

多任务卷积神经网络的电梯乘客识别方法^①



王廷银^{1,2,3}, 郭威¹, 吴允平^{1,2,3}

¹(福建师范大学 光电与信息工程学院, 福州 350007)

²(数字福建环境监测物联网实验室, 福州 350117)

³(福建省光电传感应用工程技术研究中心, 福州 350007)

通讯作者: 吴允平, E-mail: wyp@fjnu.edu.cn

摘要: 电梯安全监测系统应用中, 对于电梯乘客识别往往采用红外传感技术或是传统人脸检测算法如 Haar-like、HOG 实现, 但应用效果并非很理想. 近年来随着深度学习的发展, 基于卷积神经网络的人脸检测算法在精度上高于传统人脸检测算法, 被多个领域应用. 基于多任务级联卷积神经网络人脸检测算法模型小、运算快的特点而将其应用到电梯安全监测系统, 通过引入 Inception 模块思想, 利用不同大小卷积核并行操作增加各级网络的深度和宽度, 提升网络特征提取能力, 结合 Batch Normalization 算法提高模型训练速度和网络的分类能力. 实验结果表明, 改进后算法的精度比原算法提升了 2%, 实现高准确率的电梯乘客识别.

关键词: 卷积神经网络; 电梯安全监测; 乘客识别; Inception 模块; 批量归一化

引用格式: 王廷银, 郭威, 吴允平. 多任务卷积神经网络的电梯乘客识别方法. 计算机系统应用, 2021, 30(6): 278-285. <http://www.c-s-a.org.cn/1003-3254/7975.html>

Elevator Passenger Identification Method Based on Multi-Task Convolutional Neural Network

WANG Ting-Yin^{1,2,3}, GUO Wei¹, WU Yun-Ping^{1,2,3}

¹(College of Photonic and Electronic Engineering, Fujian Normal University, Fuzhou 350007, China)

²(Digit Fujian Internet-of-Things Laboratory of Environmental Monitoring, Fuzhou 350117, China)

³(Fujian Provincial Engineering Technology Research Center of Photoelectric Sensing Application, Fuzhou 350007, China)

Abstract: In the application of safety monitoring system of elevators, infrared sensor technology or traditional face detection algorithms involving Haar-like and HOG features are often used for the recognition of elevator passengers with poor effect though. With the development of deep learning in recent years, the face detection algorithm based on convolutional neural networks is more accurate than traditional face detection algorithms and has been applied in many fields. Moreover, the face detection algorithm based on multi-task cascaded convolutional neural networks is adopted to recognize elevator passengers in the safety monitoring system owing to its small model and fast operation. With the inception module introduced, the depth and width of networks at all levels are raised by the parallel operation of convolutional cores of different sizes for better extraction of network features; models are trained faster and network classification is enhanced through batch normalization. The experimental results show that the accuracy of the improved algorithm is 2% higher than that of the original one and can thus realize the highly accurate recognition of elevator passengers.

Key words: Convolutional Neural Network (CNN); elevator safety monitoring; passenger recognition; inception module; batch normalization

① 基金项目: 国家自然科学基金-海峡联合基金重点项目 (U1805263); 福建省引导性项目 (2019H0009, 2020H0011); 福建省自然科学基金 (2019J01427)

Foundation item: Key Project of Joint Fund for the Promotion of Scientific Cooperation across the Taiwan Straits of National Natural Science Foundation of China (U1805263); Guidance Project of Fujian Province (2019H0009, 2020H0011); Natural Science Foundation of Fujian Province (2019J01427)

收稿时间: 2020-10-10; 修改时间: 2020-11-02; 采用时间: 2020-12-02; csa 在线出版时间: 2021-06-01

1 引言

电梯是现代人类出行活动最后 50 米的交通工具^[1], 它方便人们的生活, 使得人们对电梯使用的需求也日益剧增. 截至 2019 年, 我国电梯的保有量已达到 709 万台, 位居世界第一, 但其中也包括大量的老旧电梯. 电梯作为特种设备的一种, 其信息化水平与乘梯的安全性直接挂钩, 由于老旧电梯的信息化水平普遍不足以及我国电梯的使用呈现负载高、运量大、随机性强等特点, 使得电梯事故时有发生^[2]. 当遇到电梯事故时, 如何准确了解到轿厢人员情况, 是电梯安全监测系统研究工作中重要的一环, 它有助于快速展开救援工作避免不必要的生命财产损失. 有的学者在电梯监测系统中应用红外传感器来检测识别电梯事故时是否有乘客^[3], 但由于红外传感器易受噪声影响而导致误检的情况发生^[4]. 为此, 一些学者们通过机器视觉中传统人脸检测算法如 Haar-like^[5]、HOG^[6] 识别电梯乘客, 避免红外传感易误检的问题, 例如, 2015 年, 王潇宇提出 HOG 人脸检测算法应用于电梯乘客识别, 该方法在简单实验环境下有较好的效果, 但实际应用情况不太理想^[7]; 2016 年, Kumar 利用 Haar-like 人脸检测算法进行轿厢人员的检测, 但是利用传统算法所提取的人脸特征难以充分反映真实的人脸特征, 导致在实际应用中检测的精度不高^[8].

