

# 多尺度特征融合的轻量化口罩佩戴检测算法

叶茂, 马杰, 王倩, 武麟

(河北工业大学 电子信息工程学院, 天津 300401)

**摘要:** 科学规范地佩戴口罩是预防新冠、流感等呼吸道传染病的有效方法,在当前疫情形势下,正确佩戴口罩显得尤为重要。已有的口罩佩戴检测算法多数存在结构复杂、训练难度较高和特征提取不足等问题,为此,提出一种多尺度特征融合的轻量化口罩佩戴检测算法 L-MFFN-YOLO。以 YOLOv4-Tiny 网络为基础, L-MFFN-YOLO 改进原始残差结构,使用轻量化残差模块促进模型快速收敛,在有效降低模型计算量的同时保证检测精度。在原网络  $13 \times 13$ 、 $26 \times 26$  这 2 个尺度的基础上增加  $52 \times 52$  特征分支,以增强低特征层的信息表达能力并降低小目标的漏检率。通过多层级交叉融合结构最大程度地提取有用信息,从而提高特征利用率。除佩戴和未佩戴口罩 2 种情况外,在数据集中新增口罩佩戴不正确的类别并进行手工标注,实验结果表明, L-MFFN-YOLO 算法的模型大小仅为 5.8 MB,较原始网络 YOLOv4-Tiny,其模型规模减小 76%, mAP 提高 5.25 个百分点, CPU 下的处理时间快 14 ms,能在资源受限的设备中满足口罩佩戴检测任务对准确率和实时性的要求。

**关键词:** 口罩佩戴检测;轻量化检测算法;残差结构;低特征层;多层级交叉融合

开放科学(资源服务)标志码(OSID):



**中文引用格式:** 叶茂, 马杰, 王倩, 等. 多尺度特征融合的轻量化口罩佩戴检测算法[J]. 计算机工程, 2022, 48(7): 42-50.

**英文引用格式:** YE M, MA J, WANG Q, et al. Lightweight mask-wearing detection algorithm with multi-scale feature fusion[J]. Computer Engineering, 2022, 48(7): 42-50.

## Lightweight Mask-Wearing Detection Algorithm with Multi-Scale Feature Fusion

YE Mao, MA Jie, WANG Qian, WU Lin

(School of Electronic Information Engineering, Hebei University of Technology, Tianjin 300401, China)

**[Abstract]** Standardized usage of face masks is effective as a non-pharmaceutical intervention to prevent the spread of infectious respiratory diseases, such as COVID-19 and influenza. In the current epidemic situation, wearing face masks correctly is especially important. Most existing mask-wearing detection algorithms involve problems such as complex structures, high training difficulty, and insufficient feature extraction. Therefore, this study proposes a lightweight mask-wearing detection algorithm based on multi-scale feature fusion and the YOLOv4-Tiny network, called L-MFFN-YOLO. L-MFFN-YOLO improves on the original residual structure and uses a lightweight residual module to promote rapid convergence. Moreover, it reduces the computational load while ensuring detection accuracy. Based on the original network's  $13 \times 13$  and  $26 \times 26$  feature maps,  $52 \times 52$  feature branches are added to enhance the ability of the lower feature layer to express information and reduce the false negative rate for small targets. On this basis, a Multi-level Cross Fusion (MCF) structure is used to maximally extract useful information so as to improve feature utilization. In addition to detecting mask-wearing, a category of masks worn incorrectly is added to the dataset and manually labeled. The experimental results show that the size of the proposed L-MFFN-YOLO model is only 5.8 MB, which is 76% smaller than that of the original YOLOv4-Tiny. Moreover, the mean Average Precision (mAP) of the proposed approach is 5.25 percentage points higher, and its processing time is 14 ms faster on an equivalent CPU. These results demonstrate that the proposed approach can meet the requirements of accuracy and real-time operation in resource-constrained devices to detect faces wearing masks.

**[Key words]** mask-wearing detection; lightweight detection algorithm; residual structure; lower characteristic layer; Multi-level Cross Fusion (MCF)

**DOI:** 10.19678/j.issn.1000-3428.0062231

**基金项目:** 河北省自然科学基金(F2020202045)。

**作者简介:** 叶茂(1997—),女,硕士研究生,主研方向为计算机视觉;马杰(通信作者),教授、博士;王倩、武麟,硕士研究生。

**收稿日期:** 2021-08-01 **修回日期:** 2021-09-16 **E-mail:** jma@hebut.edu.cn

## 0 概述

新型冠状病毒(COVID-19)能够通过飞沫、空气等进行传播,其极强的传染性和较长的潜伏期给人类健康和社会日常生活造成严重影响。虽然我国疫苗接种的人数越来越多,但是新冠病毒仍然在不断进化和变异,多个新冠变异毒株相继出现,对全民的生命健康造成极大威胁。

在病毒爆发初期,LIU等<sup>[1]</sup>提出佩戴口罩可以有效抑制新型冠状病毒的传播。目前,在超市、医院等多数公共场所,都需要人工提醒佩戴口罩,这种方法极大地造成了人力资源的浪费。随着深度学习的发展,利用人工智能技术可以快速地检测出人们的口罩佩戴情况。

目前,口罩佩戴检测技术作为新冠病毒的关键性预防手段之一,其相关算法研究显得尤为重要。在YOLOv3<sup>[2]</sup>的基础上,王艺皓等<sup>[3]</sup>结合跨阶段局部网络(Cross Stage Partial Network, CSPNet)<sup>[4]</sup>、改进的空间特征金字塔池化结构<sup>[5]</sup>以及路径聚合网络(Path Aggregation Network, PANet)<sup>[6]</sup>,在复杂场景下对口罩佩戴情况进行检测,其精度效高,但是检测速度仅为38 FPS。叶子勋等<sup>[7]</sup>使用改进的MobileNetv3<sup>[8]</sup>替换原有主干网络CSPDarkNet53,解决了YOLOv4<sup>[9]</sup>网络模型庞大的问题,模型大小由原来的244 MB降低至44 MB,并新增SiLU激活函数优化模型效果。曹城硕等<sup>[10]</sup>在主干网络中引入残差注意力机制提升模型对显著性特征的表达能力,然后使用特征金字塔网络(Feature Pyramid Network, FPN)<sup>[11]</sup>和PANet策略进行特征融合,但是,其计算复杂度高达 $3.098 \times 10^{10}$ 。余阿祥等<sup>[12]</sup>研究EfficientDet<sup>[13]</sup>检测算法,在其中加入多尺度注意力机制<sup>[14]</sup>并使用soft-NMS<sup>[15]</sup>代替NMS,该方法能提高检测精度,但是识别速度仅为11.8 FPS。

