



DOI:10.12404/j.issn.1671-1815.2406407

引用格式:李涛,钟寒.基于改进多任务级联卷积神经网络的人脸检测算法[J].科学技术与工程,2025,25(21):9010-9017.

Li Tao, Zhong Han. Face detection method based on improved MTCNN [J]. Science Technology and Engineering, 2025, 25(21): 9010-9017.

基于改进多任务级联卷积神经网络的人脸检测算法

李涛, 钟寒*

(中国人民公安大学信息安全学院, 北京 100038)

摘要 多任务级联卷积神经网络(multi-task convolutional neural networks, MTCNN)人脸检测算法虽然在一些人脸检测任务上取得了不错的成绩,但是面对一些复杂的小尺度、多人脸检测任务,人脸检测的准确率仍有待提高。提出一种改进的 MTCNN 算法,首先,在制造数据集时微调交并比(intersection over union, IoU) 阈值参数,来更精确地对人脸样本进行分类;其次,用卷积层替换网络的最大池化层,来提升网络性能;最后,在 O-Net 网络中引入 SE(squeeze-excitation) 注意力机制,提高网络的特征表达能力。试验结果表明:改进后的算法与原始 MTCNN 算法相比较,P-Net 网络和 R-Net 网络检测准确率提高了 1%,O-Net 网络检测准确率提高了 0.5%,且改进后的算法在实际人脸检测任务中表现更好。

关键词 MTCNN; 人脸检测; SE 注意力机制

中图分类号 TP391.41; **文献标志码** A

Face Detection Method Based on Improved MTCNN

LI Tao, ZHONG Han*

(College of Information & Cyber Security, People's Public Security University of China, Beijing 100038, China)

[Abstract] Although the multi-task convolutional neural networks (MTCNN) face detection algorithm has achieved good results in some face recognition tasks, the accuracy of face detection needs to be improved in the face of some complex small-scale and multi-person face detection tasks. An improved MTCNN algorithm was proposed. Firstly, the intersection over union (IoU) threshold parameter was fine-tuned when creating the data set to classify face samples more accurately. Secondly, replacing the max pooling layer of the network with convolutional layers can improve network performance. Finally, the squeeze-excitation(SE) attention mechanism was introduced into the O-Net network to improve the feature expression ability of the network. The test results show that compared with the original MTCNN algorithm, the detection accuracy of the P-Net network and R-Net network of the improved algorithm has increased by 1%, and the detection accuracy of the O-Net network has increased by 0.5%. Moreover, the improved algorithm performs better in the actual face detection task.

[Keywords] MTCNN; face detection; SE attention mechanism

人脸检测是识别视频或图像中是否存在人脸的过程,如果有人脸存在,则进一步检测出人脸的位置、大小和人脸面部关键点的位置信息^[1]。早期的人脸检测工作主要基于人工精心设计的局部描述进行特征提取^[2-5],但随着深度学习的蓬勃发展,逐渐演变出许多基于不同深度学习架构的人脸检测方法,极大地改善了传统人脸检测在特征提取、精确度、可扩展性方面的诸多不足^[6-9]。基于卷积神经网络的人脸检测最早可以追溯到 1994 年,随着研究的深入,基于卷积神经网络(convolutional neural networks, CNN)的人脸检测相对于其他人脸检测方法取得了更加优异的成绩,但在遮挡^[10]、光

线^[11]、复杂背景^[12-13]等情况下的人脸检测效果仍不理想。

Zhang 等^[14]在 2016 年 ECCV 会议上提出了一种多任务级联卷积神经网络(multi-task convolutional neural networks, MTCNN),MTCNN 网络结构较小,效率高,在不受约束的环境下兼顾了人脸检测和人脸对齐功能,可应用于人脸支付、安防监控、社交平台等多个领域,成为工业界的人脸检测主流算法^[15]。但研究表明,MTCNN 网络在面对一些复杂的检测任务时,检测准确率仍有待提高,为解决这一问题,一些改进的 MTCNN 网络被相继提出。蓝雯飞等^[16]通过对原始 MTCNN 算法进行迁移学习,

收稿日期:2024-08-26 修订日期:2025-04-16

基金项目:高等学校学科创新引智基地(B20087)

