类别增多我们希望分类更加精细化, 而传统分类边界过于粗糙, Am-softmax 属于度量学习策略的一种, 更加注重缩小类内距离和增大类间距离。图 4 形象地解释了传统分类和 Am-softmax 分类的区别, 对于类别 W_1 和 W_2 , 传统分类的类边界如 P_0 , 边界距离为 0。Am-softmax 划分的类边界如 P_1 和 P_2 , 不仅明确了类别的目标区域, 还产生了 margin 大小的类间距。

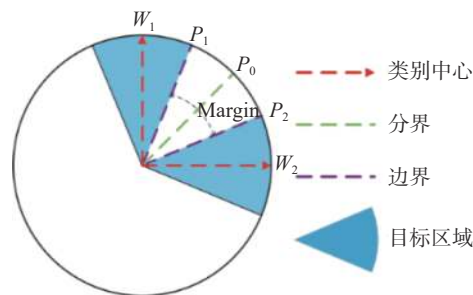


图 4 softmax 和 Am-softmax

Fig. 4 softmax and Am-softmax

Am-softmax 损失函数为

$$L_{AMS} = -\frac{1}{N} \sum_{i=1}^N \ln \frac{\exp(s(\cos \theta_{y_i} - m))}{\exp(s(\cos \theta_{y_i} - m)) + \sum_{j=1, j \neq y_i}^c \exp(s \cos \theta_j)} \quad (5)$$

式中: N 是批次的大小; s 是尺度缩放因子, 可以扩大余弦距离; y_i 是输入特征 \mathbf{x}_i 对应的模型输出; θ_{y_i} 是最后一个全连接层 y_i 所属类别的权值向量和 \mathbf{x}_i 之间的夹角; m 作为边际参数 (Additive Margin) 来扩大类间距; c 是分类的总类别数; θ_j 是第 j 类最后一个完全连接层的权重和 \mathbf{x}_i 之间的夹角。参考 ArcFace 损失, 本文设置 $m=0.35, s=30$ 。

最终的总损失为 L_{TRI} 与 L_{AMS} 直接相加的线性和, 从式 (4) 可以看出, 选择合适的三元组合使得三元损失最优化可以加速模型的收敛, 然而实际中选择合适的三元组样本是非常困难的, 这也是单独使用三元损失模型难以收敛的重要原因, 故本文采用 Am-softmax 损失函数替换传统的 softmax 作为辅助收敛的重要手段。

2 实验及结果分析

2.1 实验数据和环境

实验数据集选择开源 CASIA-WebFace^[20], 包含 10575 个人的 494414 张人脸图片, 每张图片分辨率为 250×250 。由于该数据集中的人脸并未对齐, 包含部分模糊甚至无效的照片, 故需要对数据集进行一次集中清洗。利用级联网络 MTCNN 对整个数据集进行预处理, 截取检测到的人脸并对齐, 对未检测到的人脸直接剔除, 并且缩放到分辨率为 160×160 , 最终得到 10575 个人的 452960 张人脸, 如图 5 所示。该实验数据集的 95% 用作训练, 5% 用作验证。



图 5 训练数据集
Fig. 5 Training data set

测试集选择人脸识别领域常用数据集 LFW^[21-22], LFW 人脸图片采样自不同光照、不同角度及不同的自然场景, 共包含 5749 个人的 13233 张人脸图片。测试集从中随机选择 6000 对人脸, 3000 对为同一人, 3000 对为不同人。通过测试数据集给出的一对人脸是否为同一个人, 判断模型识别人脸的准确率。

同时, 本文自制数据集, 通过收集身边同学、老师以及部分中科院(亚洲人)人脸数据集 Face v5, 共计 500 个人, 每人 2~5 张不同角度和姿态的照片, 随机进行配对, 300 对同一人脸, 300 对不同人脸, 如图 6 所示。

实验环境为 Ubuntu 18.04 操作系统, 8 GB 内存、NVIDIA GeForce RTX 2070 显卡, 在 Pycharm 2020.3.4 上利用深度学习框架 TensorFlow 训练模型。

