钟露,韩韧. 融合 EfficientViT 和 Dyhead 注意力机制的口罩佩戴检测模型[J]. 智能计算机与应用,2026,16(1):157-163. DOI:10.20169/j.issn.2095-2163.24033101

融合 EfficientViT 和 Dyhead 注意力机制的口罩佩戴检测模型

钟 露,韩 韧

(上海理工大学 光电信息与计算机工程学院,上海 200093)

摘 要:为了提高复杂环境下的口罩佩戴检测的准确率,一个新的口罩佩戴检测模型 YOLOv5-Eff-Dyhead 被提出。首先,引入 EfficientVit 模块改进骨干网络,通过将特征图切片后,对每个切片应用注意力机制进行特征关注,可以解决复杂环境中口罩特征模糊和密集人群特征相似的问题。其次,使用 Dyhead 模块改进检测头。通过对口罩的空间位置、尺度大小和任务特征的关注,检测头能够有效地处理多尺度口罩信息,从而更好地完成定位和分类任务。实验结果表明:在口罩数据集 MNW-Face 上,YOLOv5-Eff-Dyhead 的性能较之基准模型 YOLOv5 有明显提高。与当前最优秀的模型相比,该模型具有最好的精度(mAP@0.5),达到了 95.3%。同时该模型具有较低的计算复杂度,使其能够在资源受限的情况下获得良好的性能。

关键词:YOLOv5;口罩佩戴检测;EfficientVit;Dynamic Head;复杂环境

中图分类号:TP391 **文献标志码**:A **文章编号**:2095-2163(2026)01-0157-07

Mask-wearing detection method combining EfficientViT and Dyhead attention mechanisms

ZHONG Lu, HAN Ren

(School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China)

Abstract: In order to improve the accuracy of mask-wearing detection in complex environments, a new mask-wearing detection model YOLOv5-Eff-Dyhead is proposed. Firstly, the EfficientVit module is introduced to improve the backbone network. By slicing the feature map and applying attention mechanisms to each slice for feature attention, the problem of blurry mask features and similar features in dense crowds in complex environments can be solved. Secondly, the Dyhead module is used to improve the detection head. By paying attention to the spatial position, scale size, and task characteristics of masks, the detection head can better process multi-scale mask information, thereby better completing localization and classification tasks. The experimental results show that YOLOv5-Eff-Dyhead has significantly improved performance compared to the benchmark model YOLOv5 on the mask dataset MNWFace. Compared to the current best model, this model has the best accuracy(mAP@0.5), reaching 95.3%. At the same time, the model has lower computational complexity, allowing it to achieve good performance in resource constrained situations.

Key words: YOLOv5; mask-wearing detection; EfficientVit; Dynamic Head; complex environment

0 引 言

在公共场所中,人们正确地佩戴口罩是有效防止呼吸道病毒传播的重要途径[1-5]。传统依靠人力在公共场所进行口罩佩戴检测的方式存在效率低下、错误率高和浪费人力资源等问题。而随着计算机视觉和人工智能的不断发展,通过口罩图像数据集训练出检测模型的技术线路已经成为当前的主要检测方式。但在复杂环境下,口罩佩戴的图像往往呈现密集人群、高曝光或者过度昏暗等低质问题,这也给现阶段的检测模型带来了不小的挑战。

目前,YOLOv5 在目标检测领域中的检测精度和模型大小方面的平衡性好。但在复杂场景下,特别是密集人群、高曝光和昏暗的检测环境中,其性能仍有不足。针对这种情况,一个基于 YOLOv5 的新的模型 YOLOv5-Eff-Dyhead 被提出,通过融入注意力机制来更好地解决检测精度低的问题。

作者简介:钟 露(1997—),男,硕士研究生,主要研究方向:机器学习,目标检测。Email:1600869935@qq.com;韩 韧(1980—),男,博士,副教授,主要研究方向:智能计算,边缘计算。