这些年, 随着深度学习的发展^[9], 基于卷积神经网络 (Convolutional Neural Networks, CNN) 的模型层出不穷, 如 Faster R-CNN (Faster Region- Convolutional Neural Networks)、SSD (Single Shot multibox Detector)、YOLO (You Only Look Once)、MTCNN (Multi-Task Convolutional Neural Network), 它们在人脸检测的精度上比传统算法高, 从而被应用到无人机、地铁、安全监控多个领域中, 2018 年, Shen 等将 MTCNN 的人脸检测应用到无人机监控系统中, 克服了传统检测算法检测效果不理想的情况, 提高了无人机对人员检测的精度^[10]; 胡玲玲将 MTCNN 和 anchor 机制结合提高网络检测的精度同时引入奇异值分解提高网络在移动端的检测速度^[11]. 2019 年, 周慧娟等应用 YOLO 对地铁乘客进行人脸检测实现进站客流量统计, 避免传统人脸检测算法在实际客流量统计中精度不理想的问题, 有利于站管人员根据客流情况做出相应措施, 保障人流安全^[12]; 同年 9 月, 易诗等针对目前视频监控球机识别、跟踪效果不佳等问题, 运用 SSD 网络人脸检测功能提升对目标识别的精度^[13]; 同年 10 月, Halawa 等面

对停车场的人为监控效果不理想的情况, 在闭路监控中应用 Faster R-CNN 进行人脸检测, 并利用 Inception 结构优化网络, 保障对停车场人员高精度监控^[14].

相较于 Faster R-CNN、YOLO、SSD 等卷积神经网络, 多任务级联卷积神经网络算法 (MTCNN) 具有计算较快、模型较小的优点^[15]. 综上, 将 MTCNN 人脸检测算法应用于电梯乘客识别, 并对其网络结构进行改进, 实现高精度检测, 避免因红外传感容易受噪声影响导致乘客的误检和传统人脸检测算法检测精度低的问题的同时, 还能与电梯运行状态分析相融合, 对电梯运行进行安全评估, 当遇到电梯事故时, 有助于救援工作的开展.

2 MTCNN 原理

2.1 MTCNN 的网络结构

MTCNN 是一种深层次的多任务级联卷积神经网络框架^[16], 由 3 个简单的卷积神经网络级联组成, 通过由粗略到精细的方式检测人脸以及关键点的回归. 检测过程如图 1 所示, 首先对输入的图像进行尺度变换, 得到图像金字塔, 然后把图像金字塔输入到第 1 级网络 P-Net (Proposal Network) 中, 生成多尺度的人脸候选框, 通过边框回归校准候选框, 并采用非极大抑制 (NMS) 合并重叠的候选框, 将候选框图输入到第 2 级网络 R-Net (Refine Network) 中; R-Net 网络使用边框回归和 NMS 来剔除大部分错误的候选框, 将精确后的候选框图输入到第 3 级网络 O-Net (Output Network) 中; O-Net 对候选框进行再次精选得到最终人脸框图以及人脸关键点.

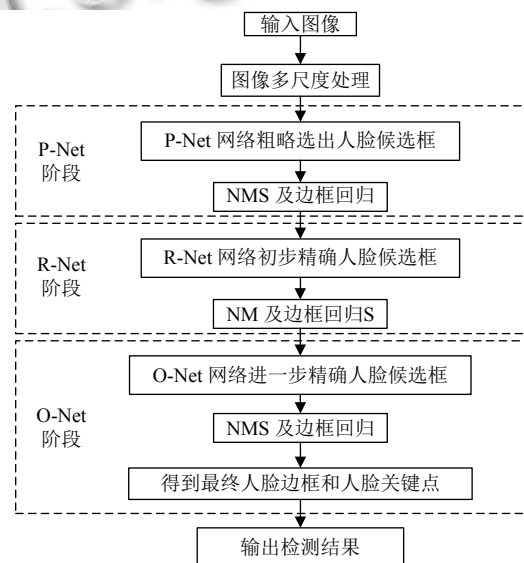


图 1 MTCNN 算法检测图

MTCNN 各级的网络结构图,如图 2 所示.它们从内容上看起来很相似,但是各自的作用还是有所不同的^[17]. P-Net 的网络结构如图 2(a) 所示, P-Net 是 3 个网络中深度最浅的,且是一个全卷积神经网络,它主要的任务是获得大量人脸区域候选框.图 2(b) 是 R-Net 的网络结构图,它的任务是过滤错误的候选框,在结构上比 P-Net 新增一层池化层与全连接层,使得 R-Net 网络