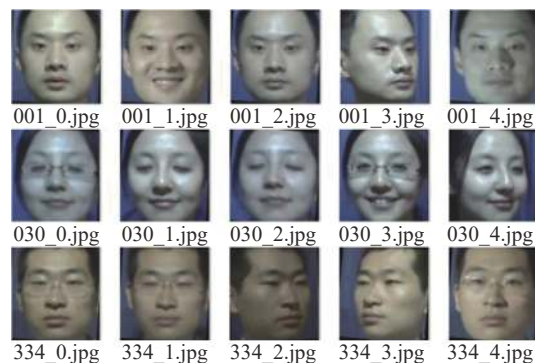


图 6 自制部分人脸数据
Fig. 6 Myself partial face data

2.2 实验过程

本次训练采用 Adam(adaptive moment estimation)^[23] 优化器指导模型训练, 模型初始化权重采用 Xavier 分布, 初始学习率为 0.01, 批训练大小为 16, 拟训练 30 个迭代, 每一迭代(覆盖训练集所有样本的一次训练为 1 个迭代)的训练集和验证集都经过了打乱重排, 一个迭代耗时约 2 h。以验证集上的损失值作为监测信号, 连续 5 个迭代的损失值不下降, 学习率衰减为原来的 1/10, 连续 10 个迭代损失值不下降则实现早停, 防止模型产生过拟合, 丢弃率经过多次实验选择 0.5。

实验设置如下:

1) 为验证本文网络结构的有效性, 采用相同的训练策略和损失函数 (softmax+Triplet 或者 Am-softmax+Triplet) 进行训练, 以 mobilenet v2 和本文的 L-mobilenet v2 网络进行两组对比实验;

2) 为验证本文改进后的联合损失函数的有效性, 采用相同的训练策略和网络结构 (mobilenet v2 或者 L-mobilenet v2), 以改进后的 Am-softmax+Triplet 损失和改进前的 softmax+Triplet 损失进行两组对比实验。

3) 以 ResNet50 作为主干特征提取网络, 使用改进前后的联合损失函数分别进行训练。

2.3 实验结果分析

图 7 为本次实验 6 种模型在测试集 LFW 的准确率折线图, 6 种模型均在第 28 代之前实现早停。折线 a 采用 mobilenet v2 和改进前的联合损失函数训练, 折线 b 采用 mobilenet v2 和改进后的联合损失函数训练, 折线 c 采用 L-mobilenet v2 和改进前的联合损失函数训练, 折线 d 采用 L-mobilenet v2 和改进后的联合损失函数训练, 折线 e 采用 ResNet50 和改进前的联合损失函数训练, 折线 f 采用 ResNet50 和改进后的联合损失函数训练。可以看到采用 L-mobilenet v2 和改进后的联合损失函数训练的 d 模型, 模型准确率是最高的。

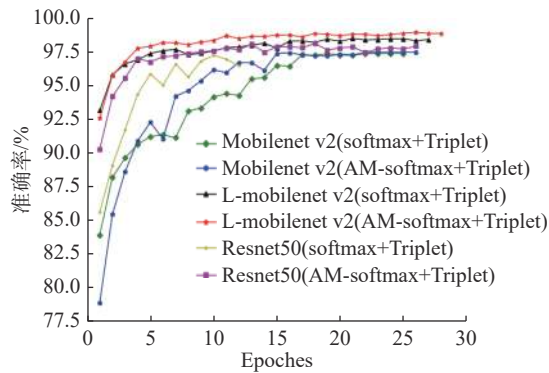


图7 6种模型在测试集LFW的准确率折线图

Fig. 7 Accuracy line chart of six models in the LFW

实验(1)的结果从a、c或b、d折线可以看出,本文L-mobilenet v2网络结构优于mobilenet v2,在LFW的识别准确率分别提高了1.1%和1.45%。