第一作者:李涛(1990—),男,汉族,河南郑州人,硕士研究生。研究方向:人脸识别。E-mail:2022212138@stu.ppsuc.edu.cn。

*通信作者:钟寒(1984—),女,汉族,山东济宁人,博士,副教授。研究方向:数据分析。E-mail:zhonghan@ppsuc.edu.cn。

使算法在自然环境中人脸检测的准确率有所提高。顾梅花等^[17]提出集成上下文信息卷积模块,对多任务卷积神经网络 MTCNN 的 R-Net 和 O-Net 网络模块进行优化,有效提升了算法在课堂场景下的小人脸检测率。周航等^[18]通过对 MTCNN 算法损失函数进行均方差函数求均值,使网络收敛更加平稳,检测准确率更高。庞殊杨等^[19]通过将 MobileNet 模块融入 MTCNN 算法中,并适当调整网络卷积核和全连接层神经元个数使模型更适用于不同尺寸的安全帽识别。刘彩云等^[20]通过设计一种新的集成图片信息卷积残差模块,并将其融入 R-Net 和 O-Net 网络层中来增大感受野,从而进一步提升了网络性能。李艳灵等^[21]将多任务卷积神经网络 MTCNN 的 R-Net 和 O-Net 网络模块的 NMS 算法优化成 Better-NMS 算法,在不影响网络鲁棒性的同时,提升了人脸检测的准确率。李丽平等^[22]采用锚框思路和深度可分离卷积思路,优化了多任务卷积神经网络 MTCNN,进一步提升了模型性能。

基于上述研究,现从训练数据和 MTCNN 网络结构两个方面入手进行优化,进一步提升模型性能,使模型在面对复杂的小尺度、多人脸检测任务时,能够准确无误的识别出图像或视频中的全部人脸。首先,在制造数据集时微调交并比(intersection over union, IoU) 阈值参数,来更精确地对人脸样本进行分类;其次,用卷积层替换网络的最大池化层来提升网络性能;最后,在 O-Net 网络中引入 SE 注

意力机制,提高模型的特征表达能力。

1 MTCNN 简介

MTCNN 网络是一种高效的人脸检测算法,通过级联的 3 个 CNN 网络(P-Net、R-Net 和 O-Net)实现了人脸的快速、准确检测。MTCNN 主要由 3 个框架组成,如图 1 所示。

P-Net(proposal network) 12×12 : 主要负责获得人脸框的回归向量和人脸 5 个关键点位置,通过非极大值抑制(non-maximum suppression, NMS)来合并重叠度高的候选框。

R-Net(refine network) 24×24 : 通过边界框回归和 NMS 进一步去掉非人脸区域,由于 R-Net 网络与 P-Net 相比较多了一个全连接层,因此能更好地抑制非人脸区域。

O-Net(output network) 48×48 : 与 R-Net 相比较,O-Net 多了一层卷积层,处理结果更加精准,同时对人脸区域进行更多监督,并输出一个人脸框和 5 个面部关键点坐标。

MTCNN 的设计理念是通过不同网络层级之间的级联,利用低复杂度的模型快速生成候选框,再利用高复杂度的模型对候选框进行筛选,从而实现速度与精度的共同提升。MTCNN 不仅能够进行人脸检测,还能够输出人脸关键点位置,可应用于人脸识别、人脸表情分析、人脸美化等多个领域,成为工业界的人脸检测主流算法。