大型网络在内存资源少、处理器性能不高以及功耗受限的设备上应用时面临巨大挑战,因此,轻量化网络应运而生。在网络模型设计方面,基于

YOLOv4-Tiny,彭成等<sup>[16]</sup>通过加入Ghost模块<sup>[17]</sup>和ShuffleConv模块<sup>[18]</sup>,大幅降低了模型参数,但是其精度较原始网络降低了0.2%。除了轻量化网络结构之外,一些研究人员还会利用模型压缩<sup>[19-20]</sup>、剪枝<sup>[21]</sup>等方式解决算法效率与存储问题。

上述算法在口罩佩戴检测任务中能发挥一定作用,但是仍然存在以下2个问题:多数网络能够大幅提升检测精度,但是模型较大,检测速度较慢;部分轻量化网络能有效降低模型参数,但是难以取得准确率和检测速度之间的平衡。

本文基于YOLOv4-Tiny网络,设计一种实时多尺度特征融合的轻量化口罩佩戴检测算法L-MFFN-YOLO。使用轻量化残差模块促进模型快速收敛,降低训练时间开销。考虑到口罩佩戴检测的目标较小,利用浅层特征分支来降低小目标漏检率,在原网络 $13 \times 13$ 、 $26 \times 26$ 这2个尺度的基础上增加 $52 \times 52$ 特征分支。同时,引入多层级交叉融合(Multi-level Cross Fusion, MCF)模块解决多尺度目标对模型精度的影响,以更好地利用深层和浅层特征信息增强有效特征的表达。

## 1 轻量化口罩佩戴检测算法

YOLOv4-Tiny<sup>[4]</sup>是YOLOv4的轻量级版本,其网络结构简单,且能有效兼顾精度和速度。YOLOv4-Tiny网络结构如图1所示,主干网络CSPDarkNet53-Tiny主要包括标准卷积、残差结构堆叠和下采样3个部分:标准卷积采用CBL结构,由卷积层Conv2D、批量标准化(Batch Normalization, BN)和激活函数Leaky ReLU组成;残差结构堆叠使用CSP(Cross Stage Partial)模块,将基础层的特征分成2个部分,第一部分直接构建残差边,第二部分通过标准卷积和上一个部分进行拼接;下采样通过最大值池化。主干网络将 $13 \times 13$ 的特征层与进行2倍上采样的特征层进行融合,最终利用YOLO Head生成预测结果。

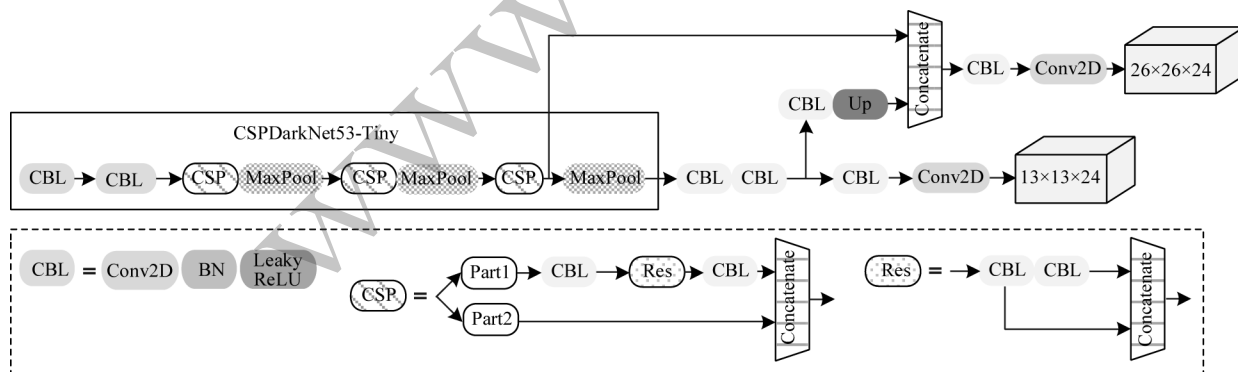


图1 YOLOv4-Tiny结构

Fig.1 The structure of YOLOv4-Tiny

为进一步提升网络性能,本文改进算法L-MFFN-YOLO的结构如图2所示。图2(a)部分采用轻量化的残差结构。由于YOLOv4-Tiny只包含2个尺度,小目标漏检情况严重,因此图2(b)部分

新增8倍下采样尺度分支,通过主干网络提取出3种不同分辨率的特征然后进行多层级交叉融合,在保证资源消耗较少的情况下增强不同分支中的特征图信息,最终提升特征提取能力。

