

多注意力机制的口罩检测网络

余阿祥¹, 李承润¹, 于书仪¹, 李洪均^{1,2}

(1.南通大学信息科学技术学院, 江苏 南通 226019)

(2.南京大学计算机软件新技术国家重点实验室, 江苏 南京 210023)

[摘要] 提出一种口罩佩戴检测模型, 引入多注意力机制, 提升了网络特征挖掘能力; 利用柔性非极大抑制方法, 消除多余目标检测框。在公共数据库上的仿真实验表明, 该模型检测人脸口罩佩戴的平均精度达到 93.81%, 帧率达到 11.8 fps, 能有效地进行人脸口罩佩戴检测。

[关键词] 口罩佩戴检测, 多注意力机制, 特征挖掘, 柔性非极大抑制

[中图分类号] TP391 **[文献标志码]** A **[文章编号]** 1672-1292(2021)01-0023-07

Multi-Attention Mechanism of Mask Wearing Detection Network

Yu Axiang¹, Li Chengrun¹, Yu Shuyi¹, Li Hongjun^{1,2}

(1.School of Information Science and Technology, Nantong University, Nantong 226019, China)

(2.State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210023, China)

Abstract: A mask in public places can effectively control the transmission of the coronavirus. To this end, a mask wearing detection model is proposed. The model introduces a multi-attention mechanism to improve the network feature mining ability and uses soft-NMS methods to eliminate redundant target detection boxes. A simulation experiment is conducted on a public database. The average accuracy of the proposed face mask wearing detection reaches 93.81%, and the frame rate reaches 11.8 fps. The experimental results show that the model can effectively detect the face mask wearing.

Key words: mask wearing test, multi-attention mechanism, feature of the mining, soft-NMS

新冠肺炎具有极强的传染性^[1], 可通过呼吸道飞沫、气溶胶和接触传播, 病毒能在自然环境下存活 5 d^[2], 公共场所佩戴口罩成为抑制病毒传播的必要手段。在国家卫健委发布的《新型冠状病毒感染肺炎预防指南》中明确要求, 个人外出前往公共场所、就医和乘坐公共交通工具时需要佩戴口罩。贯彻落实佩戴口罩的相关规定不仅需要个人自觉遵守, 也需要采取一定的手段加以监督。

考虑到口罩检测和目标检测具有很强的相似性, 首先对现有的目标检测模型进行研究。Girshick 等在 2014 年提出了 R-CNN^[3], 将神经网络用于目标检测, 真正实现了将检测任务转换为分类任务; 检测算法的主流也由传统模型转为卷积神经网络模型, 检测的平均精度均值得到了巨大的提升。随后, Girshick 等又在 2015 年相继提出 Fast R-CNN^[4] 和 Faster R-CNN^[5]。与 R-CNN 相比, Fast R-CNN 在最后一层卷积层后加了一个 ROI(region of interest)池化层; 损失函数使用了多任务损失, 将边框回归直接加入到 CNN 网络中训练。Faster R-CNN 创造性地采用卷积网络自行产生建议框, 并和目标检测网络共享卷积网络, 省去了选择性搜索生成候选框环节, 使得算法的速度得到很大的提升, 但仍达不到实时检测的效果。Redmon 等提出的 YOLO 算法^[6] 是首个单阶段目标检测算法, YOLO 将目标检测问题看作是一个回归问题, 直接从图像像素信息得到边界框的位置和类别概率, 速度快, 但准确率较低。随后 Liu 等提出 SSD 算法^[7], 利用不同尺度的特征图进行分类预测和回归预测, 在保证准确率的同时, 可达到实时检测。接着, Redmon 等又相继推出了 YOLOv2^[8]、YOLOv3^[9], 通过不同的改进使得模型精确度得到了更好的提升。2017 年 Lin 及其团队提出用 Focal Loss 替换交叉熵损失来提高检测精度^[10], 很好地改善了单阶段目标检测算法正负样本不均

收稿日期: 2020-08-08。

基金项目: 国家自然科学基金项目(61871241、61976120)、南京大学计算机软件新技术国家重点实验室基金项目(KFKT2019B015)、江苏省研究生科研与实践创新计划项目(KYCX19_2056)、南通大学大学生创新训练计划项目(2020109)。

通讯作者: 李洪均, 博士, 副教授, 研究方向: 人工智能。E-mail: lihongjun@ntu.edu.cn

衡问题,使得 RetinaNet 的精度超越了 Faster R-CNN 的目标检测算法. 2019 年, Tan 等人在 EfficientNet^[11] 的基础上,提出了名为 EfficientDet^[12] 的一种新的目标检测系列, EfficientDet 采用一种复合缩放的方法来缩放网络的宽度、深度和分辨率,其参数少,所需要的计算资源少,速度快,准确率高.