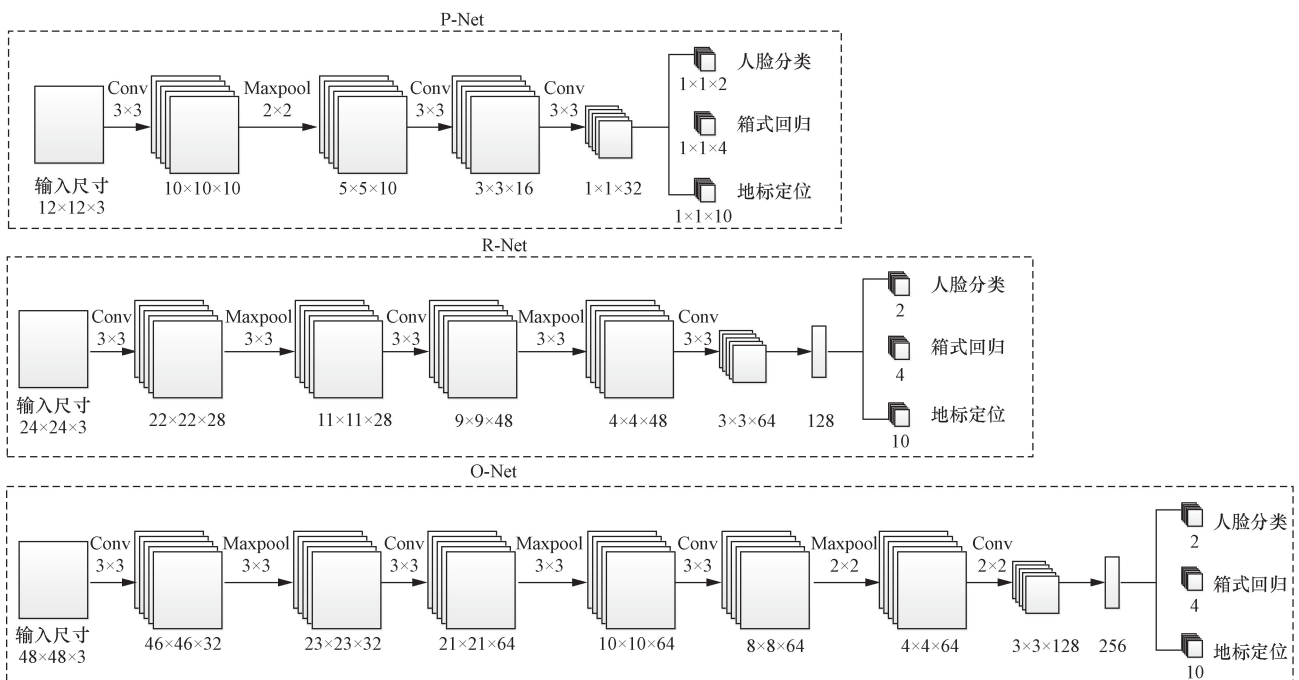


图1 MTCNN 网络结构

Fig. 1 MTCNN network structure

2 算法改进

2.1 数据集与标签制作

采用 CelebA 数据集作为基准数据集生成各个子网络训练所需的样本。CelebA 数据集是一个大型的人脸属性数据集,由香港中文大学多媒体实验室创建,包括 10 177 个身份、202 599 张面部图像、5 个地标位置以及每张图片 40 个二进制属性注释。

以 CelebA 数据集为基准生成样本后,需将样本分为正样本、部分样本、负样本对网络进行训练,使用 NMS 算法中交并比 (IoU) 对样本进行分类。IoU 的本质是真实的人脸框与用于训练的人脸框之间的重叠值,示意图如图 2 所示。

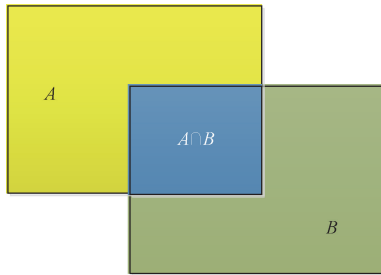


图 2 两个区域的交与并

Fig. 2 Intersection and merging of two regions

IoU 计算公式为

$$IoU = \frac{A \cap B}{A \cup B} = \frac{A \cap B}{A + B - A \cup B} \quad (1)$$

式(1)中: A 为真实的人脸框的面积; B 为用于训练的人脸框的面积; \cup 表示两者的并集; \cap 表示两者的交集。IoU 数值小则重合度低, IoU 数值大则重合度高,用于训练的样本的类别可根据 IoU 的大小去判断。

原文 IoU 阈值分别设置为 0.65、0.4、0.3,当 $IoU > 0.65$ 时,代表样本为正样本; $0.4 < IoU < 0.65$ 时,代表样本为部分样本; $IoU < 0.3$ 时,代表样本为负样本。而 IoU 阈值的设置则直接决定着样本的保留或删除,决定着样本标签的属性,若阈值设置不合理,则无法精准地划分样本类别,进而影响模型训练效果。经过反复实验证明:当 $IoU > 0.7$ 时,为正样本;当 $0.3 < IoU < 0.7$ 时,为部分样本;当 $IoU < 0.2$ 为负样本。样本分类更准确,且将负样本、正样本、部分样本、坐标样本比例设置为 7:1:1:1 时,模型训练效果更佳。

2.2 用卷积层替换最大池化层

池化层 (pooling layers) 是卷积神经网络中的一种层结构,一般使用于卷积层与卷积层之间,主要作用是增大感受野、降低特征维度、防止过拟合、缩减模型大小、提高计算速度^[23]。本文模型在连续卷

积层中间使用最大池化来降低特征图大小,提取特征图中最重要的特征,最大池化运算示例如图 3 所示。可知,最大池化虽然降低了特征图大小,减少了后续层的参数量,但是原特征图将近 8/9 的信息丢失,严重影响了网络的性能。所以在目标检测任务中,使用最大池化层会导致大量信息丢失从而影响网络检测准确率^[24],为解决这一问题,使用卷积层替换网络中的最大池化层,来提升网络的目标检测精度。

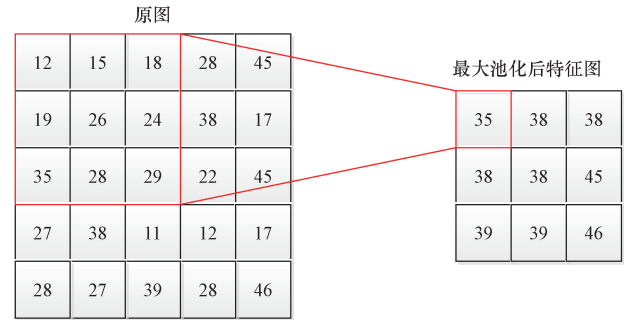


图 3 最大池化运算示例

Fig. 3 Example of maximum pooling

2.3 引入 SE 注意力机制

深度学习中的注意力机制可以模仿人类视觉和认知系统,它在处理输入数据时,可以模仿人类的神经系统,把注意力集中在重要信息上,通过引入注意力机制,神经网络能够像人类一样,拥有自主学习能力,并选择性地学习重要信息,提高网络的性能和泛化能力^[25]。为了进一步提高网络检测精度,在 O-Net 网络的第一次卷积运算后引入 SE^[26] 注意力机制,提高网络的特征表达能力。