的学习能力更强,进而可以较好地抑制错误的候选框.图 2(c) 是 O-Net 的网络结构图,它比 R-Net 增加一层卷积层与池化层,在结构上更加的复杂,从而更严格地抑制错误的候选框,输出最终人脸候选框和 5 个面部关键点.

2.2 MTCNN 的损失函数

MTCNN 中的每级网络都需要进行人脸分类、边框回归、关键点回归 3 个任务的训练.

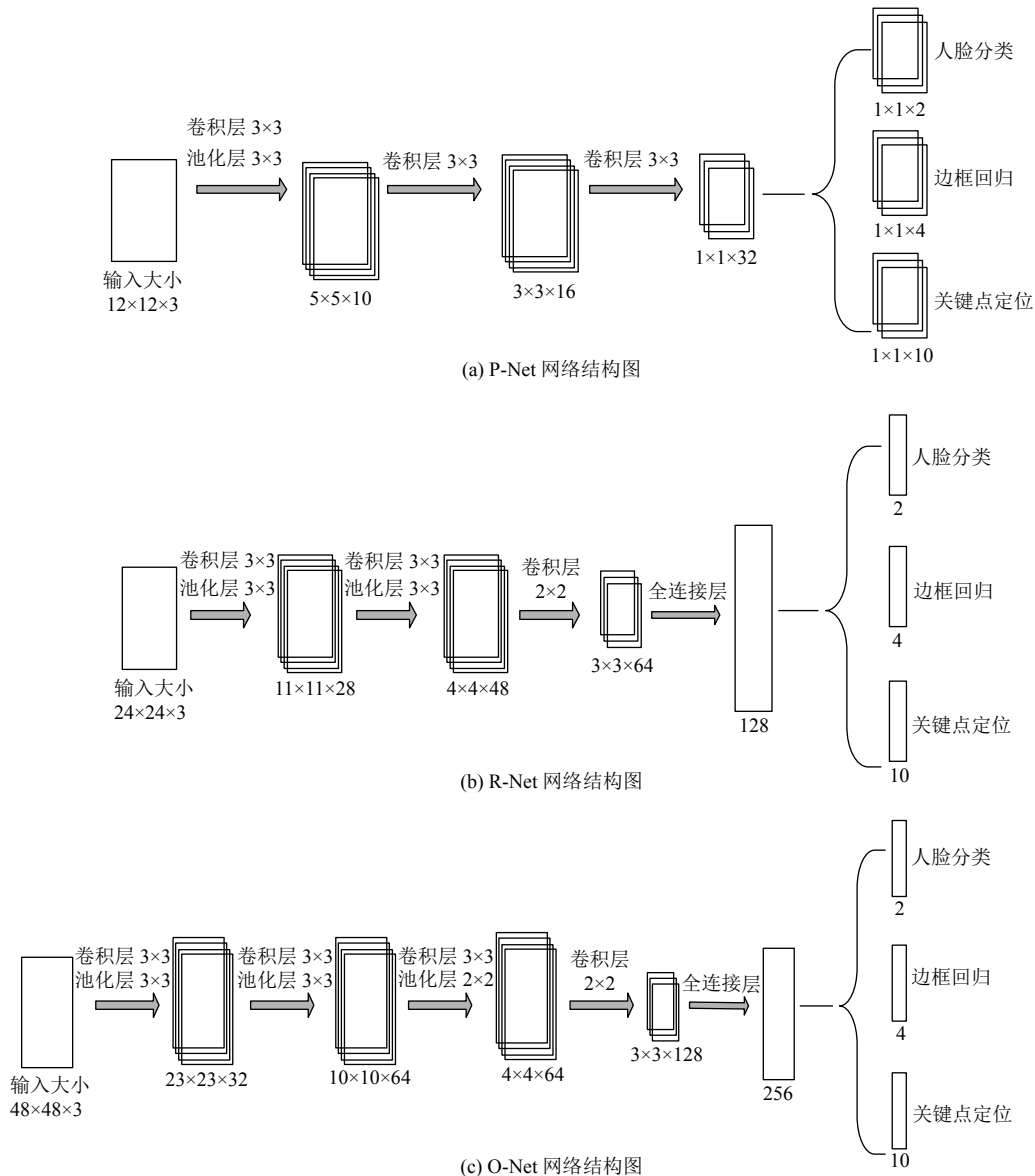


图 2 MTCNN 各级网络结构图

人脸分类属于二分类任务,通过交叉熵损失函数计算每个样本 X_i 的损失值,如式 (1) 所示.

$$L_i^{\text{det}} = -(y_i^{\text{det}} \log(p_i) + (1 - y_i^{\text{det}}) \log(1 - p_i)) \quad (1)$$

其中, p_i 表示网络预测样本属于人脸的概率, $y_i^{\text{det}} \in \{0,1\}$ 属于样本标签, 0 是非人脸, 1 是人脸.

