

基于轻量级卷积神经网络人脸识别算法的研究与应用*

胡佳玲, 施一萍, 谢思雅, 陈 藩, 刘 瑾
(上海工程技术大学 电子电气工程学院, 上海 201620)

摘 要: 针对多数卷积神经网络(CNN)人脸识别算法在追求高精度的同时, 加大网络层数, 造成网络参数过多、计算量过大、训练速度缓慢且对设备要求高, 增加了成本的问题, 提出了一种改进型轻量级卷积神经网络 MobileNet 的人脸识别算法。首先, 将 MobileNet 中的 SoftMax 层为 L-SoftMax 层, 避免了过度拟合, 实现更好的分类效果。其次, 将改进的 MobileNet 和区域生成网络(RPN)融合, 并在 Jetson Nano 小型设备上训练。实验表明: 所提算法与传统的卷积神经网络人脸识别算法相比, 在 LFW 人脸数据库和自建的小型数据库上训练测试, 模型的参数量减少了 88%, 识别准确率与原 MobileNet 相比增加了 0.2%, 达到了 97.54%。运行速度较原 MobileNet 网络提高了 21.3%。

关键词: 人脸识别; 卷积神经网络; MobileNet 模型; L-SoftMax 层; 区域生成网络(RPN)

中图分类号: TP391.4 **文献标识码:** A **文章编号:** 1000-9787(2022)01-0153-04

Research and application of face recognition algorithm based on lightweight CNN*

HU Jialing, SHI Yiping, XIE Siya, CHEN Fan, LIU Jin

(School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201620, China)

Abstract: Aiming at the problem that many convolutional neural network (CNN) face recognition algorithms pursue high precision, at the same time, increase the number of network layers, resulting in excessive network parameters, too much calculation, slow training speed and high equipment requirements, which increases the cost, an improved lightweight CNN face recognition algorithm is proposed. It is mainly implemented by Mobilenet. Firstly, the SoftMax layer in MobileNet is replaced by L-SoftMax layer. L-SoftMax avoids over fitting and has better classification effect. Secondly, the MobileNet will be integrated with the region proposal network (RPN). Experimental results show that, compared with the traditional CNN face recognition algorithm, the proposed algorithm can reduce the number of parameters by 88% and increase the recognition accuracy rate by 0.2% compared with the original MobileNet algorithm, reaching 97.54% after training and testing on LFW face database and its own small database. Compared with the original MobileNet, the running speed is increased by 21.3%.

Keywords: face recognition; convolutional neural network (CNN); MobileNet model; L-SoftMax layer; region proposal network (RPN)

0 引 言

近年来, 人脸识别^[1]一直是很多学者们研究的热点问题。人脸识别算法目前主要分为两种: 一种是机器学习算法, 本质主要是通过设定人脸特征再结合 Adaboost 等分类算法加以实现; 第二种是近年来比较流行的结合深度学习^[2]的人脸识别算法。基于深度学习的人脸识别算法一般采用卷积神经网络(convolution neural network, CNN)来实现。发展至今, 基于 CNN^[3]的人脸识别算法在准确率上已经趋于 100%, 并且有越来越多效果很好的网络相继出现。2014 年, 香港中文大学的团队利用深度学习人脸识别

算法取得了准确率为 97.45% 的好成绩; 准确率的提升往往伴随着 CNN 深度和复杂度的增加, 在提高精度的同时, 意味着网络深度、参数和计算量的大幅增加, 训练时间也相对延长。例如 Google Inception Net 有 22 层, 且 ResNet 达到了 152 层。其次, 随着小型设备的应用不断广泛, 越来越多的研究不仅仅局限于靠强大的服务器去训练网络。

针对上述问题, 本文选用了一种轻量级 MobileNet^[4]模型, 将改进的 MobileNet 网络和区域生成网络(region proposal network, RPN)^[5]融合构成一个用于人脸识别的专用

收稿日期: 2020-07-08

* 基金项目: 国家自然科学基金资助项目(61701296)

网络,并对其进行了改进,有效地提高了网络的运行速度,降低了网络的参数量。最后用 Jetson Nano 嵌入式开发板作为训练的载体,测试网络的具体性能,并应用到实际场景中。