SE 注意力机制通过自主学习的方式赋予有用特征通道高的权重,赋予无用特征通道低的权重,以此来获取更多有价值信息^[27]。SE 注意力机制主要采用压缩和激励两种操作来实现所述功能,模块结构如图 4 所示。

步骤 1 Transformation (F_{tr}): 给定一个输入的特征图 X , 通过 F_{tr} 操作, 生成特征图 U , 公式如下所示。

$$U = v_c * X = \sum_{s=1}^{c'} v_c^s * X^s \quad (2)$$

式(2)中: X 为输入特征图; v_c 为二维空间卷积核; $*$ 表示卷积操作。

步骤 2 Squeeze (F_{sq}): 通过 F_{sq} 操作, 将特征图 U 压缩为一个 $1 \times 1 \times C$ 的向量, 公式如下所示。

$$Z_c = F_{sq}(U) = \frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j) \quad (3)$$

式(3)中: U 为特征图; C 为特征图通道数; HW 为特征图空间维度; u_c 为单个通道的特征图。

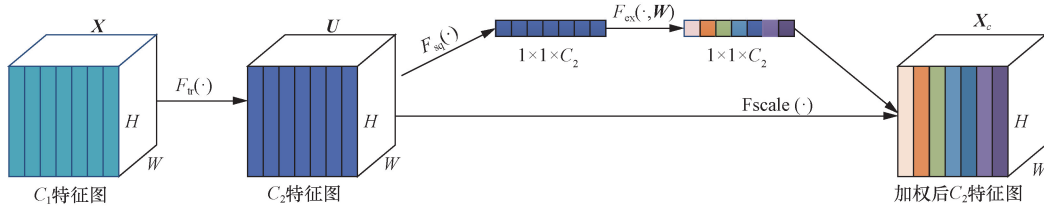


图4 SE 模块

Fig. 4 SE module

步骤3 Excitation (F_{ex}): 对上一步得到的向量通过两个全连接层进行处理, 得到通道权重值, 赋予不同通道不同的权重, 用不同的权重值表示不同通道的重要程度, 公式如下所示。

$$S = F_{ex}(Z, W) = \sigma[g(Z, W)] \\ = \sigma[W_2 \delta(W_1, Z)] \quad (4)$$

式(4)中: Z 为特征图; W_1 、 W_2 为两个全连接层的权重矩阵; δ 为 ReLU 激活函数; σ 为 Sigmoid 激活函数。

步骤4 Scale (F_{scale}): 将上一步得到的权重向量对特征图 U 进行权重赋值, 完成对原始特征的重标定, 生成最终的特征图, 公式如下所示。

$$X_c = F_{scale}(u_c, s_c) = s_c u_c \quad (5)$$

式(5)中: u_c 为单个通道特征图; s_c 为通道权重。

优化后的 MTCNN 网络结构如图 5 所示。

3 实验结果与分析

3.1 实验环境

实验环境为 Windows11, 搭载 Inter Core i7 12700H 处理器, 内存为 16 G, 采用 Nvidia RTX 3060 laptop, 显存为 6 G, 采用 Python3.7 进行代码编译。

3.2 损失函数

由于 MTCNN 网络由 3 个子网络组成, 因此其损失函数也由 3 部分组成。针对人脸识别问题, 利用交叉熵损失函数来判断是否存在人脸, 对于人脸框回归和人脸关键点位置定位, 使用 L_2 损失函数, 最后将这 3 个损失函数的损失和自身的权重相乘然后累加起来, 形成最后的总损失。在 P-Net 和 R-Net 训练的过程中, 更加注重目标框的准确度, 而较少关注关键点判定的损失, 因此目标框损失所占权重较大, 而关键点损失所占权重较小。对于 O-Net 的训练则恰恰相反, 更加注重关键点的位置, 因此关键点损失所占权重较大, 目标框损失所占权重较小。