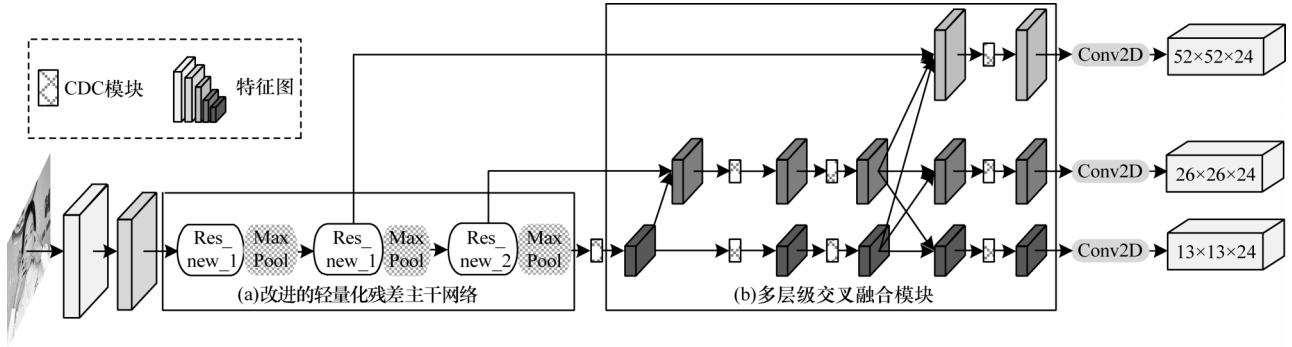


图2 L-MFFN-YOLO结构

Fig.2 The structure of L-MFFN-YOLO

### 1.1 改进的轻量化残差结构

在一些资源受限的设备中,较深的卷积网络模型容易导致反向传播过程中出现梯度弥散、梯度爆炸等问题,使得口罩佩戴检测模型无法很好地收敛。运算量较少的卷积操作在保证模型有效性的同时可以缓解上述问题。深度可分离卷积(Depth-Wise Separable Convolution)<sup>[22]</sup>可以减少算法参数,提高训练速度,并且对检测精度影响较小,能够解决网络加深引起的性能退化问题。深度可分离卷积可分成深度卷积(Depth-

Wise Convolution, DW Conv)和逐点卷积(Point-Wise Convolution, PW Conv),即 $1 \times 1$ 标准卷积,其结构如图3(a)所示。在此基础上,本文引入一种能保证检测性能并进一步降低参数数量的CDC模块,如图3(b)所示。首先,使用 $1 \times 1$ 标准卷积生成少量原始内部特征图,减少大量的卷积运算并将通道信息重新进行整合;然后,针对原始内部特征,通过深度卷积在不损失特征信息的情况下实现较小的计算代价;最后,将原始内部特征与经过深度卷积后的特征进行融合。

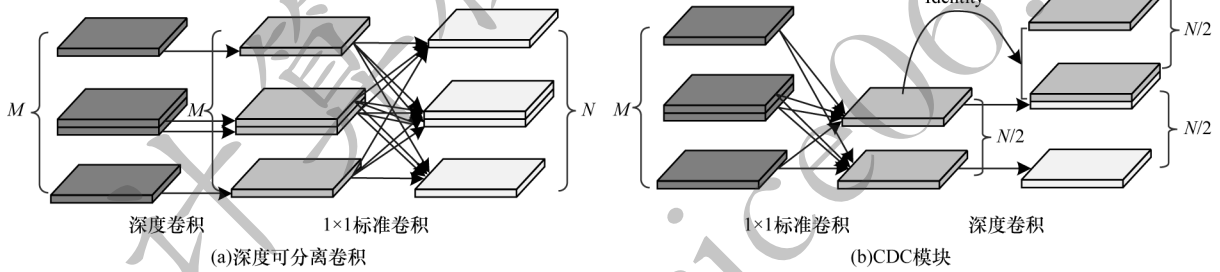


图3 不同的卷积过程

Fig.3 Different convolution processes

设2种不同卷积的输入特征大小为 $H \times W \times M$ ,输出特征大小为 $H_o \times W_o \times N$ ,其中, $M$ 和 $N$ 分别代表输入和输出通道。图3(a)中DW Conv卷积核尺寸为 $D_k \times D_k \times 1 \times M$ ,PW Conv卷积核尺寸为 $1 \times 1 \times M \times N$ ,CDC中2个卷积核的尺寸分别为 $1 \times 1 \times M \times \frac{N}{2}$ 和 $D_k \times D_k \times 1 \times \frac{N}{2}$ ,则CDC模块与深度可分离卷积的参数量比值为 $\frac{M \times N + D_k \times D_k \times N}{2(D_k \times D_k \times M + M \times N)}$ 。

在本文中,使用 $D_k \times D_k = 5 \times 5$ 的卷积扩大有效感受野,并假设 $M = N$ ,根据参数量比值可知CDC模块的参数量更少。

上述2种卷积形式都可以在检测任务中大幅降低模型参数数量,但是改进模块在硬件条件有限的情况下更具优势。图4所示为不同残差模块结构的对比。在YOLOv4-Tiny中,采用图4(a)所示的CSP模块进行残差结构堆叠;图4(b)中Res\_new\_X是本文主干网络的重要组成部分,其用CDC模块替换原有CSP结构中的部分卷积。

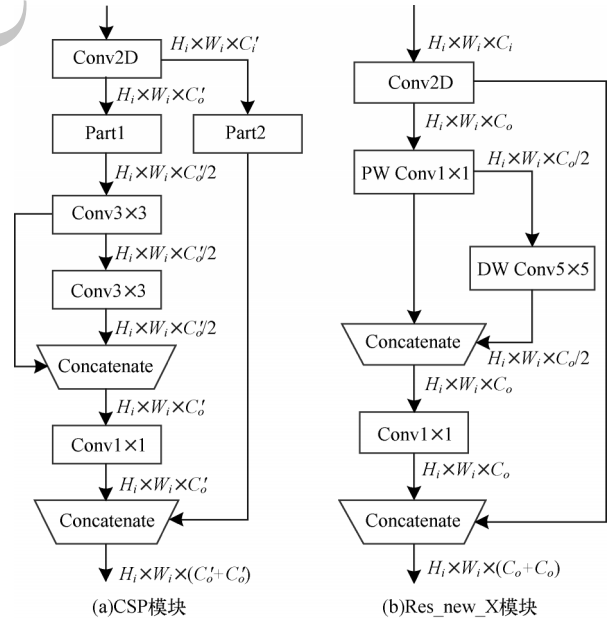


图4 不同的残差模块

Fig.4 Different residual modules

设 CSP 与 Res\_new\_X 输入特征图的大小分别为  $H_i \times W_i \times C'_i$  和  $H_i \times W_i \times C_i$ , Conv2D 是卷积核大小为  $K \times K$  的标准卷积。Res\_new\_X 与 CSP 模块的参数