收稿日期:2024-03-31

1 相关研究

通常来说,检测口罩佩戴检测有 2 种算法:其中,一阶段算法和二阶段算法。其中,一阶段算法采用回归技术能直接对输入图像进行分类和定位,检测速度很快。代表性算法有:YOLO 系列、SSD 系列[6] 和 RetinaNet 系列[7]。如:He 等学者[8] 使用带有 DIoU_nms 和 α-CIoU 损失函数的 YOLOv5 模型来检测口罩,提高了遮挡重叠目标的检测准确性。Xu 等学者[9] 将 ShuffleNetV2 网络和坐标注意力相融合,改善了在公共场所进行口罩检测时面临的高精度和实时性能问题。二阶段算法则是将检测分为 2 个步骤:首先生成可能包含对象的候选区域,然后进一步对区域进行分类和校准以获得最终结果。该算法更注重检测精度,但检测速度相对较慢。代表性算法有:RCNN 和 Faster-RCNN[10]。如:Hanna 等学者[11] 使用 R-CNN 算法在 MaskedFace-Net 数据集上检测戴口罩的人脸。此外,引入 DS-Face 以提高小尺度目标的检测准确性,并使用 LHMD 算法增强了低分辨率目标的检测精度。李泽琛等学者[12] 在 Faster-RCNN 中引入了残差结构和注意力机制,构建了空间-通道注意力残差模块,进一步提高了模型对小物体的检测精度。

上述工作大多数都是针对单个人的图像进行检测,较少关注公共场所中人群的口罩佩戴检测方法。由于公共场所环境复杂,摄像头可能受到干扰,拍摄出的图像质量往往不佳。图像中的密集人群会出现重叠现象,口罩的尺度也会呈现出多种情况。这些因素都会影响检测算法的性能。

为了解决这些问题,提出一个口罩佩戴检测模型 YOLOv5-Eff-Dyhead。该模型能提高复杂环境下密集人群的口罩佩戴检测精度。对多种尺度的口罩有很好的检测性,对于过度曝光和昏暗的环境下也有高检测精度,能够在资源受限情况下高效地进行口罩佩戴检测。

2 用于复杂环境的口罩佩戴检测网络

2.1 YOLOv5-Eff-Dyhead 模型

二阶段算法适用于单目标检测。针对人群检测的情况,一阶段算法更适合多目标的检测。此外在一阶段模型中,考虑到 YOLO 系列具有检测速度和精度达到平衡的特点,因此采用 YOLOv5 作为基准模型进行改进并提出 YOLOv5-Eff-Dyhead 模型。该模型主要由骨干网络(Backbone)、颈部结构和头部结构(Detection Head)构成。具体网络结构如图 1 所示。

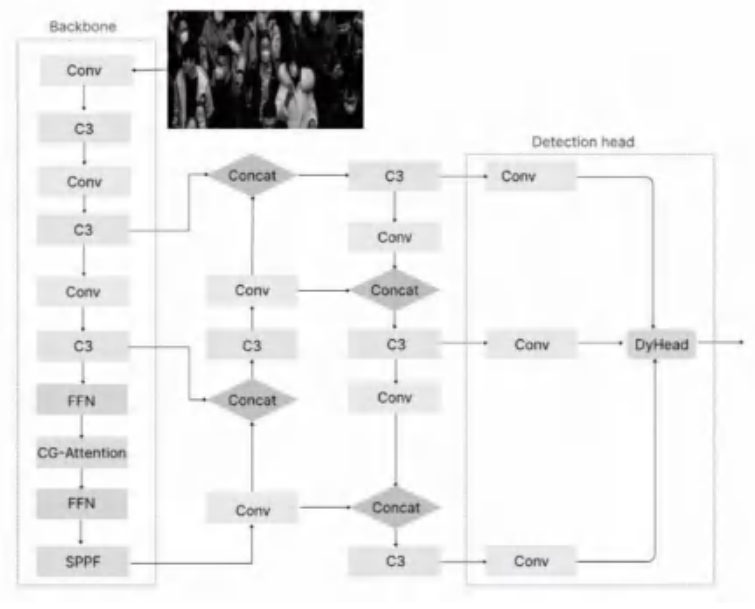

图 1 YOLOv5-Eff-Dyhead 网络结构图

Fig. 1 YOLOv5-Eff-Dyhead network structure diagram

骨干网络部分负责提取图像的特征信息。通过一系列的卷积操作(Conv)和 C3 模块来提取图像的不同层次的特征。其中,C3 模块由多个 Conv 构成,采用 ResNet 结构,可以增加网络的深度和感受野,从而提高特征提取能力。骨干网络中引入 EfficientVit[13] 的构建块去解决低质量图片的特征模