1 人脸识别系统

如图1所示,为本文设计构建的人脸识别模型,系统的硬件采用嵌入式开发板 Jetson Nano,将搭建好的改进型轻量级人脸识别网络在开发板上训练好,并投入使用。

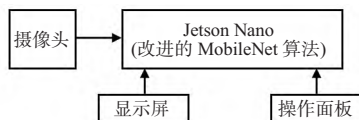


图1 人脸识别系统模型

2 改进型 MobileNet 网络

本文设计的模型具体从两个方面进行改进:首先,将 MobileNet 网络中 SoftMax^[6] 层改用 L-SoftMax^[7] 层替换;其次,网络融合入 RPN,提高网络的识别效率。

2.1 MobileNet 网络

2017年,Google 提出 MobileNet 深度学习网络,与传统卷积神经网络不同在于,MobileNet 网络采用深度可分离卷积^[8] (depthwise separable convolution),即把传统卷积分为深度卷积 (depthwise convolution) 和逐点卷积 (pointwise convolution)。深度可分离卷积可以在保证网络不损失大量精确度的情况下大大降低网络参数和计算量。卷积的具体分解过程如图2所示。

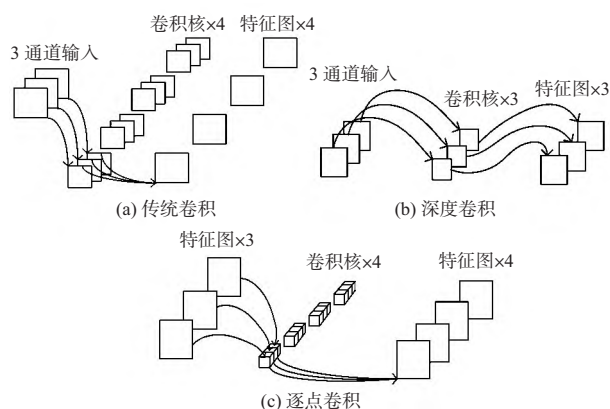


图2 传统卷积、深度卷积和逐点卷积

如图2(a)传统卷积所示:卷积核 channel=输入特征矩阵 channel; 输出特征矩阵 channel=卷积核个数。而图2(b)中的深度卷积:卷积核 channel=1; 输入特征矩阵 channel=卷积核个数=输出特征矩阵 channel。

1×1 的卷积核在深度可分离卷积中被称为逐点卷积。特征映射 F 的大小为 (D_F, D_F, M) , K 为 (D_K, D_K, M, N) , 如图2(a)传统卷积所示。其输出的特征映射 G 大小为 (D_G, D_G, N) 。传统卷积的计算公式如下

$$G_{k,l,n} = \sum_{i,j,m} K_{i,j,m,n} \cdot F_{k+i-1,l+j-1,m} \quad (1)$$

传统卷积的计算量为 $D_K \cdot D_K \cdot M \cdot N \cdot D_F \cdot D_F$ 。现将传统卷积 (D_K, D_K, M, N) 进行拆分,拆分为深度卷积和逐点卷积:

1) 滤波由深度卷积负责,尺寸为 $(D_K, D_K, 1, M)$, 如图2(b)的深度卷积所示,输出的特征为 (D_G, D_G, M) 。

2) 转换通道由逐点卷积负责,尺寸为 $(1, 1, M, N)$, 如图2(c)的逐点卷积所示,得到最终的输出为 (D_G, D_G, M) 。深度卷积的卷积公式为

$$\hat{G}_{k,l,n} = \sum_{i,j} \hat{K}_{i,j,m} \cdot F_{k+i-1,l+j-1,m} \quad (2)$$

式中 \hat{K} 为深度卷积,其卷积核为 $(D_K, D_K, 1, M)$ 。在上述卷积核中有约 m_{th} 个卷积核运用在 F 中第 m_{th} 个通道上,将产生 \hat{G} 上第 m_{th} 个通道输出。因此,卷积的计算量为

$$D_K \cdot D_K \cdot M \cdot D_F \cdot D_F + M \cdot N \cdot D_F \cdot D_F \quad (3)$$

因此,深度可分离卷积与传统卷积比值为

$$\frac{D_K \cdot D_K \cdot M \cdot D_F \cdot D_F + M \cdot N \cdot D_F \cdot D_F}{D_K \cdot D_K \cdot M \cdot N \cdot D_F \cdot D_F} = \frac{1}{N} + \frac{1}{D_K^2} \quad (4)$$