对于人脸识别的问题, 将输入的样本 x_i , 使用交叉熵代价函数, 公式为

$$L_i^{det} = -[y_i^{det} \ln p_i + (1 - y_i^{det})(1 - \ln p_i)] \quad (6)$$

式(6)中: y_i^{det} 为样本的真实标签; p_i 为样本为人脸的概率。

对于目标框的回归, 采用欧氏距离, 公式为

$$L_i^{box} = \left\| \hat{y}_i^{box} - y_i^{box} \right\|_2^2 \quad (7)$$

式(7)中: \hat{y}_i^{box} 为通过校正之后得到的边界框的坐标; y_i^{box} 为目标的真实边界框。

对于关键点, 也采用的是欧氏距离, 公式为

$$L_i^{land} = \left\| \hat{y}_i^{land} - y_i^{land} \right\|_2^2 \quad (8)$$

式(8)中: \hat{y}_i^{land} 为网络输出之后校正得到的边界框的坐标; y_i^{land} 为目标的真实边界框。

把上述 3 个损失函数按照不同的权重联合起来, 得

$$\min \sum_{i=1}^N \sum_{j \in \{det, box, land\}} \alpha_j \beta_i^j L_i^j \quad (9)$$

式(9)中: N 为训练样本的总数; $\beta_i^j \in \{0, 1\}$ 为样本类型指示符; α_j 为各个损失所占的权重。在训练 P-Net 和 R-Net 网络时, 把 α_{det} 、 α_{box} 和 α_{land} 分别设置为 1、0.5、0.5, 在训练 O-Net 网络时, 把 α_{det} 、 α_{box} 和 α_{land} 分别设置为 1、0.5 和 1。

3.3 实验与分析

MTCNN 网络采用级联的思想一步步筛选出符合标准的样本, 将一个大的网络拆解成 3 个小型网络, 即 P-Net、R-Net、O-Net。在模型训练过程中, 需要先用大小为 12×12 的数据训练 P-Net 网络, 然后利用训练好的 P-Net 网络生成训练 R-Net 网络所需的大小为 24×24 的数据。生成的大小为 24×24 的数据将 R-Net 网络训练成功以后, 利用训练好的 R-Net 网络生成训练 O-Net 网络所需的大小为 48×48 的数据, 最后用生成的大小为 48×48 的数据训练 O-Net 网络。

在分别制作 P-Net、R-Net、O-Net 数据时, 用 IoU 阈值对数据进行分类, 当 $IoU > 0.7$ 时, 为正样本; 当 $0.3 < IoU < 0.7$ 时, 为部分样本; 当 $IoU < 0.2$ 时, 为负样本。分类完成后, 将负样本、正样本、部分样本、坐标样本比例设置为 7:1:1:1。