表2 模型性能综合对比

Table 2 Comparison of model comprehensive performance

| 模型 | 存储尺寸/MB | 参数量/ 10^6 | 准确率/% | | 速率/($\text{f}\cdot\text{s}^{-1}$) (Jetson Nano) |
|--------------------------------------|---------|-------------|-------|-------|--|
| | | | LFW | 自制数据集 | |
| Inception-resnet-v1 | 88.1 | 22.8 | 99.05 | 96.66 | 4 |
| ResNet50(softmax+Triplet) | 96.5 | 23.85 | 96.85 | 86 | 4 |
| ResNet50(AM-softmax+Triplet) | 96.5 | 23.85 | 97.73 | 92.83 | 4 |
| Lightfacenet ^[24] | — | 1.1 | 99.32 | — | — |
| mobilenet v2(softmax+Triplet) | 9.68 | 2.42 | 97 | 87.9 | 11 |
| mobilenet v2(AM-softmax+Triplet) | 9.68 | 2.42 | 97.11 | 88.16 | 11 |
| 本文L-mobilenet v2(softmax+Triplet) | 2.83 | 0.67 | 98.10 | 95.1 | 15 |
| 本文L-mobilenet v2(AM-softmax+Triplet) | 2.83 | 0.67 | 98.56 | 95 | 15 |

从表2可知, L-mobilenet v2模型参数量67万,模型存储尺寸2.83MB,在LFW测试集准确率为98.56%,比Inception-resnet-v1模型准确率99.05%略有下降,但模型参数量和存储尺寸巨幅减少;相比ResNet50,本文模型在存储尺寸、参数量、识别准确率以及Jetson Nano的识别帧率上具有明显优势;2020年文献[24]提出了一种轻量化的Lightfacenet网络,训练集采用MS-Celeb-1M^[25]数据集,该数据集人物ID(10万)是本文训练集的10倍,图片数量(1000万)是本文训练集的20倍,训练迭代次数(15000次)约是本文迭代次数的500倍,在LFW测试集准确率仅高于本文0.76%,模型参数量却大于本文64%,说明本文模型在基本不损失识别精度的前提下,更便于在嵌入式端进行部署。

如前文表1中,本文对mobilenet v2网络结构

实验(2)、(3)的结果从a、b或c、d或e、f折线可以看出,改进后的联合损失函数优于传统的损失函数,测试集LFW模型准确率分别提高了0.11%、0.46%和0.88%。同时,从第1代开始c、d模型准确率就已经超过92%,而采用原始mobilenet v2训练的a、b模型第1代准确率还不到85%,且训练准确率波动相对较大,平稳性不如本文模型。

为了综合比较各模型性能,表2给出了6种模型的存储尺寸、参数量、LFW测试集和自制测试数据集上的识别准确率、NVIDIA嵌入式平台Jetson Nano上单张人脸识别的速率。同时加入谷歌基于Inception-resnet-v1网络的FaceNet模型,该模型使用和本文同样的输入尺寸160像素 \times 160像素,同样的训练集CASIA-WebFace进行训练。

进行了优化,经实验将原倒置残差结构中的维度扩展因子由6下降为2,使得模型存储尺寸缩减了70.7%,参数量缩减了72.3%,但模型精度有所损失,故采用改进后的联合损失使模型精度得以回升。相比于改进前的模型,本文模型在LFW识别准确率提高1.56%,自制测试数据集上提高了7.1%,在Jetson Nano上的单张人脸识别帧率提高了36.3%,充分说明本文模型网络结构以及改进后的损失函数的优越性。

表2中各模型在自制数据集上的识别准确率低与LFW测试集,原因在于本文所用的训练集多为欧美人种,缺乏亚洲人脸数据,所以训练出的模型在LFW测试集表现优于自制的数据集。图8是利用本文模型对一张合成的多人人脸图片在嵌入式设备Jetson nano平台上进行识别的结果。