考虑到佩戴口罩检测需要在自然场景下运用,人与人之间的重叠程度高,人脸区域模糊,本文在 EfficientDet 的基础上加入多尺度注意力机制,以便更好地挖掘戴口罩人脸的特征,进而提高人脸佩戴口罩的检测准确度. 同时,利用柔性非极大抑制的方法消除多余检测框,避免重叠人脸检测框被误删,减小模型的误检率.

1 EfficientDet 网络

EfficientDet 网络提出一种新的特征加强网络(BiFPN),可实现多尺度特征融合. 图 1 显示了 EfficientDet 的总体体系结构. EfficientDet 采用 EfficientNet 作为主干网络, BiFPN 作为特征加强网络,从主干网络中获取 3-7 级 { P3, P4, P5, P6, P7 } 输入特征,并分别应用自上而下和自下而上的双向功能融合,融合的特征被送到分类网络和检测框预测网络,分别生成目标的类和目标检测框.

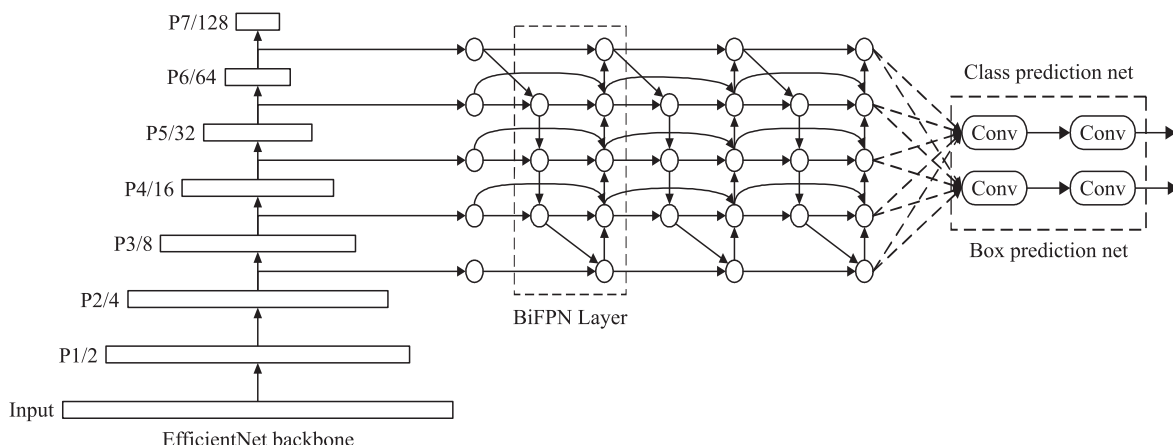


图 1 EfficientDet 目标检测网络

Fig. 1 EfficientDet target detection network

1.1 主干特征提取网络

目标检测算法中常用主干网络有 ResNet^[13]、VGGNet^[14] 等. VGGNet 主要通过 3×3 卷积核来提取图像特征, ResNet 由若干个卷积模块组成,增加了跳层连接,解决了网络加深引起的网络退化问题,使得网络加深的情况下预测效果依然得到提升. 但网络加深会导致参数增加及训练成本增加. EfficientNet 重新思考了在原有网络结构基础上增加模型精度的方法:增加网络的深度、宽度和分辨率. 任一参数的增加都能提高模型精度,但是随着参数的增加,模型的精度增加变得缓慢,于是模型在总参数量不变的情况下,找出了 8 种不同的缩放模型,分别为 EfficientNet-B0~EfficientNet-B7. 下面以 EfficientNet-B0 为例加以说明.

EfficientNet-B0 由 1 个 Stem 模块及若干个 Block 模块组成. Stem 模块包含 1 个 3×3 步长为 2 的卷积层、批量标准化层及激活函数层. Block 为一种通用模块,如图 2 所示.