量比值为  $\frac{K \times K \times C_i + \frac{3}{2}C_o + \frac{25}{2}}{K \times K \times C'_i + C'_o \left( \frac{K \times K}{2} + 1 \right)}$ 。设输入和输出

通道相同,即  $C_i = C'_i$  且  $C_o = C'_o$ ,则输出通道越大,轻量化残差结构的参数量比 CSP 模块减少得越多。

本文采用 Res\_new\_X 模块替换 YOLOv4-Tiny 中的 3 个 CSP 模块,前 2 个使用卷积核大小为  $3 \times 3$  的 Res\_new\_1 模块,增强特征提取,第 3 个采用  $1 \times 1$  的 Res\_new\_2,减少参数量和计算量,加快推理速度,从而更好地实现口罩佩戴实时检测。

## 1.2 多层级交叉融合

自制数据集包含很多不同尺度的口罩目标,部

分图片中目标较集中,特征差异性较小,实验结果表明,Res\_new\_X 中引入了一定的检测误差,因此,单一尺度的卷积核无法适应多角度、多尺度变化的图片。浅层网络分辨率较大,包含较清晰的位置信息,深层特征则包含丰富的语义信息,不同尺度的特征层包含的特征信息不同,对不同大小的目标适应性较强。为了解决多角度、多尺度目标降低模型精度的问题,本文基于 HRNet<sup>[23]</sup> 的并行网络结构构建一种轻量化的多层级交叉融合模块 MCF: 将 3 种不同分辨率的子网以并行的方式进行连接;在每一个子网上通过上采样或下采样操作转换特征的大小;利用拼接操作将所有的特征信息进行融合,使得网络能够较好地利用深层语义信息。

MCF 模块仅增加极少的内存,却能很好地保留特征图的细粒度信息并大幅提升检测准确度,不同分辨率的特征图的融合过程如图 5 所示。

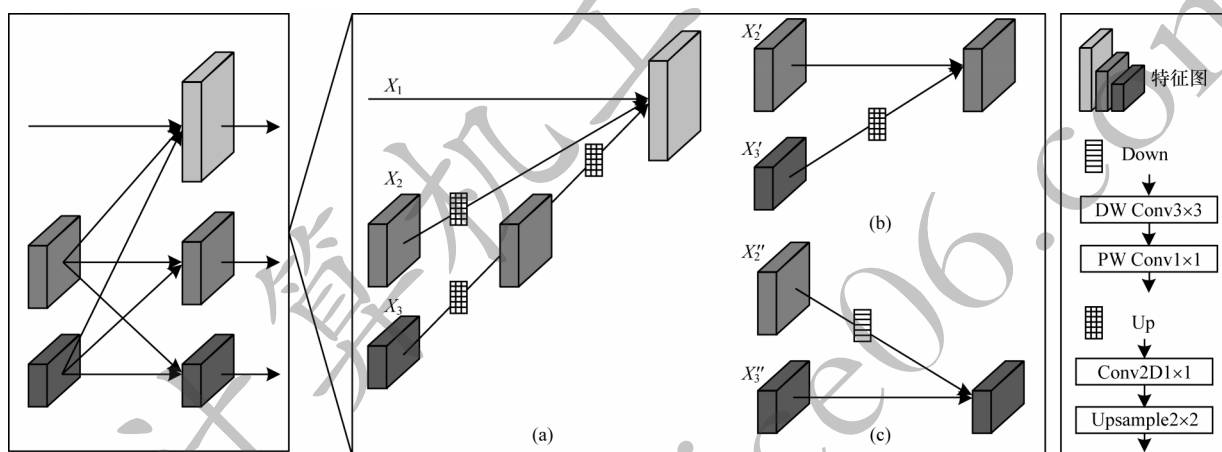


图 5 不同分辨率的特征图的融合过程

Fig.5 Fusion process of feature images with different resolutions

图 5 中的 (a)~(c) 分别表示  $52 \times 52$ 、 $26 \times 26$ 、 $13 \times 13$  分辨率的信息交互环节。首先,尽可能多地提取位置信息和语义信息;然后,通过一系列上采样或下采样以及拼接操作进行多尺度融合,从而提升多尺度特征的融合效果,增强检测性能;最终,建立一种多尺度特征融合的轻量化口罩佩戴检测网络。(a) 中  $X_1$ 、 $X_2$ 、 $X_3$  表示不同分辨率的特征图, $X_3$  先通过 2 倍上采样形成与  $X_2$  相同维度的特征图,再经过 Up 结构与  $X_2$  融合以输出新特征信息; $X_2$  通过上采样与  $X_1$  拼接。(b) 中  $X_3$  经过 Up 进行 2 倍上采样与  $X_2$  的特征进行拼接。(c) 对  $X_2$  经过 Down 进行 2 倍下采样并与  $X_3$  融合。完整的多层级交叉融合模块结构如图 2(b) 所示,最终生成 3 个不同尺度大小的特征图。其中,Up 是经过  $1 \times 1$  卷积再 2 倍邻近上采样的过程,Down 是由深度可分离卷积组成的 2 倍下采样,DW Conv 采用步长为 2 的  $3 \times 3$  卷积,再经过点卷积重新整合通道。

多层级交叉融合能够提取更多有利信息并有效进行信息交互,并且只增加少量的计算量,更适用于

存在目标遮挡、多目标以及目标较小的复杂场景。

## 2 实验结果与分析

### 2.1 实验设置

#### 2.1.1 实验平台

本文模型训练于 Ubuntu20.04 操作系统,使用 Keras 框架,CPU 为 Intel 酷睿 i7-8700K, GPU 为 NVIDIA GeForce GTX 1080Ti (11 GB),软件环境为 CUDA10.0、CuDNN7.6、Python3.6。输入图像统一缩放为  $416 \times 416$  的大小。