边框回归是回归问题. 边框包含了左上角坐标 (X,Y) 以及边框的宽和高 4 个数值,通过解算网络预测

边框坐标与真实标注边框坐标的欧几里德损失值纠正坐标位置,如式(2)所示。

$$L_i^{\text{box}} = \|\hat{y}_i^{\text{box}} - y_i^{\text{box}}\|_2^2 \quad (2)$$

其中, \hat{y}_i^{box} 表示网络预测的边框坐标, y_i^{box} 表示真实标注的边框坐标。

关键点回归也是属于回归问题。关键点包含左右眼、鼻子、左右嘴角 5 个坐标,共 10 个数值,通过解算网络预测的脸部关键点坐标和真实脸部关键点坐标的欧几里德损失值从而修正坐标位置,如式(3)所示。

$$L_i^{\text{landmark}} = \|\hat{y}_i^{\text{landmark}} - y_i^{\text{landmark}}\|_2^2 \quad (3)$$

其中, $\hat{y}_i^{\text{landmark}}$ 表示网络预测的脸部关键点坐标, y_i^{landmark} 表示真实脸部关键点坐标。

在训练时,各个网络的总损失函数通过上述 3 种损失函数以加权形式表示,如式(4)所示。

$$L = \sum_{i=1}^N \sum_{j \in \{\text{det}, \text{box}, \text{landmark}\}} \alpha_j \beta_i^j L_i^j \quad (4)$$

其中, N 表示训练样本数量, α_j 是各任务权重系数,表示该任务在网络中的重要程度, $\beta_i^j \in \{0, 1\}$ 表示样本类型标签,在计算各个网络的损失值时,根据网络对各个任务的注重程度不同来设置权重系数^[18]。

3 MTCNN 算法改进优化

本文通过引入 Inception 模块和 BN 算法进行 MTCNN 算法的改进优化研究。

3.1 引入 Inception 模块

由于不同大小的卷积核对图片处理能力不同, Szegedy 等因此而设计 Inception 模块并应用在 GoogLeNet 中^[19],其特点是并行使用多个大小不同的卷积核,增强整体网络的特征提取能力,图 3 就是典型的 Inception 模块结构图。由于轿厢人员的人脸在图片中的比例大小不固定,在卷积神经网络中使用固定大小的卷积核会影响人脸检测的准确率,所以将 GoogLeNet 中 Inception 模块的思想与 MTCNN 网络结合,增强网络多尺度特征提取的能力,进而提高人脸检测的准确性。

3.2 引入 BN 层

在训练卷积神经网络的时候,各参数的初始化设置会影响到训练的结果。而 BN (Batch Normalization) 算法可以避免此影响^[20],有效加快收敛过程,同时,BN 算法还具有正则化的作用可增强网络分类效果,所以在 MTCNN 网络中加入 BN 层。BN 算法的传导过程

如图 4 所示,其中①是对数据进行标准正态分布处理,加快收敛;②是进行线性变换处理,恢复数据特征,避免标准正态分布对数据特性以及网络表达能力的影响。

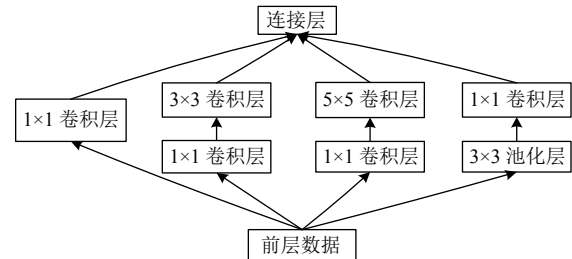


图 3 Inception 模块结构图

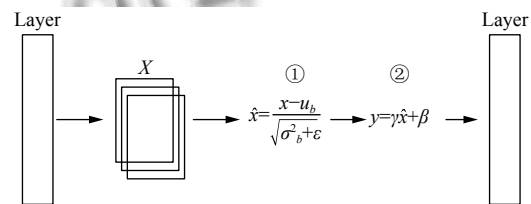


图 4 BN 算法传导过程图

3.3 改进后的 MTCNN 网络结构图

改进后的 MTCNN 网络结构图如图 5 所示,通过并行不同大小的卷积核来对特征图进行处理,并在每个卷积层后面添加 BN 层,增加网络的宽度和深度,同时也增强网络的分类能力。