即计算量为传统卷积的 $1/N + 1/D_K^2$ 。

一般情况下, N 的取值较大。假设用的卷积核为 3×3 大小,则这里计算量大约会减少9倍左右,训练网络的时间也大大缩短。

2.2 网络层优化

虽然,MobileNet 的计算量比传统卷积神经网络高许多倍,但其检测准确率低于层数较多的卷积神经网络。MobileNet 在训练时使用 SoftMax 分类器。但研究发现采用 L-SoftMax 的分类效果比 SoftMax 更好。最初,人脸识别采用 SoftMax Loss 方法,但实验发现,这种方式只能确保让其组与组间具有区分性,而对组内的约束是很小的。所以,鉴于此,L-SoftMax 增加了一个代表分类间隔的超参数 m ,它决定了接近地面真值类的强度,产生一个角边距,使得学习特征变得更加紧凑和良好分离。

SoftMax Loss 函数公式如式(5)所示

$$L = \frac{1}{N} \sum_i L_i = \frac{1}{N} \sum_i -\log(s_j) \quad (5)$$

$$s_j = \frac{e^{f_{ji}}}{\sum_j e^{f_{ji}}}$$

式中 s_j 为 SoftMax 函数,代表属于各个类别的概率。

而 L-SoftMax Loss 的公式如下

$$L_i = -\log \left(\frac{e^{\|W_{y_i}\| \|x_i\| \psi(\theta_{y_i})}}{e^{\|W_{y_i}\| \|x_i\| \psi(\theta_{y_i})} + \sum_{j \neq y_i} e^{\|W_j\| \|x_i\| \cos(\theta_j)}} \right) \quad (6)$$

其中, $\psi(\theta) = \begin{cases} \cos(m\theta), 0 \leq \theta \leq \frac{\pi}{m} \\ D(\theta), \frac{\pi}{m} \leq \theta \leq \pi \end{cases}$

式中 $D(\theta)$ 要求是单调递减函数, $D(\pi/m)$ 应等于 $\cos(\pi/m)$ 。 m 是一个与分类裕度密切相关的整数, 随着 m 的增大, 分类裕度变大, 学习目标也变难。而这种方式, 会迫使模型学习到类间距离更大, 而类内距离更小的特征。因此, 会使得分类效果更清晰。这一点在迭代次数多的情况下效果会更明显。表 1 为改进后的网络结构。

表 1 改进型 MobileNet 网络结构

类型/步距	卷积核尺寸	输入
Conv/s2	3 × 3 × 32	224 × 224 × 3
Conv dw/s1	3 × 3 × 32 dw	112 × 112 × 32
Conv/s1	1 × 1 × 32 × 64	112 × 112 × 32
Conv dw/s2	3 × 3 × 64 dw	112 × 112 × 64
Conv/s1	1 × 1 × 64 × 128	56 × 56 × 64
Conv dw/s1	3 × 3 × 128 dw	56 × 56 × 128
Conv/s1	1 × 1 × 128 × 128	56 × 56 × 128
Conv dw/s2	3 × 3 × 256 dw	56 × 56 × 128
Conv/s1	1 × 1 × 128 × 256	28 × 28 × 128
Conv dw/s1	3 × 3 × 256 dw	28 × 28 × 256
Conv/s1	1 × 1 × 256 × 256	28 × 28 × 256
Conv dw/s2	3 × 3 × 256 dw	28 × 28 × 256
Conv/s1	1 × 1 × 256 × 512	14 × 14 × 256
5x	Conv dw/s1	3 × 3 × 512
	Conv/s1	1 × 1 × 512 × 512
	Conv dw/s2	3 × 3 × 512 dw
	Conv/s1	1 × 1 × 512 × 1024
	Conv dw/s2	3 × 3 × 1024 dw
Conv/s1	1 × 1 × 1024 × 1024	7 × 7 × 1024
Avg Pool/s1	Pool 7 × 7	7 × 7 × 1024
FC/s1	1024 × 1000	1 × 1 × 1024
L-SoftMax	Classifier	1 × 1 × 1000