在训练过程中,采用 K-means 聚类得到新的先验框,分别为: (6, 10), (11, 20), (18, 31); (29, 47), (46, 69), (59, 121); (88, 101), (120, 172), (205, 253)。训练参数设置如下:初始学习率为  $1 \times 10^{-3}$ , Batch size 为 32,在第 50 个 epoch 之后,学习率下降到  $1 \times 10^{-4}$ , Batch size 为 16,在实验中使用早停法 (Early Stopping) 的训练方式避免模型过拟合。

YOLOv4-Tiny 与 L-MFFN-YOLO 的损失函数值随训练轮数的变化曲线如图 6 所示。

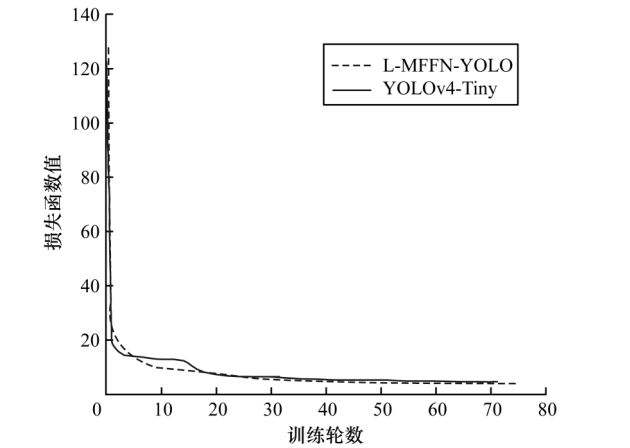


图6 YOLOv4-Tiny和L-MFFN-YOLO的损失函数对比

Fig.6 Comparison of loss function of YOLOv4-Tiny and L-MFFN-YOLO

2.1.2 数据集

考虑口罩类型、制造商、颜色等因素,本文通过手机拍摄、网络爬取等方法收集图片作为数据集,共计

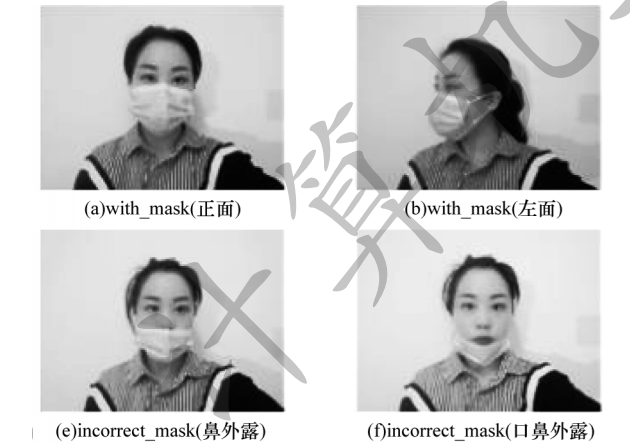


图7 数据集示例

Fig.7 Examples of the dataset

2.1.3 评价指标

为了验证本文算法的检测性能,将精确率  $P$  (Precision)和召回率  $R$  (Recall)作为定量评估指标,两者的计算公式分别如式(1)、式(2)所示:

$$P = \frac{T_p}{T_p + F_p} \tag{1}$$

$$R = \frac{T_p}{T_p + F_n} \tag{2}$$

其中:  $T_p$ 、 $F_n$ 和  $F_p$ 分别表示目标是正确的、目标被检测错误以及未被检测出的样本数量。

本文利用平均精确度 (Average Precision, AP)来评价模型在测试集上的检测性能,其计算如式(3)所示。多类别的检测结果通常采用平均精确度均值 (mean Average Precision, mAP)来衡量,其计算如式(4)所示。

$$A^{AP} = \int_0^1 P(R) dR \tag{3}$$

7 910张,每张图片中的人脸都对应一个标签,每个标签对应一个序号。将检测结果分为3类:序号0对应“佩戴口罩(with\_mask)”,表示已佩戴口罩;序号1对应“不正确佩戴口罩(incorrect\_mask)”,表示佩戴口罩不正确;序号2对应“人脸(face)”,表示未佩戴口罩。数据集中的不同类别样本分布如表1所示。

表1 数据集中的样本分布

Table 1 Sample distribution in dataset

类别	训练集	验证集	测试集
with_mask	1 956	251	554
incorrect_mask	1 908	249	505
face	1 752	212	523
total	5 616	712	1 582

图7(a)~图7(d)分别表示标准佩戴口罩的正面、左面、右面、上面示例,图7(e)~图7(h)分别展示鼻外露、口鼻外露、口鼻下巴外露、下巴外露4种常见的不正确的口罩佩戴方式。

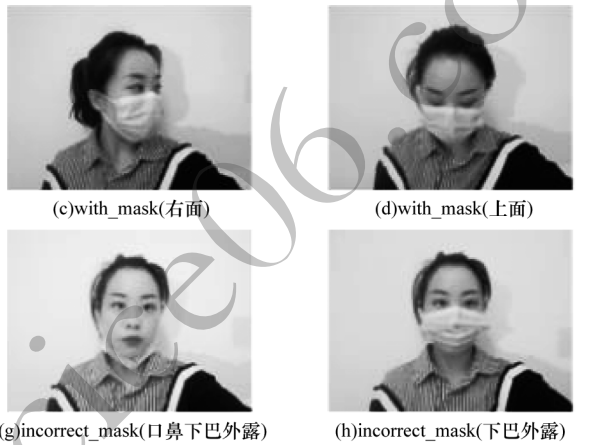


图7 数据集示例

Fig.7 Examples of the dataset

$$m^{mAP} = \frac{A^{AP}_{with\_mask} + A^{AP}_{incorrect\_mask} + A^{AP}_{face}}{3} \tag{4}$$