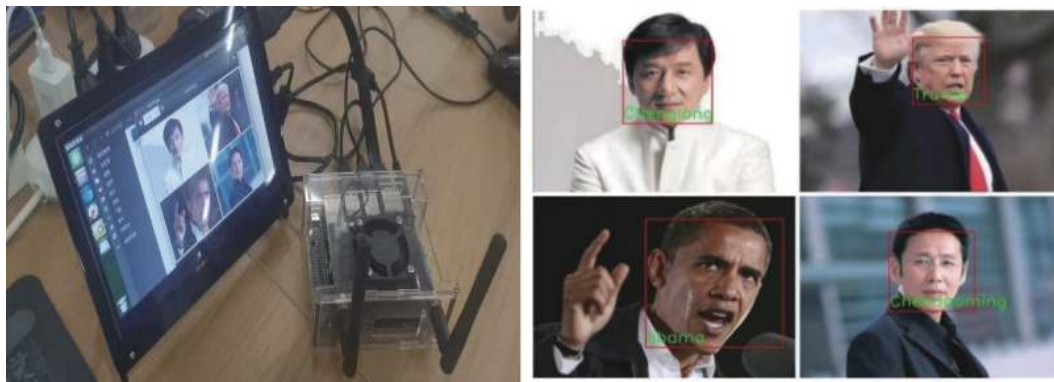


图 8 L-mobilenet v2 在 Jetson Nano 的人脸识别结果

Fig. 8 Face recognition result of L-mobilenet v2 in Jetson Nano

3 结束语

本文提出了一种基于改进 mobilenet v2 的人脸识别算法, 将原模型参数量减小 72.3% 的前提下, 在 LFW 测试集识别准确率提高了 1.56%, 自制数据集准确率提高了 7.1%, 加速了人脸识别算法前向推理速度, 可以在嵌入式平台 Jetson Nano 上实时运行。实验中还加入谷歌 Inception-resnet-v1 和 ResNet50 网络的对比, 充分说明本文 L-mobilenet v2 是极其轻量化的人脸识别网络, 可以方便的部署在计算资源有限的终端设备上。但由于实验训练集缺乏亚洲人脸数据, 在自制数据集上的表现不佳, 后续可以增加亚洲人脸数据进行训练以增强模型的表现力。

参考文献:

- [1] 李东博, 黄铝文. 重加权稀疏主成分分析算法及其在人脸识别中的应用[J]. 计算机应用, 2020, 40(3): 717-722.
- [2] 徐竞泽, 吴作宏, 徐岩, 等. 融合 PCA、LDA 和 SVM 算法的人脸识别[J]. 计算机工程与应用, 2019, 55(18): 34-37.
- [3] LIU Weiyang, WEN Yandong, YU Zhiding, et al. SphereFace: Deep Hypersphere Embedding for Face Recognition[C]//Proceedings of the 2017 IEEE conference on computer vision and pattern recognition. Piscataway: IEEE, 2017: 212-220.
- [4] WANG Hao, WANG Yitong, ZHOU Zheng, et al. CosFace: large margin cosine loss for deep face recognition[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 5265-5274.
- [5] DENG Jiankang, GUO Jia, YANG Jing, et al. Arcface: Additive angular margin loss for deep face recognition[J]. IEEE transactions on pattern analysis and machine intelligence, 2021, 44(10): 5962-5979.
- [6] SUN Yifan, CHENG Changmao, ZHANG Yuhan, et al. Circle loss: a unified perspective of pair similarity optimization[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 6397-6406.
- [7] 才华, 孙俊, 朱瑞昆, 等. 基于自适应圆边界的深度人脸识别算法[J]. 兵工学报, 2021, 42(11): 2424-2432.
- [8] CAI Hua, SUN Jun, ZHU Ruikun, et al. Depth face recognition algorithm based on adaptive circle margin[J]. Acta armamentarii, 2021, 42(11): 2424-2432.
- [9] HOWARD A G, ZHU MENGLONG, CHEN BO, et al. MobileNets: efficient convolutional neural networks for mobile vision applications[EB/OL]. (2017-04-17) [2021-11-29]. <https://arxiv.org/abs/1704.0486>.
- [10] ZHANG Xiangyu, ZHOU Xinyu, LIN Mengxiao, et al. ShuffleNet: an extremely efficient convolutional neural network for mobile devices[EB/OL]. (2017-07-04) [2021-11-29]. <https://arxiv.org/abs/1707.01083>.
- [11] TAN Mingxing, LE Q V. EfficientNet: rethinking model scaling for convolutional neural networks[EB/OL]. (2019-05-28) [2021-11-29]. <https://arxiv.org/abs/1905.11946>.
- [12] SANDLER M, HOWARD A, ZHU Menglong, et al. MobileNetV2: inverted residuals and linear bottlenecks[EB/OL]. (2018-01-13) [2021-11-29]. <https://arxiv.org/abs/1801.04381>.
- [13] GIRSHICK R. Fast R-CNN[C]//2015 IEEE International Conference on Computer Vision. Santiago: IEEE, 2016: 1440-1448.