3.3.1 P-Net 网络

在训练 P-Net 网络时, 先用微调后的 IoU 阈值对数据进行分类, 并将负样本、正样本、部分样本、坐

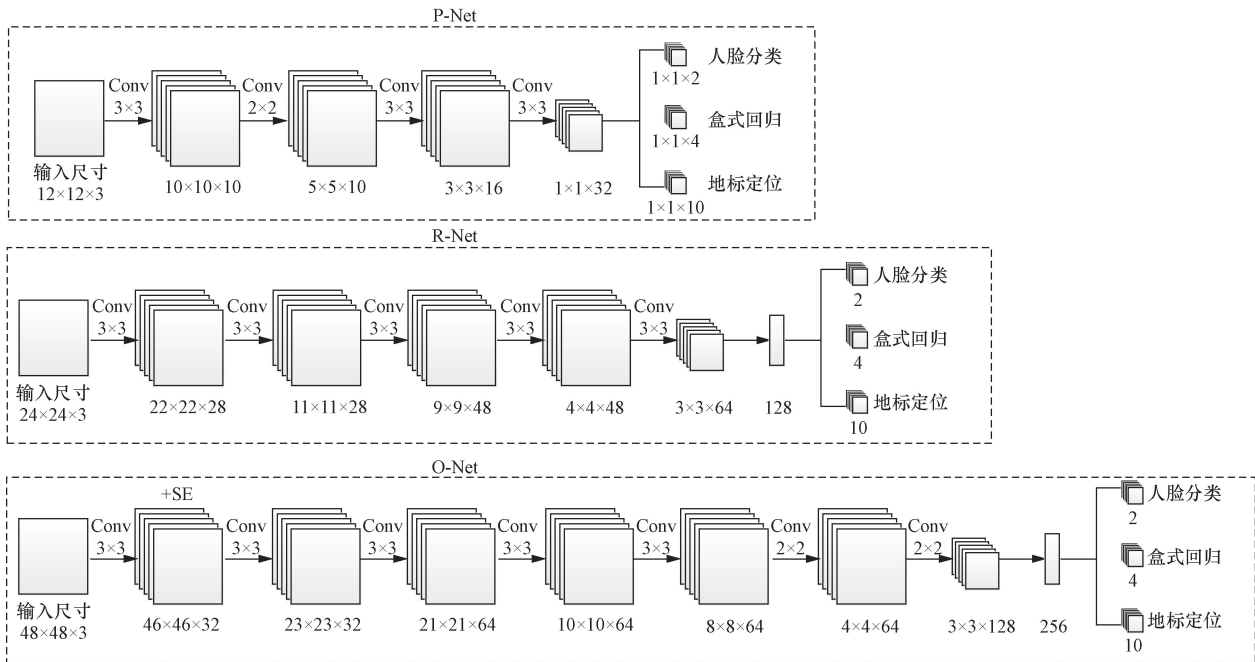


图5 优化后的MTCNN网络结构

Fig. 5 Optimized MTCNN network structure

标样本比例设置为 7:1:1:1,然后用普通卷积替代 P-Net 网络的最大池化层,用优化后的数据训练优化后的 P-Net 网络,实验结果如图 6 所示,可知,P-Net 网络在原始 MTCNN 算法训练过程中准确率保持在 92.5% ~ 93%,而本文算法 P-Net 网络在训练过程中准确率保持 94%,相比原始 MTCNN 算法准确率更高。

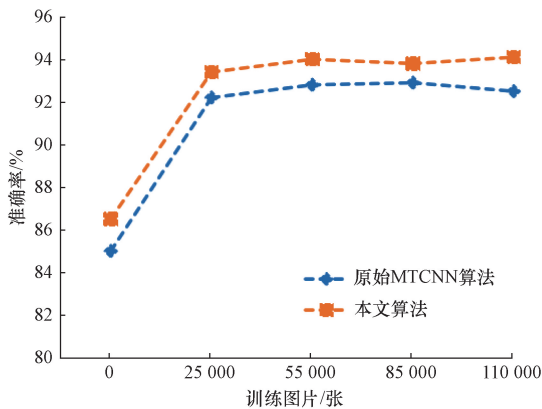


图6 P-Net 网络准确率

Fig. 6 The accuracy of the P-Net

3.3.2 R-Net 网络

在训练 R-Net 网络时,同样先用微调后的 IoU 阈值对数据进行分类,并将负样本、正样本、部分样本、坐标样本比例设置为 7:1:1:1,然后用普通卷积替代 R-Net 网络的最大池化层,用优化后的数据训练优化后的 R-Net 网络,实验结果如图 7 所示,可知,原始 MTCNN 算法 R-Net 网络在训练过程中准确

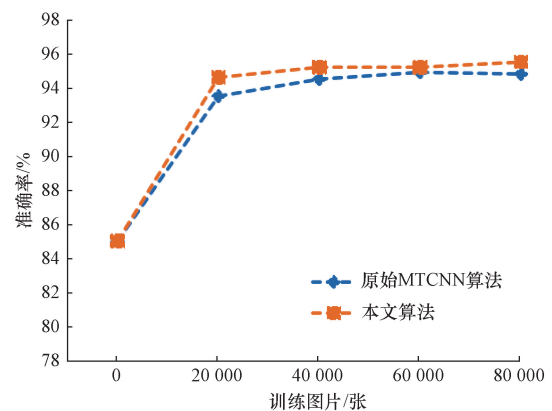


图7 R-Net 网络准确率

Fig. 7 The accuracy of the R-Net

率保持在 94% ~ 95%,而本文算法 R-Net 网络在训练过程中准确率保持在 95% ~ 96%,相比原始 MTCNN 算法准确率更高。

3.3.3 O-Net 网络

在训练 O-Net 网络时,同样先用微调后的 IoU 阈值对数据进行分类,并将负样本、正样本、部分样本、坐标样本比例设置为 7:1:1:1,然后用普通卷积替代 O-Net 网络的最大池化层,并在第一层卷积后加入 SE 注意力机制,最后用优化后的数据训练优化后的 O-Net 网络。实验结果如图 8 所示。可知,原始 MTCNN 算法 O-Net 网络在训练过程中准确率保持在 97% ~ 97.5%,而本文算法 O-Net 网络在训练过程中准确率保持 97.5% ~ 98%,相比原始 MTCNN 算法的准确率更高。

3.3.4 实际效果对比

把用原始 MTCNN 算法数据训练好的模型和本文模型训练好的模型进行多人脸检测,来对比优化前后模型的实际效果,原始 MTCNN 算法检测结果如图 9 所示,本文模型检测结果如图 10 所示。对比检测结果可知,原始 MTCNN 算法检测图片中有 5 个戴口罩人脸、一个严重遮挡人脸、两个既戴口罩又戴帽子人脸检测失败,本文模型检测图片中只有两个既戴口罩又戴帽子人脸检测失败。通过对比可知,本文模型的实际人脸检测效果要优于原始 MTCNN 算法模型。