4 实验与分析

4.1 实验环境

本实验中软件环境为: Python 2.7、VS 2015、OpenCV、Caffe、Android Studio; 硬件环境为: NVIDIA960 显卡的主机和 RK3288 嵌入式开发板。

4.2 实验数据

使用公开的 WINDERFACE 人脸检测数据集进行人脸分类任务和人脸边框回归任务训练。WINDERFACE 人脸数据集包含 32 203 张图片,共 393 703 张人脸,具有各种尺度、姿势、表情、光照等特征,非常匹配于电梯场景的乘客人脸情况,将其中 16 106 张图片(包含 198 485 张人脸)有人脸标注的作为训练集。使用文献 [21] 提供的数据集进行人脸关键点回归任务训练,该数据集有 13 466 张人脸图像,每张人脸都标注 5 个关键点,其中 5590 张来自 LFW 数据集,另外 7876 张来自互联网。使用 Fddb 数据集对网络进行测试评估,该数据集常用于人脸检测算法的验证,共有 2 485 张图片,共 5 171 张人脸。

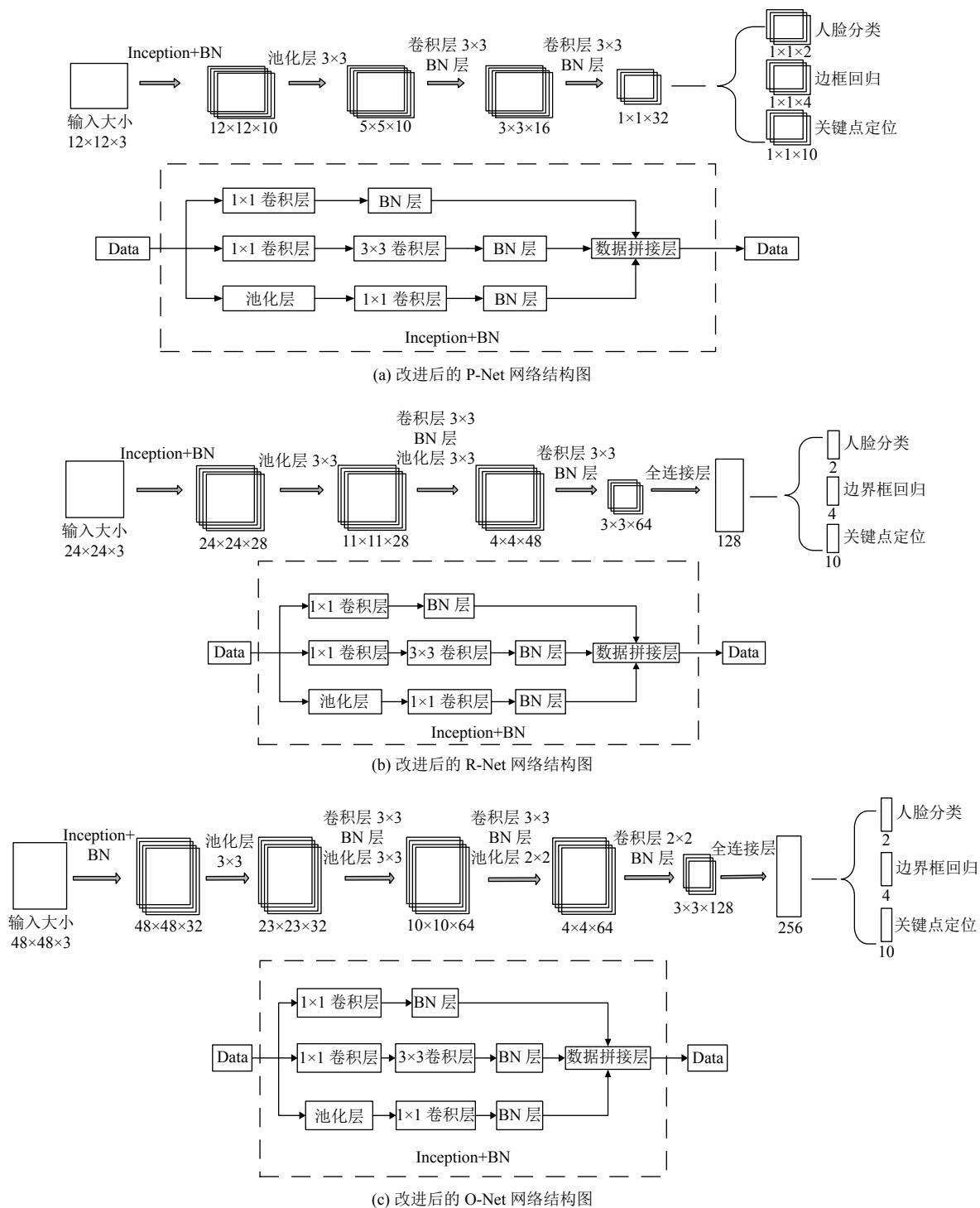


图5 改进后的 MTCNN 网络结构图

4.3 网络训练

将训练的数据集划分成 4 个不同类型的数据集用于对应任务的训练, 分别为正样本集、负样本集、半人脸样本集、关键点样本集。其中, 负样本集: 与真实人脸框的重叠率 IOU 小于 0.3; 正样本集: 与真实人脸

框的重叠率 IOU 大于 0.65; 半人脸样本集: 与真实人脸框的重叠率 IOU 在 0.3 和 0.65 之间; 关键点样本集: 标注了人脸面部 5 个关键点位置。其中, 正样本集和负样本集用于训练人脸分类任务, 正样本集和半人脸样本集用于训练人脸边框回归任务, 关键点样本集用于