EfficientNet-B0 总体设计思路是反转残差结构^[15]和残差结构,在 3×3 或 5×5 网络结构前利用 1×1 卷积升维,在 3×3 或 5×5 网络结构后增加一个关于通道方向的注意力机制,最后通过 1×1 卷积降维后增加一个残差边.

在每个模块中,对于输入 \mathbf{x} ,对应的输出 $H(\mathbf{x})$ 可通过残差边的连接建立如下公式:

$$H(\mathbf{x}) = f(\mathbf{x}, \{w_i\}) + \mathbf{x}, \quad (1)$$

式中, \mathbf{x} 为输入向量; w_i 为权重参数; $f(\mathbf{x}, \{w_i\})$ 为需要学习的残差映射,对于其余层的连接单元,可由式 (2) 表示:

$$f(\mathbf{x}, \{w_i\}) = w_3(\delta_1(w_1(\delta_1(w_0\mathbf{x}))) \times \delta_2(w_2(\text{Avgpool}(\delta_1(w_1(\delta_1(w_0\mathbf{x})))))), \quad (2)$$

式中, δ_1 为 Swish 激活函数; δ_2 为 Sigmoid 激活函数; w_i 为各部分对应权重参数; Avgpool(\cdot) 为全局平均池化操作.

1.2 BiFPN 特征增强网络

BiFPN 网络将 PANet^[16] 和 NAS-FPN^[17] 融合,其主要思想是有效的双向跨尺度连接和加权特征融合. 首先,删除那些只有 1 个输入边的节点,得到简化的双向网络. 其次,若原始输入节点与输出节点处于同一级别,则在其间添加一条额外的边. 最后,将每个双向路径视为 1 个要素网络层,并重复相同的层多次以启用更多高级功能融合. 由于特征融合过程中,每个特征图贡献不同,因此加入了一个可学习的加权平均的权重. 图 3 为一个 BiFPN Layer 的结构.

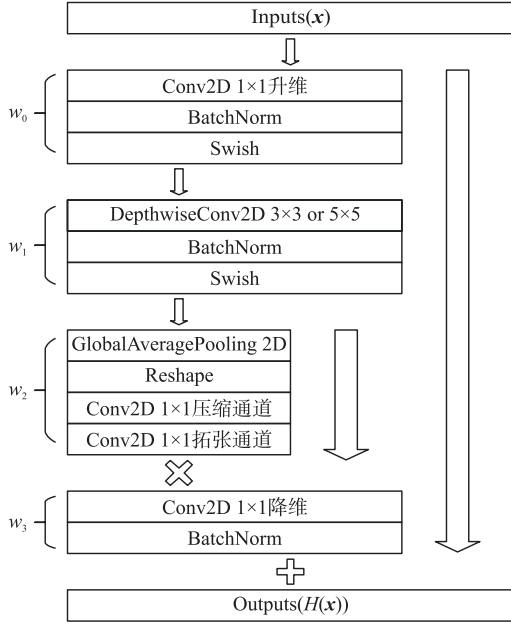


图 2 EfficientNet Block 内部具体结构

Fig. 2 The internal structure of EfficientNet Block

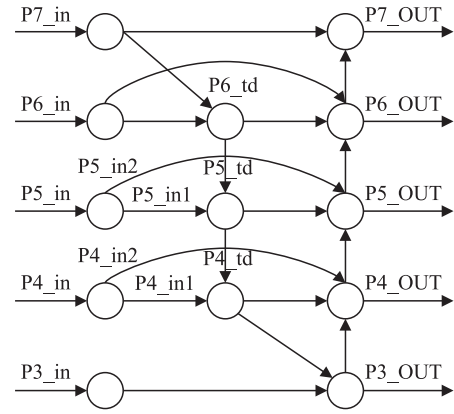


图 3 BiFPN Layer 具体结构

Fig. 3 The specific structure of BiFPN Layer

式(3)介绍了如何得到特征图 p6_td 和 p6_out:

$$\begin{aligned} P6_td &= \text{Conv} \left(\frac{w_1 \cdot P6_in + w_2 \cdot \text{Resize}(P7_in)}{w_1 + w_2 + \varepsilon} \right), \\ P6_out &= \text{Conv} \left(\frac{w'_1 \cdot P6_in + w'_2 \cdot P6_td + w'_3 \cdot \text{Resize}(P5_out)}{w'_1 + w'_2 + w'_3 + \varepsilon} \right), \end{aligned} \quad (3)$$

式中, w 为权重; $\varepsilon = 0.0001$ 为最小值以避免数值不稳定; $\text{Resize}(\cdot)$ 表示将特征图尺寸变换到与本层相对应的尺寸. EfficientNet-B0 共有 3 个 BiFPN Layer, 该层的输出为下一次 BiFPN Layer 的输入.