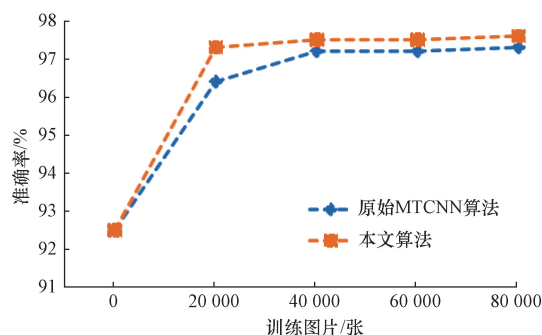


图 8 O-Net 网络准确率

Fig. 8 The accuracy of the O-Net



图 9 原始 MTCNN 算法模型检测结果

Fig. 9 Detection results of the original MTCNN algorithm



图 10 本文模型检测结果

Fig. 10 Detection results of the model in this paper

4 结论

针对在一些复杂的小尺度、多人脸检测任务中人脸检测的准确率低的问题,提出一种改进的 MTCNN 算法。首先,在制造数据集时微调 IoU 阈值参数,来获取更多、更精确的人脸样本;其次,用卷积层替换网络的最大池化层,来提升网络性能;最后,在 O-Net 网络中引入 SE 注意力机制,提高模型的特

征表达能力。试验结果表明:改进后的算法与原始 MTCNN 算法相比较,P-Net 网络和 R-Net 网络检测准确率提高了 1%,O-Net 网络检测准确率提高了 0.5%,且改进后的算法在实际人脸检测任务中表现更好。在未来的研究中,将致力于研究遮挡情况下人脸检测数据集和遮挡情况下如何提升 MTCNN 算法人脸检测准确率,提升 MTCNN 算法在复杂环境下的性能。

参 考 文 献

- [1] Ho H, Nguyen L, Le T, et al. Face detection using eigenfaces: a comprehensive review[J]. *IEEE Access*, 2024, 12: 118406-118426.
- [2] 赵立新, 邢润哲, 白银光, 等. 深度学习在目标检测的研究综述[J]. *科学技术与工程*, 2021, 21(30): 12787-12795.
Zhao Lixin, Xing Runzhe, Bai Yinguang, et al. Review on survey of deep learning in target detection[J]. *Science Technology and Engineering*, 2021, 21(30): 12787-12795.
- [3] Li J M, Xie H T, Li J H, et al. Frequency-aware discriminative feature learning supervised by single-center loss for face forgery detection[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Los Alamitos: IEEE Computer Society Press, 2021: 6458-6467.
- [4] Albiero V, Chen X Y, Yin X, et al. Img2pose: face alignment and detection via 6DoF, face pose estimation[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Los Alamitos: IEEE Computer Society Press, 2021: 7617-7627.
- [5] 张阳婷, 黄德启, 王东伟, 等. 基于深度学习的目标检测算法研究与应用综述[J]. *计算机工程与应用*, 2023, 59(18): 1-13.
Zhang Yangting, Huang Deqi, Wang Dongwei, et al. Review on research and application of deep learning-based target detection algorithms[J]. *Computer Engineering and Applications*, 2023, 59(18): 1-13.
- [6] Chaudhary B, Aghdaie P, Soleymani S, et al. Differential morph face detection using discriminative wavelet sub-bands[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Los Alamitos: IEEE Computer Society Press, 2021: 1425-1434.
- [7] Gao H X, Tao W, Wen D C, et al. DupNet: towards very tiny quantized CNN with improved accuracy for face detection[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Los Alamitos: IEEE Computer Society Press, 2019: 168-177.
- [8] 徐铭, 李华. 基于改进YOLOv5s-face的Face5系列人脸检测算法[J]. *重庆理工大学学报(自然科学)*, 2024, 38(6): 194-202.
Xu Ming, Li Hua. Face5 series face detection algorithm based on improved YOLOv5s-face[J]. *Journal of Chongqing University of Technology(Natural Science)*, 2024, 38(6): 194-202.
- [9] Vedapant N, Wang B Y. Cface: confidence ranker for model-agnostic face detection refinement[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Los Alamitos: IEEE Computer Society Press, 2021: 1674-1684.
- [10] 戴惠丽. 融合深度字典学习和特征重建的遮挡人脸检测研究[J]. *计算机应用与软件*, 2024, 41(11): 228-233.
Dai huili. An occlusion face detection method combining depth dictionary learning and feature reconstruction[J]. *Computer Applications and Software*, 2024, 41(11): 228-233.
- [11] Wang W J, Yang W H, Liu J Y. HLA-Face: joint high-low adaptation for low light face detection[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Los Alamitos: IEEE Computer Society Press, 2021: 16195-16204.
- [12] 丁雅眉. 复杂场景下的人脸检测方法研究[D]. 西安: 西安邮电大学, 2024.
Ding Yamei. Research on face detection methods in complex Scenes[D]. Xi'an: Xi'an University of Posts and Telecommunications, 2024.
- [13] 董子平, 陈世国, 廖国清. 基于YOLOv5s的密集多人脸检测算法[J]. *计算机工程与科学*, 2023, 45(10): 1838-1846.
Dong Ziping, Chen Shiguo, Liao Guoqing. A dense multi-face detection algorithm based on YOLOv5s[J]. *Computer Engineering and Science*, 2023, 45(10): 1838-1846.
- [14] Zhang K, Zhang Z, Li Z, et al. Joint face detection and alignment using multitask cascaded convolutional networks[J]. *IEEE Signal Process*, 2016, 23(10): 1499-1503.
- [15] 刘晓军, 类成玉, 张晓玲, 等. 基于卷积神经网络的人脸检测综述[J]. *兵工自动化*, 2021, 40(6): 58-63.
Liu Xiaojun, Lei Chengyu, Zhang Xiaoling, et al. Face detection survey based on convolutional neural network[J]. *Ordnance Industry Automation*, 2021, 40(6): 58-63.
- [16] 蓝雯飞, 张盛兰, 朱容波, 等. 基于改进MTCNN的人脸检测算法[J]. *中南民族大学学报(自然科学版)*, 2020, 39(6): 637-641.
Lan Wenfei, Zhang Shenglan, Zhu Rongbo, et al. Face detection algorithm based on improved MTCNN[J]. *Journal of South-Central University for Nationalities(Natural Science Edition)*, 2020, 39(6): 637-641.
- [17] 顾梅花, 冯婧, 杨娜. 一种融合上下文信息特征的改进MTCNN人脸检测算法[J]. *西安工程大学学报*, 2021, 35(6): 114-120.
Gu Meihua, Feng Jing, Yang Na. Improved MTCNN face detection algorithm fused with context features[J]. *Journal of Xi'an Polytechnic University*, 2021, 35(6): 114-120.
- [18] 周航, 蔡茂国, 吴涛, 等. 一种改进的多任务级联神经网络人脸检测算法研究[J]. *智能计算机与应用*, 2021, 11(3): 172-176.
Zhou Hang, Cai Maoguo, Wu Tao, et al. Research on an improved multi-task cascade network face detection algorithm[J]. *Intelligent Computer and Applications*, 2021, 11(3): 172-176.
- [19] 庞殊杨, 芦莎. 基于改进MTCNN的多尺度安全帽识别[J]. *计算机应用研究*, 2021, 38(6): 1907-1912, 1916.
Pang Shuyang, Lu Sha. Multi-scale safety helmet detection based on improved MTCNN[J]. *Application Research of Computers*, 2021, 38(6): 1907-1912, 1916.
- [20] 刘彩云, 李雅雯, 刘倩. 一种改进的多任务级联卷积神经网络人脸检测算法[J]. *长江大学学报(自然科学版)*, 2021, 18(6): 111-118.
Liu Caiyun, Li Yawen, Liu Qian. An improved face detection algorithm based on multi-task cascaded convolutional neural network[J]. *Journal of Yangtze University(Natural Science Edition)*, 2021, 18(6): 111-118.
- [21] 李艳灵, 王莎莎, 杨志鹏. 一种改进的多任务级联卷积神经网络人脸检测算法[J]. *计算机应用研究*, 2021, 38(6): 1907-1912, 1916.