训练人脸关键点回归的任务. 在训练时, 由于每个网络对各个任务的注重程度不同, 所以各个网络损失函数的权重系数也有所差异, 其中 P-Net 和 R-Net 中人脸分类的任务更为重要, 所以人脸分类任务、人脸边框回归任务、人脸关键点回归任务三者权重比为 1:0.5:0.5, 而 O-Net 增大了人脸关键点回归的重要程度, 3 个任务的权重比为 1:0.5:1.

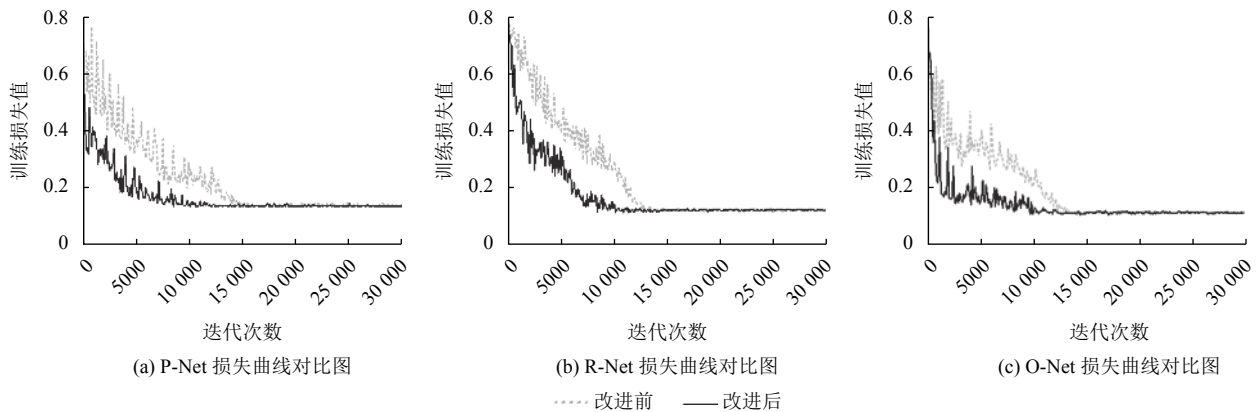


图6 网络改进前后损失曲线对比图

通过在网络中使用大小不同卷积核并行的方式, 结合 Batch Normalization 算法的优化策略, 对各级网络结构进行调整. 表 1 是各级改进前后的网络在训练集上人脸检测精度的对比结果. 由表 1 可见, 本文所述的优化改进方案对人脸检测准确率有一定的提高.

表 1 改进前后各个网络训练集上准确率对比 (%)

模型	P-Net		R-Net		O-Net	
	改进前	改进后	改进前	改进后	改进前	改进后
准确率	94.6	95.2	95.4	96.1	95.4	96.6

在 Fddb 数据集上使用 ROC 曲线对改进前后算法、文献 [11] 算法、Faster R-CNN、Haar-like 进行测试评估, 如图 7 所示, 其中横坐标假正例数 FP(False Positives) 表示的是误检数目, 即把非人脸检测为人脸, 纵坐标真正例率 TPR (True Positives Rate) 为数据集中标注的人脸被检测出的比例, 图中 MTCNN 算法、本文算法、文献 [11] 算法以及 Faster R-CNN 算法它们的准确率都要远高于传统算法 Haar-like, 本文算法比 MTCNN 和文献 [11] 算法有更好的检测效果, 虽然三者的准确率都低于 Faster R-CNN, 但是 Faster R-CNN 的网络较深、模型较大, 难以部署在移动端并实时检测.