1.3 损失函数

EfficientNet-B0 的损失包括分类损失和边界框回归损失. 分类损失使用 Focal Loss 损失函数^[10] 在交叉熵损失函数的基础上进行修改, 通过调整类别权重 α 缓解正负样本不平衡问题. 实际检测中, 大多数先验框中不包含物体, 导致正负样本不平衡, 分类的准确性下降. 引入权重 α 能很好地解决正负样本不平衡问题:

$$\text{FL}(p, y) = \begin{cases} -\alpha(1-p)^\gamma \log p, & y = 1; \\ -(1-\alpha)p^\gamma \log(1-p), & y = 0. \end{cases} \quad (4)$$

采用 Smooth_{L1} 函数作为边界框回归损失来调整检测框的位置^[4], 其公式为:

$$\text{Smooth}_{L1}(x) = \begin{cases} 0.5x^2, & \text{if } |x| < 1; \\ |x| - 0.5, & \text{otherwise.} \end{cases} \quad (5)$$

其曲线平滑, 易于求导, 梯度稳定, 可以控制梯度的量级, 从而更易训练.

2 多注意力机制的高效网络

2.1 多尺度注意力机制

注意力机制被广泛用于自然语言处理^[18]、计算机视觉领域^[19]. 在计算机视觉中, 注意力机制可提高

特征的挖掘能力,进而提高图像分类和目标检测的准确性.在佩戴口罩检测的任务上,口罩种类多,目标较小,引入注意力机制能提高网络的特征挖掘能力,提升口罩检测的平均精度.本文在 BiFPN 模块之前加入注意力机制模块,其结构受 SE-Net^[20] 的启发,如图 4 所示.

输入特征图先经过全局平均池化,得到 $1 \times 1 \times C$ 向量,经两个全连接层和一个激活函数层得到的向量与输入特征图相乘,将施加注意力机制后的特征图输入到 BiFPN 特征加强网络进行特征加强.其过程可由式(6)表示:

$$f(\mathbf{x}) = (\delta_2(w_2 \delta_1((w_1(\text{AvgPool}(\mathbf{x})))))) \cdot \mathbf{x}, \quad (6)$$

式中, w_i 为全连接层对应的权重; δ_2 为 Sigmoid 激活函数; $\text{Avgpool}(\cdot)$ 表示全局平均池化; δ_1 为 Swish 激活函数; \mathbf{x} 表示输入特征图; $f(\mathbf{x})$ 表示输出特征图.

考虑到高层特征图具有较高的语义信息而缺少像素信息,较低的特征具有较强的像素信息而缺少语义信息,将注意力机制施加在 P3、P4、P5 这 3 个特征图.