3 改进型 MobileNet + RPN 识别算法

3.1 RPN

为了缩小与深层卷积神经网络在识别率上的差距, 本文还加入了 RPN, 用来确定人脸检测窗口。

RPN 原理上是在卷积神经网络的卷积层后生成的特征图上产生建议选区。因其巧妙地采用了 Anchor 机制, 所以, 可以产生多尺度长宽比的建议选区。RPN 网络的具体实现过程如下: 首先, 用一个小网络在特征图上进行滑动, 一般的滑动窗口设置为 3 × 3 大小。这个滑动窗口会做全卷积操作, 并根据最后卷积层的通道数, 生成一个对应数目的向量。其次, 将向量连向分类层, 分类层将判断该区域是场景中的哪部分, 并给出具体的分数。原 Fast R-CNN^[9]一

共设置了 9 种 Anchor。由于人脸的特殊性, 一般人脸的比例是固定的。本文对 LFW(labeled faces in the wild) 中人脸比例情况做了粗略统计, 结果表明, 人脸的长宽比基本都是 1:1 和 2:1, 大部分在 1.5 左右。所以, 本文的 Anchor 比例设置为 1:1 和 2:1, 尺寸和原 RPN 一样, 都采用 3 种尺寸, 所以, 一个 3 × 3 的区域只需要产生 6 个 Anchor, 达到了减少计算量的目的, 同时, 加入的 RPN 也提升了本文算法的识别率。

3.2 改进后的模型

如图 3, 第一部分为 MobileNet 的卷积层, 第二部分为加入的 RPN, 第三部分为全卷积层后对原 SoftMax 改进为 L-SoftMax。改进后的算法为 MobileNet-L + RPN 算法。

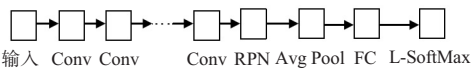


图 3 整体网络结构

4 实验结果与分析

4.1 实验数据集

本次实验采用的人脸数据库为 LFW, 其中的图片均来自网络, 共包含 13 233 张人脸图片, 5 749 位名人, 每位名人有一张或多张人脸图片且都被标注了姓名和编号。

4.1.1 数据集分类情况

针对 LFW 数据集, 选取 10 000 张用于训练集, 3 000 张用于验证集。除此外, 本文还准备了 1 000 张测试集, 测试集来源于网络人脸和现实拍照, 包括身边的同学、朋友等。

4.1.2 图像预处理

1) 灰度化^[10]: 为使后续的图像计算量变少, 本文预先对数据库进行统一的灰度化处理, 采用的是传统的灰度化处理方法。

2) 直方图均衡化^[11]: 增强灰度图像的对比度, 减少不相关因素的影响。

4.2 实验结果及数据分析

Jetson Nano 支持的系统为 Ubuntu 系统, 本次安装的版本为较新的 Ubuntu 18.04.2 LTS。软件部分采用 Python 3.6 版本下的 Anaconda, 采用 Pytorch 框架。网络训练参数的设置如下: 初始学习率为 0.001, 学习策略采用按需调整学习率, 权重衰减为 0.0004。

本文比较了 3 种网络模型: CNN, MobileNet, MobileNet-L + RPN。首先在带有 GPU 服务器的电脑设备上, 进行模型的评价, 实验数据如图 4 所示: 三种模型迭代 15 000 次的准确率折线图, 从图中可以看出: 采用 CNN 在迭代 4 000 ~ 5 000 次时, 准确率趋于收敛, 准确率最终稳定在 98 % 左右。MobileNet 在迭代 10 000 左右趋于收敛, 准确率稳定在

95%左右。采用 MobileNet (L-SoftMax) + RPN 模型在迭代 8000 次左右趋于收敛,准确率最终稳定在 97%左右。对比传统的 CNN,准确率还是有所下降。但比原 MobileNet 网络,准确率提高了 2 个百分点。因此,本文提出的改进 MobileNet 网络切实可行,且在准确率上较原 MobileNet 网络有提升。虽然没有达到和传统 CNN 一样或者更高的精确度,但这对于大大提升计算速度和大大降低网络参数而言,在性能上有了很大的提升。