- 网络人脸检测算法[J]. 信阳师范学院学报(自然科学版), 2022, 35(4): 651-655.
- Li Yanling, Wang Shasha, Yang Zhipeng. An improved multi-task cascaded convolutional neural network face detection algorithm [J]. Journal of Xinyang Normal University (Natural Science Edition), 2022, 35(4): 651-655.
- [22] 李丽平, 许营坤, 王嘉航. 基于锚框稀疏图像金字塔的MTCNN人脸检测方法[J]. 浙江工业大学学报, 2023, 51(3): 298-304, 318.
- Li Liping, Xu Yingkun, Wang Jiahang. Research on MTCNN face detection method of sparse image pyramid based on anchor box [J]. Journal of Zhejiang University of Technology, 2023, 51(3): 298-304, 318.
- [23] Ma Z, Chang D, Xie J, et al. Fine-grained vehicle classification with channel max pooling modified CNNs[J]. IEEE Transactions on Vehicular Technology, 2019, 68(4): 3224-3233.
- [24] Springenberg T J, Dosovitskiy A, Brox T, et al. Striving for simplicity: the all convolutional net[J]. ArXiv, 2014: 1412.6806.
- [25] Woo S, Park J, Lee J Y. CBAM: convolutional block attention module[C]//European Conference on Computer Vision (ECCV). Munich: Springer, 2018: 11211.
- [26] Hu J, Li S, Albanie S, et al. Squeeze-and-excitation networks [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Los Alamitos: IEEE Computer Society Press, 2018: 7132-7141.
- [27] 汪建伟, 游疆, 万敏, 等. 复杂背景下的低空无人机检测与跟踪算法[J]. 强激光与粒子束, 2023, 35(7): 158-169.
- Wang Jianwei, You Jiang, Wan Min, et al. Low-altitude UAV detection and tracking algorithms in complex backgrounds[J]. High Power Laser and Particle Beams, 2023, 35(7): 158-169.