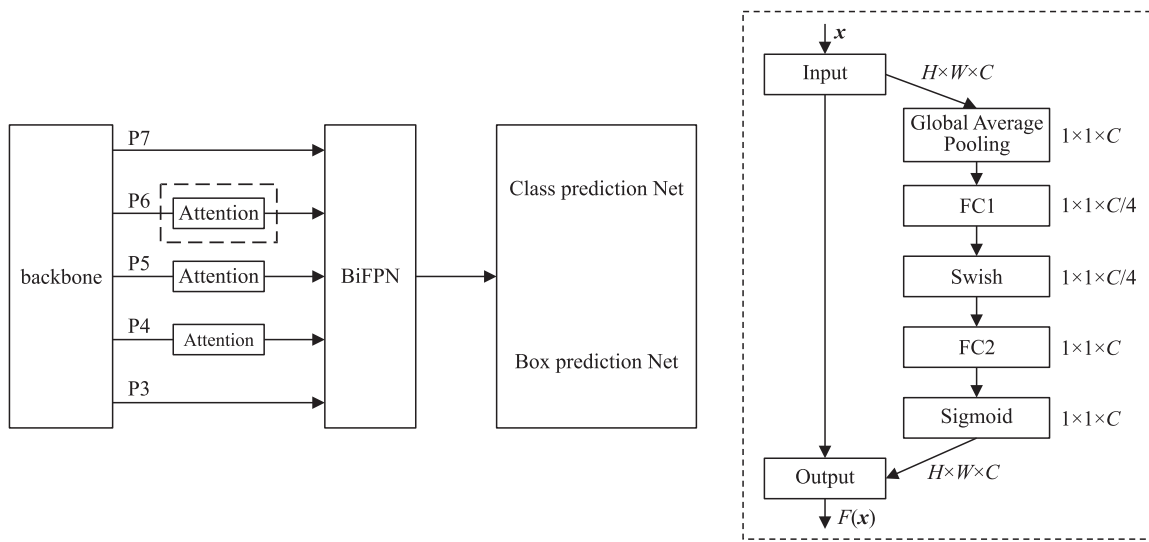


图 4 加入多尺度注意力机制的网络口罩佩戴检测模型

Fig. 4 Network mask wearing detection model with multi-scale attention mechanism

2.2 柔性非极大抑制

图片经过训练目标检测网络,会在一个物体上产生很多检测框,通常采用非极大抑制(NMS)算法来去除多余的检测框,保留得分最大的框,其原理可由式(7)表示:

$$s_i = \begin{cases} s_i, & \text{iou}(M, b_i) < N_t; \\ 0, & \text{iou}(M, b_i) \geq N_t. \end{cases} \quad (7)$$

式中, $B = \{b_1, \dots, b_N\}$ 是初始检验框的列表; $S = \{s_1, \dots, s_N\}$ 包含相应的检测分数; N_t 为 NMS 阈值; M 表示得分最高的检测; $\text{iou}(M, b_i)$ 表示两者对应的 iou.

当图片中被检测的物体稠密,物体与物体之间靠得太近或部分重叠时,漏检的可能性就大大增加.针对这个问题, Bodla 等提出了一种柔性非极大抑制^[21](soft-NMS).该算法减小与 M 重合的检测框的分数,而不是直接删除.因此,和 M 相距较远的检测框不会受到影响,和 M 重叠的检测框会被分到更大的惩罚,其过程如式(8)所示:

$$s_i = \begin{cases} s_i, & \text{iou}(M, b_i) < N_t; \\ s_i(1 - \text{iou}(M, b_i)), & \text{iou}(M, b_i) \geq N_t. \end{cases} \quad (8)$$

以上公式在重叠方面是不连续的,当重叠达到 N_t 时,会突然惩罚.为了使重叠连续,若惩罚函数是连续的,则为理想的状况,否则可能导致检测排名的突然变化.为了消除这种情况,引入高斯惩罚函数进行修剪:

$$s_i = s_i \exp\left(-\frac{\text{iou}(M, b_i)^2}{\delta}\right), \forall b_i \in D. \quad (9)$$

式中, D 为最终检测框集合; δ 为系数.

EfficientDet-D0 采用 NMS 除去多余的检测框,考虑到本实验中存在人脸重叠及人脸相距比较近的情况,将 NMS 改成 soft-NMS.

3 实验仿真与分析

3.1 数据库

本实验数据来源于 MAFA dataset^[22] 和 WIDER FACE dataset^[23]. 其中,来自 MAFA 的照片有 4 065 张,来自 WIDER Face 的照片有 3 894 张. 本数据集采用 VOC2007 格式进行标注,数据集中人脸分别标注为 face 或 face_mask. face 表示不戴口罩的人脸,face_mask 表示戴口罩的人脸. 将数据集随机划分为训练集、验证集和测试集,其中,训练集有 5 508 张照片,验证集有 612 张照片,测试集有 1 839 张照片.