此外,本文采用每秒处理图片数量,即每秒帧率 (Frames Per Second, FPS)来衡量算法的检测速度。在面向一些内存受限的设备时,还需要评估模型的参数量、模型大小以及每秒10亿次的浮点运算量 (Billion Floating-point Operations Per second, BFLOPs)。

2.2 对比实验结果分析

2.2.1 检测算法模型参数分析

口罩佩戴检测系统需要快速高效地判断出人们是否佩戴口罩以及佩戴口罩是否正确,因此,更加适合部署在内存资源受限的设备上。为验证模型的有效性,实验对比分析不同主流检测模型的精度、参数量、检测速度、在 GPU 和 CPU 下的推理时间以及每秒10亿次的浮点运算量,结果如表2所示,最优结果加粗表示。

表 2 不同模型的性能对比结果

Table 2 Performance comparison results of different models

模型	mAP/%	输入大小/像素	参数量/ $10^6$	模型大小/MB	GPU 处理时间/ms	CPU 处理时间/ms	BFLOPs
Faster R-CNN with ResNet50 <sup>[24]</sup>	<b>87.59</b>	600×600	28.3	110.0	319.00	6 469.00	179.00
SSD-MobileNetv1 <sup>[25]</sup>	67.88	300×300	7.1	29.6	17.63	99.35	11.50
SSD-MobileNetv2 <sup>[22]</sup>	70.69	300×300	6.7	27.4	13.38	93.47	8.00
YOLOv3-Tiny <sup>[2]</sup>	75.73	416×416	8.7	33.2	7.46	86.34	5.57
YOLOv4-Tiny <sup>[4]</sup>	80.83	416×416	5.9	23.7	<b>6.82</b>	81.61	6.79
L-MFFN-YOLO	86.08	416×416	<b>1.4</b>	<b>5.8</b>	6.86	<b>67.21</b>	<b>4.13</b>

从表 2 可以看出:大型网络 Faster R-CNN<sup>[24]</sup> 的检测精度比本文 L-MFFN-YOLO 高 1.51 个百分点,但是其参数量和模型较大;用 MobileNetv1 和 MobileNetv2 替换 SSD 主干网络的 2 个一阶段目标检测器,在一定程度上可以降低计算复杂度,但是模型整体性能均低于本文 L-MFFN-YOLO;相较 YOLOv3-Tiny 和 YOLOv4-Tiny,在输入尺寸为 416×416 像素时,本文 L-MFFN-YOLO 的计算复杂度分别降低 26% 和 39%,模型大小和参数量也大幅减少,而精度分别提高 10.35 和 5.25 个百分点,在推理时间上,本文 L-MFFN-YOLO 在 CPU 上分别快约 19 ms 和 14 ms,在 GPU 上比 YOLOv3-Tiny 快 0.6 ms,但是比 YOLOv4-Tiny 慢 0.04 ms,主要原因是

L-MFFN-YOLO 计算层数提高,增加了一定的访问量,在 GPU 这种并行处理器下不能发挥出计算量较小的优势,但是延长时间较少,可以忽略不计。实验结果表明,本文 L-MFFN-YOLO 能降低设备资源内耗,并且在较少参数冗余的情况下具有较高的检测速度。

2.2.2 YOLO 类模型在口罩佩戴测试集上的表现

L-MFFN-YOLO 是在 YOLOv4-Tiny 的基础上进行的改进,因此,需要对不同的轻量化 YOLO 模型在口罩佩戴测试集上的检测结果进行对比,如表 3 所示。从表 3 可以看出,相较 YOLOv3-Tiny 和 YOLOv4-Tiny, L-MFFN-YOLO 在  $T_p$  中达到最大值,在  $F_N$  中达到最小值,其精确率和召回率也达到最佳,即对样本具有良好的检测性能。

表 3 不同的轻量化 YOLO 模型在口罩佩戴测试集上的检测结果

Table 3 Test results of different lightweight YOLO models on mask-wearing test set

模型	分类	输入大小/像素	目标数	$T_p$	$F_p$	$F_N$	$P/\%$	$R/\%$
YOLOv3-Tiny <sup>[2]</sup>	with_mask	416×416	286	484	54	70	89.96	87.36
	incorrect_mask		294	407	38	98	91.46	80.59
	face		297	416	68	107	85.95	79.54
	total		877	1 307	160	275	89.12	82.50
YOLOv4-Tiny <sup>[4]</sup>	with_mask	416×416	554	490	49	64	90.91	88.45
	incorrect_mask		505	421	32	84	92.94	83.37
	face		523	418	60	105	87.45	79.92
	total		1 582	1 329	141	253	90.43	83.91
L-MFFN-YOLO	with_mask	416×416	554	499	50	55	90.89	90.01
	incorrect_mask		505	432	21	73	95.36	85.54
	face		523	420	45	103	90.32	80.30
	total		1 582	<b>1 351</b>	<b>116</b>	<b>231</b>	<b>92.19</b>	<b>85.28</b>

为进一步直观地展示本文改进算法在口罩佩戴检测任务中的效果,分别对 YOLO 系列算法在常规、多人脸、遮挡及背景虚化、光线较弱等场景下进行测试,结果如图 8 所示(彩色效果见《计算机工程》官网 HTML 版)。在图 8(a)场景下,3 种方法均能实时检测出口罩佩戴情况,但是,YOLOv3-Tiny 和 YOLOv4-Tiny 的检测性能低于 L-MFFN-YOLO;在图 8(b)场景下,YOLOv3-Tiny 和 YOLOv4-Tiny 的漏检和错检情况严重,而 L-MFFN-YOLO 不仅能够解

决漏检问题,还在识别精度上有很大提高;在图 8(c)场景下,本文 L-MFFN-YOLO 由于加入了 MCF 结构,能够融合深层语义信息和浅层位置信息,因此对远处虚化目标的检测性能较好,有效降低了小目标漏检率;在图 8(d)场景下,YOLOv3-Tiny 漏检情况严重,YOLOv4-Tiny 虽然能识别大部分目标,但是检测精度低于 L-MFFN-YOLO。综上,本文提出的 L-MFFN-YOLO 算法能够在复杂场景中实现高效的目标检测。