糊和重叠人群的特征相似带来的影响。

颈部结构由特征金字塔网络(FPN)和路径聚合网络(PAN)组成。其中,FPN 采用自顶向下的方式提取特征,PAN 则采用自底向上方式来强化特征提取。颈部结构通过多个卷积操作(Conv)和池化操作进一步提取图像的特征信息。将提取到的特征与骨干网络中提取的特征进行融合来增强特征信息。

头部结构负责对图像进行分类和定位。通过 3 个 Conv 提取不同尺度的口罩佩戴特征图。再通过 Dynamic Head[14](简称 Dyhead)模块进行口罩尺度、口罩空间位置和口罩任务三个维度的特征融合,进一步提高口罩佩戴的推理能力。

2.2　融合的 EfficientViT 模块

图 2 为 EfficientVit 模块结构图。模块采用“三明治”构建块,由 2 个前馈神经网络(FFN)和 1 个级联分组注意力模块(CG-Attention)构成。通过前馈神经网络对特征进行处理,随后通过 Split 对特征图进行切片,每个切片由自注意力机制进行特征提取,实现对特征图的多维度关注。Concat 块中进行通道数叠加操作,融合后的特征传递给第 2 个前馈网络进行处理。此外,级联分组注意力模块中采用特征金字塔结构。每层提取的特征和下一层切片进行融合再提取,增强了特征表达能力。

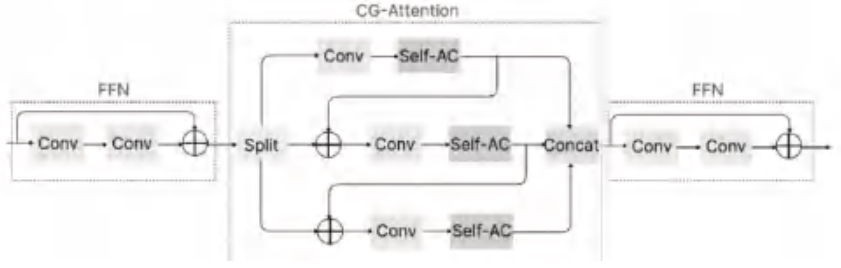

图 2　EfficientVit 模块

Fig. 2　EfficientVit module

骨干网络中融入 EfficientVit 模块能对输入的特征进行切片处理,输入到不同的注意力头部,实现了对口罩特征的多维度关注。即,通过每个注意力头部捕捉输入特征的不同方面,来增强检测模型对重叠部分特征的理解能力。

2.3　融合的 DyHead 模块

经过骨干网络后,冗余的背景特征被减少,但关键的口罩特征也可能被丢失。口罩的定位和分类的结果可能受到影响,从而导致检测精度下降。为了解决这个问题,DyHead 检测头被引入。该模块能增强口罩的尺度、空间位置和类别信息表达。Dyhead 的网络结构如图 3 所示,输入特征图经过尺度感知注意力块(Scale-Aware Attention)、空间感知注意力块(Spatial-Aware Attention)和任务感知注意力块(Task-Aware Attention)的一轮特征提取后,特征图再次送入尺度注意力块中进行第 2 轮的特征提取。重复多轮次特征提取过程后,模型的推理准确性得到提升。DAI 等学者[14]在实验中证明 DyHead 模循环次数(n)为 6 时有较好提升效果。