3.2 实验环境及模型训练

本实验在 Windows10 操作系统下进行,编程环境为 Python3.7. 深度学习框架为 TensorFlow1.13.2, Keras2.1.5. 深度学习加速工具为 CUDA10.0, cudnn7.4.1.5. 硬件环境主要包括 CPU 为 Intel(R) Xeon(R) CPU E5-2696 v4 @ 2.20GHz, GPU 为 NVIDIA GeForce GTX 1080Ti@ 11GB, 内存(RAM) 32GB. 模型采用 Adam 优化算法进行训练^[24],其中 lr 为 0.001, β_1 为 0.9, β_2 为 0.999, epsilon 为 $1e-08$. 训练设置为 100 个 epoch. 前 50 个 epoch 训练时冻结主干网络,后 50 个 epoch 解冻主干网络;前 50 个 epoch 的批量为 8,后 50 个 epoch 批量为 4. 前 50 个 epoch 初始学习率为 0.001,当连续两个 epoch 验证集 loss 不发生变化时,学习率缩小一半;当连续 6 个 epoch 验证集 loss 不变时,提前终止训练. 后 50 个 epoch 初始学习率为 0.000 05,学习率衰减方式与前者相同.

主干网络初始化权重采用在 ImageNet 数据集训练的权重,其余层的初始化方式为“Xavier 方法^[25]”初始化.

3.3 实验评价指标

本实验采用平均精度(AP)和识别速度作为实验的评价指标来客观评价人脸佩戴口罩算法. AP 为单一目标的检测效果,其计算如式(10)所示:

$$A_{AP} = \int_0^1 p(r) dr, \quad (10)$$

式中, $p(r)$ 表示准确率 P (Precision) 和召回率 R (Recall) 之间的曲线. 准确率 P 和召回率 R 公式为:

$$P = \frac{TP}{TP+FP}, \quad (11)$$

$$R = \frac{TP}{TP+FN} \quad (12)$$

式中, TP 表示正样本被预测为正样本; FP 负样本被预测为正样本; FN 表示负样本被预测为负样本.

除了检测准确度,目标检测算法的另外一个重要性能指标是速度,识别速度的常用指标是每秒帧率 (frame per second, fps),即每秒内可以处理的图片数量.

3.4 实验结果分析

当 IOU 阈值设置为 0.5 时, EfficientDet 人脸佩戴口罩检测的平均精度达到了 93.1%, 本文提出的算法在人脸口罩佩戴口罩检测方面的平均精度达到了 93.81%, 如表 1 所示. 表 1 中, EfficientDet 表示不作任何改进, Attention 表示在 EfficientDet 的基础上单独加入多注意力机制, soft-NMS 表示在 EfficientDet 的基础上单独加入了柔性非极大抑制, Ours 为本文提出的算法.

由表 1 可知, 本文提出的模型相较于 EfficientDet 在人脸佩戴口罩检测方面和识别速度方面均得到了提升, 其中人脸佩戴口罩方面平均精度提升了 0.71%, 识别速度提升了 22.9%. 可见本文提出的算法能有效提高人脸口罩佩戴口罩检测的准确度和效率.

此外, 单独加入多注意力机制或柔性非极大抑制

表 1 实验结果对比

算法	face_mask AP/%	fps
EfficientDet	93.10	9.60
Attention	93.74	10.90
Soft-NMS	93.25	10.88
Ours	93.81	11.80

都能提高口罩佩戴检测的准确率和效率. 单独加入注意力机制时,人脸佩戴口罩检测的平均精度提高了 0.64%,说明加入注意力机制能更好地挖掘戴口罩人脸的特征,使得网络模型对于人脸佩戴口罩识别能力有一定的提高. 单独加入柔性非极大抑制,人脸佩戴口罩检测的平均精度提高了 0.15%,柔性非极大抑制能避免重叠检测框的误删,避免重叠人脸检测框的误删,从而提高人脸佩戴口罩检测的平均精度.

图 5 为实验效果图. 由图可知,本文提出的算法能够在复杂环境下区分佩戴口罩的人脸和未佩戴口罩的人脸,有效减少了模型的误检率. 同时,能更好地识别侧脸的情况,在侧脸方面的识别率更高. 该模型面对模糊及小人脸方面具有很好的鲁棒性,能够应对较为复杂的场景.



图 5 口罩佩戴检测实际效果图

Fig. 5 Actual effect diagram of mask wearing test

4 结论

本文通过对 EfficientDet 目标检测算法的研究,提出了一种改进 EfficientDet 的口罩佩戴检测模型,将其应用于口罩佩戴检测任务上. 加入了多尺度注意力机制,经实验验证在不同环境下均具有很好的识别效果和鲁棒性. 通过预训练好的 EfficientNet 主干网络加速模型的训练,并使用 soft-NMS 代替 NMS. 训练后的模型在测试集上对佩戴口罩人脸的检测平均精度达到 93.81%,识别速度达到 11.80 fps,验证了将多尺度注意力机制运用到目标检测模型来进行人脸口罩佩戴检测的可行性. 未来的研究将在保证准确率的前提下简化主干网络和特征增强网络,提高识别速度,加大训练数据集,进一步提高模型识别准确率.