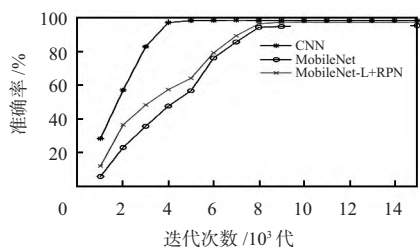


图4 三种模型准确率比较折线图

三个模型试验结果对比如表2所示,由表2可以看出,本文提出的模型在降低网络参数和训练时间的同时,识别的准确率较原 MobileNet 网络也有提升。

表2 三种模型参数量及识别准确率比较

网络结构	参数量	准确率/%
CNN	25 000 000	98.90
MobileNet	3 600 000	95.53
MobileNet-L+RPN	3 000 000	97.54

其次,本文又将原 MobileNet 模型和改进后的轻量级模型分别在 Jetson Nano 嵌入式平台上进行测试,最终得到的数据如表3所示。由表3可以看出,本文提出的模型在运行速度上比原 MobileNet 快了 21.3%左右。

表3 模型在 Jetson Nano 上运行速度对比 ms

网络结构	最小 (Min)	最大 (Max)	平均 (Avg)
MobileNet	130.39	141.67	136.03
MobileNet-L+RPN	100.63	113.54	107.09

5 结束语

本文提出了一种轻量级人脸识别算法,对 MobileNet 网络进行网络层的改进,并融合 RPN,在提升网络识别速度的同时,网络的准确率也得到提升,使得网络能够在小型计

算平台上得以应用(本文以 Jetson Nano 为例)。实验结果表明:在 LFW 数据库和自建的数据库上训练的人脸识别准确率达到 97.54%。同时,本文将改进的 MobileNet 人脸识别运用到具体的 Jetson Nano 设备上,可构成完整的识别系统,运用于多个现实场景。如室外安防、打卡系统、查寝系统等。

参考文献:

- [1] 薛建明,刘宏哲,袁家政,等.基于 CNN 与关键区域特征的人脸表情识别算法[J].传感器与微系统,2019,38(10):146-149,153.
- [2] 周刊,胡士强,吴桐.基于深度度量学习的行人重识别方法[J].传感器与微系统,2020,39(5):61-64.
- [3] 曹璐,陈明,秦玉芳.基于多通道稀疏卷积神经网络的行人检测算法[J].传感器与微系统,2018,37(10):130-132,136.
- [4] 汤高扬.基于 MobileNet 的目标跟踪算法研究与实现[D].广州:广东工业大学,2019.
- [5] 许超.基于改进 CNN 的人脸实名登记系统的研究与设计[D].武汉:湖北工业大学,2018.
- [6] 万磊,佟鑫,盛明伟,等. SoftMax 分类器深度学习图像分类方法应用综述[J].导航与控制,2019,18(6):1-9,47.
- [7] 姜轩.基于深度学习的场景图像分类算法研究[D].北京:北京邮电大学,2019.
- [8] 王丹峰,陈超波,马天力,等.基于深度可分离卷积的 YOLOv3 行人检测算法[J].计算机应用与软件,2020,37(6):218-223.
- [9] 姚万业,李金平.基于改进 Faster R-CNN 的行人检测算法[J].科学技术与工程,2020,20(4):1498-1503.
- [10] 顾梅花,苏彬彬,王苗苗,等.彩色图像灰度化算法综述[J].计算机应用研究,2019,36(5):1286-1292.
- [11] 陈永亮.灰度图像的直方图均衡化处理研究[D].合肥:安徽大学,2014.

作者简介:

胡佳玲(1995-),女,硕士研究生,研究方向为深度学习和图像处理。

施一萍(1964-),女,通讯作者,副教授,硕士研究生导师,主要研究领域为深度学习和智能控制等。

(上接第 149 页)

- [14] ALIPPI C, DISABATO S, ROVERI M, et al. Moving convolutional neural networks to embedded systems: The alexnet and VGG-16 case[C]// Information Processing in Sensor Networks, 2018: 212-223.
- [15] HOWARD A G, ZHU M, CHEN B, et al. MobileNets: Efficient convolutional neural networks for mobile vision applications[EB/

OL]. (2017-04-17). [2019-05-06]. <https://arxiv.org/pdf/1704.04861.pdf>.

作者简介:

赵九霄(1994-),男,硕士研究生,研究方向为图像处理。

刘毅(1969-),男,博士,教授,硕士研究生导师,研究领域为智能信息处理。

李国燕(1984-),女,博士,讲师,研究领域为下一代网络技术。