图8 不同场景下的检测效果对比

Fig.8 Comparison of detection effects in different scenes

2.3 消融实验结果分析

2.3.1 不同残差模块的对比分析

1.1节分析了深度可分离卷积、CSP模块和CDC模块的参数数量,较少参数数量和较低计算复杂度的模型更适合部署在资源受限的设备中。将YOLOv4-Tiny中的CSP模块替换为由深度可分离卷积构建的残差结构和由CDC构建的Res\_new\_X,不同残差模块对算法性能的影响如表4所示。

表4 不同残差模块对算法性能的影响

Table 4 Influence of different residual modules on algorithm performance

模块	参数量/ $10^6$	BFLOPs	GPU 处理 时间/ms	mAP/%
深度可分离卷积 模块 <sup>[22]</sup>	5.60	5.79	6.323	<b>81.03</b>
CSP 模块 <sup>[4]</sup>	5.89	6.79	6.245	80.83
CDC 模块	<b>5.02</b>	<b>5.05</b>	<b>5.595</b>	80.96

从表4可以看出,改进模块精度较深度可分离卷积模块降低0.07个百分点,但是参数量和计算量降低最多,GPU处理时间较优,因此,其降低的精度可以忽略不计。

2.3.2 不同多尺度特征融合网络的对比分析

上文验证了由CDC模块构建的主干网络Res\_new\_X的有效性,本节将在Res\_new\_X的基础上,首先将FPN和PANet中的卷积操作替换成CDC模块,大幅降低模型参数量,然后再分析不同的多尺度融合模块对网络性能的影响。分别在原始主干网络Res\_new\_X中加入轻量化FPN结构(构成Res\_new\_X-A)、轻量化PANet(构成Res\_new\_X-B)以及多层级交叉融合MCF(构成Res\_new\_X-C),输入图片尺寸为 $416\times416$ ,将3种模型在口罩佩戴数据集上进行训练和测试并与原始主干网络作对比,实验结果如表5所示。从表5可以看出:Res\_new\_X-A的mAP较Res\_new\_X提高了3.23个百分点,表明多尺度特征融合在一定程度上可以更好地检测出口罩

佩戴情况;Res\_new\_X-B由于结合自上而下和自下而上的融合机制,导致网络的参数量增加,较Res\_new\_X-A,其FPS下降2,mAP提升不明显,其检测性能不具备优势;Res\_new\_X-C采用MCF模块,

通过重复融合不同分辨率的特征信息,很好地解决了多角度和多尺度目标剧烈变化而带来的精度下降问题,从而提高了检测性能,mAP达到86.08%,FPS的下降可忽略不计。

表5 不同多尺度融合网络对算法性能的影响

网络	FPS	参数量/10 <sup>6</sup>	AP/%			mAP/%
			with_mask	incorrect_mask	face	
Res_new_X	168	0.737	88.62	87.24	63.28	79.71
Res_new_X-A	156	0.997	92.80	89.66	66.36	82.94
Res_new_X-B	154	1.168	92.99	89.87	66.91	83.26
Res_new_X-C	150	1.381	94.64	92.23	71.38	86.08

2.3.3 不同模块对改进算法性能的影响

本文采用消融实验来分析不同模块结构对整个网络性能的影响,实验分为4组分别进行训练:

- 1)实验1,使用的主干网络为YOLOv4-Tiny CSPDarkNet53-Tiny<sup>[4]</sup>并去掉上采样;
- 2)实验2,使用的网络在实验1的基础上增加52×52尺度分支。
- 3)实验3,使用的主干网络用CDC模块构成的Res\_new\_X。
- 4)实验4,使用的网络在实验3的基础上加入多层级交叉融合模块MCF。

实验结果如表6所示,其中,“√”表示加入了该模块。

表6 改进模块对检测性能的影响

Table 6 Influence of improved module on detection performance							
实验分组	52×52分支	CDC	MCF	mAP/%	FPS	参数量/10 <sup>6</sup>	模型大小/MB
实验1				76.29	163	3.84	15.5
实验2	√			79.43	157	4.58	18.5
实验3	√	√		79.71	168	0.74	3.1
实验4	√	√	√	86.08	150	1.38	5.8

从表6可以看出:实验2的mAP比实验1提高了3.14个百分点,证明增加浅层预测分支能在较少计算开销的条件下增强算法的学习能力;实验3用CDC模块构成的Res\_new\_X替换原始主干网络,mAP达到79.71%,模型大小比实验2减少15.4 MB,从而验证了Res\_new\_X的有效性;实验4在实验3的基础上引入多层级交叉融合模块MCF,该模块使用卷积运算,较实验3增加了0.64×10<sup>6</sup>的参数量,减慢了检测速度,FPS有所下降,但其检测精度提高了6.37个百分点。

3 结束语

针对现有口罩佩戴检测算法网络结构复杂、计算参数冗余等问题,本文提出一种多尺度特征融合的轻量化口罩佩戴检测算法L-MFFN-YOLO,其可以实现多尺度、遮挡、多目标场景下的口罩佩戴实时检测,且适合部署于资源受限的检测设备。在YOLOv4-Tiny的基础上,通过轻量化主干特征提取网络缓解模型规模较大的问题。针对数据集图像大小不一的现象,新增一个高分辨率预测分支来提升小目标的检测能力。在此基础上,使用多层级交叉融合模块提高模型的检测精度,增强语义信息和位置信息的表达。实验结果表明,该算法的参数量与模型大小均较低,且检测精度较轻量级模型YOLOv4-Tiny提升5.25个百分点,其能够兼顾检测精度和速度,具有较好的工程应用价值。但是,在遮挡、背景虚化和光线较弱的场景中,有一些目标未被本文算法检测出,因此,下一步将在遮挡、能见度低的情况下,通过学习人脸眼睛、耳朵等细节特征来增加人脸定位,从而利用本文轻量化算法快速准确地完成口罩佩戴检测任务。

参考文献

[ 1 ] LIU Y, GAYLE A A, WILDER-SMITH A, et al. The reproductive number of COVID-19 is higher compared to SARS coronavirus[J]. Journal of Travel Medicine, 2020, 27(2): 1-4.