[参考文献] (References)

- [1] 赵文明,宋述慧,陈梅丽,等. 2019 新型冠状病毒信息库[J]. 遗传,2020,42(2):212-221.
- [2] 白浪,王铭,唐小琼,等. 对新型冠状病毒肺炎诊疗中的热点问题的思考[J]. 华西医学,2020,35(2):125-131.
- [3] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Columbus, USA: IEEE Computer Society, 2014:580-587.
- [4] GIRSHICK R. Fast R-CNN[C]//IEEE International Conference on Computer Vision (CVPR). Santiago, Chile: IEEE Computer Society, 2015:1440-1448.
- [5] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.

- [6] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Las Vegas, USA: IEEE Computer Society, 2016: 779–788.
- [7] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox detector[C]//European Conference on Computer Vision(ECCV). Amsterdam, Netherlands: ECCV, 2016: 21–37.
- [8] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]//IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Honolulu, USA: IEEE Computer Society, 2017: 6517–6525.
- [9] REDMON J, FARHAD A. YOLOv3: An incremental improvement[EB/OL]. [2020-08-08]. <https://arxiv.org/abs/1804.02767>.
- [10] LIN T, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]//IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Honolulu, USA: IEEE Computer Society, 2017: 2999–3007.
- [11] TAN M, LE Q V. EfficientNet: rethinking model scaling for convolutional neural networks[C]//International Conference on Machine Learning(ICML). California, USA: IMLS, 2019: 6105–6114.
- [12] TAN M, PANG R, LE Q V, et al. EfficientDet: scalable and efficient object detection[C]//IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Seattle, USA: IEEE Computer Society, 2020: 10781–10790.
- [13] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Las Vegas, USA: IEEE Computer Society, 2016: 770–778.
- [14] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[C]//International Conference of Learning Representation. San Diego, USA, 2015.
- [15] SANDLER M, HOWARD A, ZHU M, et al. MobileNetV2: inverted residuals and linear bottlenecks[C]//IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Salt Lake City, USA: IEEE Computer Society, 2018: 4510–4520.
- [16] LIU S, QI L, QI H, et al. Path aggregation network for instance segmentation[C]//IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Salt Lake City, USA: IEEE Computer Society, 2018: 8759–8768.
- [17] GHIASI G, LIN T, LE Q V, et al. NAS-FPN: learning scalable feature pyramid architecture for object detection[C]//IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Long Beach, USA: IEEE Computer Society, 2019: 7036–7045.
- [18] 石磊, 王毅, 成颖, 等. 自然语言处理中的注意力机制研究综述[J]. 数据分析与知识发现, 2020, 4(5): 1–14.
- [19] 王文冠, 沈建冰, 贾云得. 视觉注意力检测综述[J]. 软件学报, 2019, 30(2): 416–439.
- [20] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Salt Lake City, USA: IEEE Computer Society, 2018: 7132–7141.
- [21] BODLA N, SINGH B, CHELLAPPA R, et al. Soft-NMS—Improving object detection with one line of code[C]//IEEE International Conference on Computer Vision(ICCV). Venice, Italy: IEEE Computer Society, 2017: 5562–5570.
- [22] GE S, LI J, YE Q, et al. Detecting masked faces in the wild with LLE-CNNs[C]//IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Honolulu, USA: IEEE Computer Society, 2017: 426–434.
- [23] YANG S, LUO P, LOY C C, et al. WIDER FACE: a face detection benchmark[C]//IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Las Vegas, USA: IEEE Computer Society, 2016: 5525–5533.
- [24] KINGMA D P, BA J. Adam: a method for stochastic optimization[C]//The 3rd International Conference for Learning Representations. San Diego, USA, 2015.
- [25] GLOYOT X, BENGIO Y. Understanding the difficulty of training deep feed forward neural networks[J]. Journal of Machine Learning Research, 2010, 9: 249–250.

[责任编辑: 严海琳]