- [13] LIU Wei, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox detector[C]//European Conference on Computer Vision. Cham: Springer, 2016: 21–37.
- [14] DENG Jiankang, GUO Jia, ZHOU Yuxian, et al. RetinaFace: single-stage dense face localisation in the wild [EB/OL]. (2019–03–02) [2021–11–29]. <https://arxiv.org/abs/1905.00641>.
- [15] ZHANG Kaipeng, ZHANG Zhanpeng, LI Zhifeng, et al. Joint face detection and alignment using multitask cascaded convolutional networks[J]. *IEEE signal processing letters*, 2016, 23(10): 1499–1503.
- [16] SCHROFF F, KALENICHENKO D, PHILBIN J. FaceNet: a unified embedding for face recognition and clustering[EB/OL]. (2015–03–12)[2021–11–29]. <https://arxiv.org/abs/1503.03832>.
- [17] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition[EB/OL]. (2015–12–10) [2021–11–29]. <https://arxiv.org/abs/1512.03385>.
- [18] WEN Yandong, ZHANG Kaipeng, LI Zhifeng, et al. A discriminative feature learning approach for deep face recognition[C]//European Conference on Computer Vision. Cham: Springer, 2016: 499–515.
- [19] WANG Feng, LIU Weiyang, LIU Haijun, et al. Additive margin softmax for face verification[EB/OL]. (2018–01–17) [2021–11–29]. <https://arxiv.org/abs/1801.05599>.
- [20] YI Dong, LEI Zhen, LIAO Shengcai, et al. Learning face representation from scratch[EB/OL]. (2014–11–28) [2021–11–29]. <https://arxiv.org/abs/1411.7923>.
- [21] HUANG G B, LEARNED-MILLER E. Labeled faces in the wild: updates and new reporting procedures[R]. Amherst: University of Massachusetts, 2014.
- [22] HUANG G B, MATTAR M, BERG T, et al. Labeled faces in the wild: a database for studying face recognition in unconstrained environments[R]. Amherst: University of Massachusetts, 2007.
- [23] KINGMA D P, BA J. Adam: a method for stochastic optimization[EB/OL]. (2014–12–22) [2021–11–29]. <https://arxiv.org/abs/1412.6980>.
- [24] 张典, 汪海涛, 姜瑛, 等. 基于轻量级网络的实时人脸识别算法研究 [J]. *计算机科学与探索*, 2020, 14(2): 317–324.
- ZHANG Dian, WANG Haitao, JIANG Ying, et al. Research on real-time face recognition algorithm based on lightweight network[J]. *Journal of frontiers of computer science and technology*, 2020, 14(2): 317–324.
- [25] GUO Yandong, ZHANG Lei, HU Yuxiao, et al. MS-celeb-1M: a dataset and benchmark for large-scale face recognition[EB/OL]. (2016–07–27)[2021–11–29]. <https://arxiv.org/abs/1607.08221>.

作者简介:



屈东东, 硕士研究生, 主要研究方向为计算机视觉、深度学习。



贺利乐, 教授, 博士生导师, 主要研究方向为机器人智能化技术、机器学习。2015 年获陕西省高等学校科学技术奖二等奖, 2016 年获陕西省科学技术奖三等奖。获发明专利授权 5 项, 出版专著 1 部, 教材 4 部, 发表学术论文 86 篇。



何林, 讲师, 主要研究方向为深度学习。