[ 2 ] REDMON J, FARHADI A. YOLOv3: an incremental improvement[EB/OL]. [ 2021-07-05 ]. <https://arxiv.org/abs/1804.02767>.

[ 3 ] 王艺皓,丁洪伟,李波,等. 复杂场景下基于改进YOLOv3的口罩佩戴检测算法[J]. 计算机工程, 2020, 46(11): 12-22.  
WANG Y H, DING H W, LI B, et al. Mask wearing detection algorithm based on improved YOLOv3 in complex scenes[J]. Computer Engineering, 2020, 46(11): 12-22. (in Chinese)

[ 4 ] WANG C Y, MARK LIAO H Y, WU Y H, et al. CSPNet: a new backbone that can enhance learning capability of CNN[C]//Proceedings of IEEE/CVF Conference on Computer

- Vision and Pattern Recognition. Washington D. C. , USA: IEEE Press, 2020; 1571-1580.
- [ 5 ] HE K M, ZHANG X Y, REN S Q, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.
- [ 6 ] LIU S, QI L, QIN H F, et al. Path aggregation network for instance segmentation[C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA: IEEE Press, 2018; 8759-8768.
- [ 7 ] 叶子勋, 张红英. YOLOv4口罩检测算法的轻量化改进[J]. 计算机工程与应用, 2021, 57(17): 157-168.
- YE Z X, ZHANG H Y. Lightweight improvement of YOLOv4 mask detection algorithm[J]. Computer Engineering and Applications, 2021, 57(17): 157-168. (in Chinese)
- [ 8 ] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection[EB/OL]. [ 2021-07-05 ]. <https://arxiv.org/abs/2004.10934>.
- [ 9 ] HOWARD A, SANDLER M, CHU G, et al. Searching for MobileNetv3 [C]//Proceedings of IEEE International Conference on Computer Vision. Washington D. C. , USA: IEEE Press, 2019; 1314-1324.
- [ 10 ] 曹城硕, 袁杰. 基于YOLO-Mask算法的口罩佩戴检测方法[J]. 激光与光电子学进展, 2021, 58(8): 211-218.
- CAO C S, YUAN J. Mask-wearing detection method based on YOLO-Mask[J]. Laser & Optoelectronics Progress, 2021, 58(8): 211-218. (in Chinese)
- [ 11 ] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA: IEEE Press, 2017; 936-944.
- [ 12 ] 余阿祥, 李承润, 于书仪, 等. 多注意力机制的口罩检测网络[J]. 南京师范大学学报(工程技术版), 2021, 21(1): 23-29.
- YU A X, LI C R, YU S Y, et al. Multi-attention mechanism of mask wearing detection network[J]. Journal of Nanjing Normal University(Engineering and Technology Edition), 2021, 21(1): 23-29. (in Chinese)
- [ 13 ] TAN M, PANG R, LE Q V, et al. EfficientDet: scalable and efficient object detection [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA: IEEE Press, 2020; 10781-10790.
- [ 14 ] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA: IEEE Press, 2018; 7132-7141.
- [ 15 ] BODLA N, SINGH B, CHELLAPPA R, et al. Soft-NMS—improving object detection with one line of code [C]//Proceedings of IEEE International Conference on Computer Vision. Washington D. C. , USA: IEEE Press, 2017; 5562-5570.
- [ 16 ] 彭成, 张乔虹, 唐朝晖, 等. 基于YOLOv5增强模型的口罩佩戴检测方法研究[J]. 计算机工程, 2022, 48(4): 39-49.
- PENG C, ZHANG Q H, TANG Z H, et al. Research on mask wearing detection method based on YOLOv5 enhancement model [J]. Computer Engineering, 2022, 48(4): 39-49. (in Chinese)
- [ 17 ] HAN K, WANG Y H, TIAN Q, et al. GhostNet: more features from cheap operations[C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA: IEEE Press, 2020; 1577-1586.
- [ 18 ] MA N N, ZHANG X Y, ZHENG H T, et al. ShuffleNet V2: practical guidelines for efficient CNN architecture design [EB/OL]. [ 2021-07-05 ]. <https://arxiv.org/pdf/1807.11164.pdf>.
- [ 19 ] 何涛, 俞舒曼, 徐鹤. 基于条件生成对抗网络与知识蒸馏的单幅图像去雾方法[J]. 计算机工程, 2022, 48(4): 165-172.
- HE T, YU S M, XU H. Single image dehazing method based on conditional generative adversarial network and knowledge distillation[J]. Computer Engineering, 2022, 48(4): 165-172. (in Chinese)
- [ 20 ] FAN S T, ZHANG X M, SONG Z H. Reinforced knowledge distillation: multi-class imbalanced classifier based on policy gradient reinforcement learning [J]. Neurocomputing, 2021, 463: 422-436.
- [ 21 ] 曹远杰, 高瑜翔. 基于GhostNet残差结构的轻量化饮料识别网络[J]. 计算机工程, 2022, 48(3): 310-314.
- CAO Y J, GAO Y X. Lightweight beverage recognition network based on GhostNet residual structure[J]. Computer Engineering, 2022, 48(3): 310-314. (in Chinese)
- [ 22 ] SANDLER M, HOWARD A, ZHU M L, et al. MobileNetV2: inverted residuals and linear bottlenecks[C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA: IEEE Press, 2018; 4510-4520.
- [ 23 ] SUN K, XIAO B, LIU D, et al. Deep high-resolution representation learning for human pose estimation [C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA: IEEE Press, 2019; 5686-5696.
- [ 24 ] 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.
- [ 25 ] HOWARD A G, ZHU M L, CHEN B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications[EB/OL]. [ 2021-07-05 ]. <https://arxiv.org/abs/1704.04861>.