对随机拍摄到的 100 张电梯人员人脸图片, 分别使用改进前后的 MTCNN 算法进行检测, 将检测的准

4.4 结果分析

通过在各级网络中加入 Batch Normalization 算法对训练过程进行优化, 图 6 中的 3 个子图分别是改进前后 P-Net、R-Net 以及 O-Net 的损失曲线对比图. 从图中可以看出, 改进后各网络的收敛速度更快, 训练损失值更加的稳定, 迭代步数也相对减少, 从而减少不必要的训练时间.

准确率和模型大小以及在基于嵌入式平台开发的电梯信息终端上电梯乘客人脸检测的速度进行综合性能对比, 如表 2 所示, 虽然增大了模型但对检测速度没有造成过大的影响, 而且改进后的模型检测的准确率比原来高. 图 8 是两种算法电梯乘客人脸检测的效果对比图, 图 9 是两种算法是电梯乘客多人人脸检测的效果对比图, 可见本文算法对电梯乘客识别效果更好.

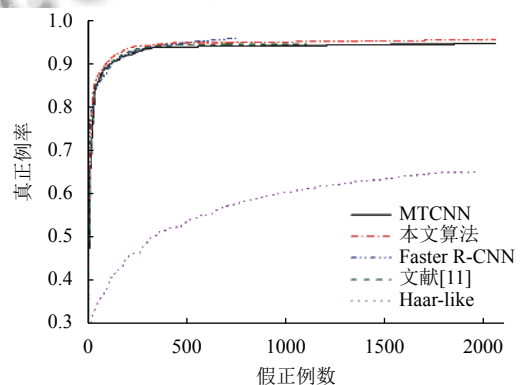


图7 Fddb 数据集的 ROC 曲线图

表 2 MTCNN 改进前后算法模型的性能综合对比

模型	模型大小(MB)	速度(fps)	准确率(%)
改进前	1.9	22	94
改进后	2.0	20	96

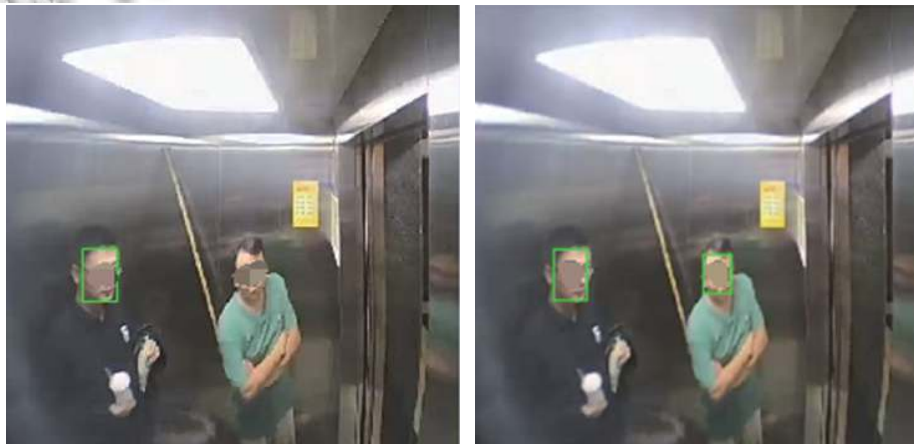


(a) 模型改进前电梯乘客人脸检测效果图



(b) 模型改进后电梯乘客人脸检测效果图

图8 电梯乘客人脸检测的效果对比图



(a) 模型改进前电梯乘客多人人脸检测效果图

(b) 模型改进后电梯乘客多人人脸检测效果图

图9 电梯乘客多人人脸检测的效果对比图

5 总结

在电梯安全监测系统中使用传统人脸检测算法识别电梯乘客,但检测精度并不高难以满足实际情况的应用,本文将 MTCNN 人脸检测算法应用于电梯人员检测,并在此基础上,结合 Inception 模块的设计思想,通过并行操作多个大小不同的卷积核,来加深加宽各级网络的结构,增强各级网络的表达能力,同时基于 BN 算法能够避免训练时人工参数选择的随机性提高训练的收敛速度以及 BN 算法具有一定增强网络的分类能力的优点,将二者引入到 MTCNN 的网络结构中,提高网络对人脸检测的准确率,使得改进后的 MTCNN

人脸检测算法对电梯乘客的识别更加理想,在电梯安全监测系统应用中,能准确了解轿厢乘客具体情况,当电梯出现异常时,有利于救援工作的展开,进一步提高了乘梯安全。

参考文献

- 1 郭威, 吴允平, 王廷银. MEMS 四元数卡尔曼滤波算法的电梯姿态估计. 计算机系统应用, 2020, 29(3): 246-252. [doi: 10.15888/j.cnki.csa.007289]
- 2 李东风. 基于物联网技术的电梯安全运行监控系统. 通讯世界, 2017, (17): 8-9. [doi: 10.3969/j.issn.1006-4222.2017.17.005]

- 3 张凯, 庞玉凯. 基于 ZigBee 网络的电梯安全监测系统. 中国科技信息, 2015, (23): 24–25. [doi: [10.3969/j.issn.1001-8972.2015.23.005](https://doi.org/10.3969/j.issn.1001-8972.2015.23.005)]
- 4 宋文迪, 巩凯强. 电梯轿厢人员检测方法. 信息通信, 2020, (1): 269–271. [doi: [10.3969/j.issn.1673-1131.2020.01.135](https://doi.org/10.3969/j.issn.1673-1131.2020.01.135)]
- 5 Viola P, Jones MJ. Robust real-time face detection. International Journal of Computer Vision, 2004, 54(2): 137–154.
- 6 Dalal N, Triggs B. Histograms of oriented gradients for human detection. Proceedings of 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego, CA, USA. 2005:886–893.
- 7 王潇宇. 基于 HOG 的电梯乘客人脸检测方法研究 [硕士学位论文]. 沈阳: 沈阳建筑大学, 2016.
- 8 Kumar NVR, DhanaSekar G, Dennis M. Application of face detection system for passenger counting in lifts using HAAR features. ARPN Journal of Engineering and Applied Sciences, 2016, 11(13): 8336–8341.
- 9 蒋纪威, 何明祥, 孙凯. 基于改进 YOLOV3 的人脸实时检测方法. 计算机应用与软件, 2020, 37(5): 200–204. [doi: [10.3969/j.issn.1000-386x.2020.05.035](https://doi.org/10.3969/j.issn.1000-386x.2020.05.035)]
- 10 Shen QF, Jiang LF, Xiong HL. Person tracking and frontal face capture with UAV. Proceedings of the 2018 IEEE 18th International Conference on Communication Technology. Chongqing, China. 2018. 1412–1416.
- 11 胡玲玲. 基于移动端的人脸检测算法研究 [硕士学位论文]. 杭州: 浙江理工大学, 2018.
- 12 周慧娟, 张强, 刘羽, 等. 基于 YOLO2 的地铁进站客流量检测方法. 计算机与现代化, 2019, (10): 76–82. [doi: [10.3969/j.issn.1006-2475.2019.10.015](https://doi.org/10.3969/j.issn.1006-2475.2019.10.015)]
- 13 易诗, 陈鑫凯, 宋瑞源, 等. 适用于智能球机的高鲁棒性侵入跟踪方法. 电子科技大学学报, 2019, 48(5): 754–758. [doi: [10.3969/j.issn.1001-0548.2019.05.015](https://doi.org/10.3969/j.issn.1001-0548.2019.05.015)]
- 14 Halawa LJ, Adi W, Ferda E. Face recognition using Faster R-CNN with inception-V2 architecture for CCTV camera. Proceedings of the 2019 3rd International Conference on Informatics and Computational Sciences. Semarang, Indonesia. 2019. 1–6.
- 15 Zhang KP, Zhang ZP, Li ZF, *et al.* Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Signal Processing Letters, 2016, 23(10): 1499–1503. [doi: [10.1109/LSP.2016.2603342](https://doi.org/10.1109/LSP.2016.2603342)]
- 16 朱含杉. 无人机系统下基于卷积神经网络的人脸检测技术研究 [硕士学位论文]. 南京: 南京航空航天大学, 2019.
- 17 刘宇明, 凌志祥, 吴强, 等. 基于多任务卷积网络的参会人员人数统计算法. 计算机应用, 2018, 38(S2): 51–54.
- 18 陈雨薇. 基于改进 MTCNN 模型的人脸检测与面部关键点定位 [硕士学位论文]. 上海: 东华大学, 2019.
- 19 Szegedy C, Liu W, Jia YQ, *et al.* Going deeper with convolutions. Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA. 2015. 1–9.
- 20 Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. Proceedings of the 32nd International Conference on Machine Learning. Lille, France. 2015. 448–456.
- 21 Sun Y, Wang XG, Tang XO. Deep convolutional network cascade for facial point detection. Proceedings of 2013 IEEE Conference on Computer Vision and Pattern Recognition. Portland, OR, USA. 2013. 3476–3483.