3 种注意力机制的提取特征过程如图 4 所示[14]。尺度感知注意力增强了对多种尺度口罩的感知能力。空间感知注意力加强了对环境中口罩的空间位置的感知能力。任务感知注意力增加了对通道维度的关注,增强了对不同口罩佩戴情况的感知能力。具体来说,图 4(a)中,尺度注意力通过平均池化提取最大特征值。使用 1×1 卷积对特征图的维度进行压缩。经过激活函数 Relu 和 Sigmoid 的非线性变换后,输出大小一致的特征图。图 4(b)中,空间注意力通过由步长控制(Index)、卷积块和 Offset 组成的可变形卷积[15]学习稀疏注意力。其中,Offset 是对特征图进行可变步长的卷积操作。通过 Offset 获取空间位置偏移特征。图 4(c)中,任务注意力用池化操作降低特征图的维度。此后经过 2 个全连接层(FC)和 1 个归一化层(Normalize),再经由 Sigmoid 函数将输出值限定在[-1,1]的范围内。

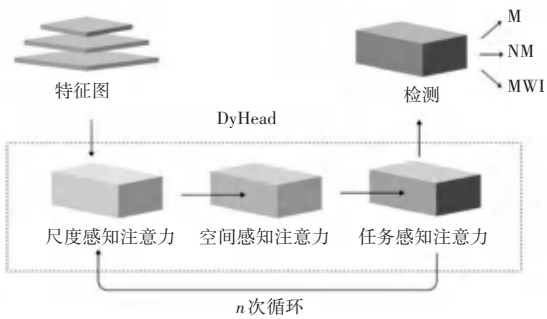

图 3　Dyhead 模块

Fig. 3　Dyhead module

检测头中融入 Dyhead 模块后,通过对图像中的复杂环境进行特征图上的尺度感知、空间位置感知和任务感知,可以更有效地推理与识别有效口罩和佩戴情况信息,并获得更准确的检测结果。

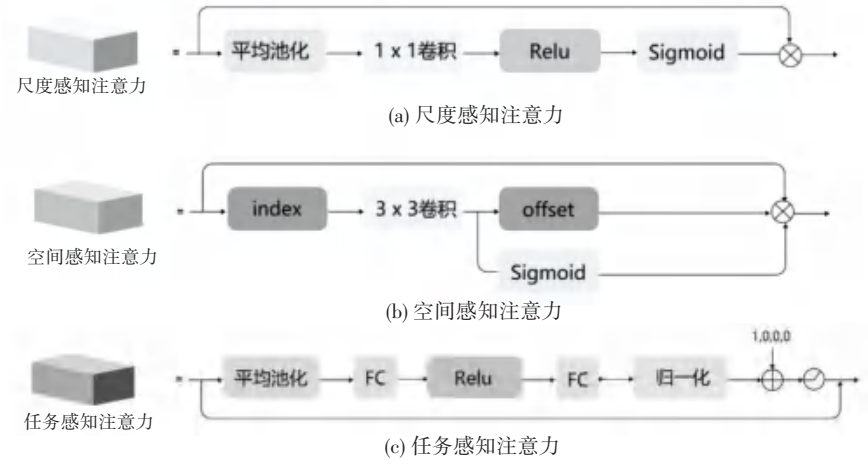

图 4 Dyhead 模块中 3 种注意力机制[14]

Fig. 4 Three attention mechanisms in the Dyhead module[14]

3 实验

3.1 数据集 MNWFace

MNWFace 数据集由在 Kaggle 上发布的 COVID-19 医学口罩检测数据集(MFMD)[16]和 Kaggle 上公开的口罩人脸数据集(FMD)[17]组成。考虑到不正确佩戴口罩的样本数量较少,在 MaskedFace-Net 数据集[18]中随机选择了 500 个不正确佩戴的口罩样本,添加到 MNWFace 中。新数据集的特征如图 5 所示。图 5(a)~(d)分别展示了 MNWFace 中的密集人群图像、拍摄角度和距离导致的多尺度口罩图像、以及曝光和昏暗的复杂环境图像。

图 5 MNWFace 数据集特征

Fig. 5 Characteristics of the MNWFace dataset

3.2 数据预处理

为防止过拟合,采用数据增强来增加数据集样本的数量和多样性。除了增加曝光值、降低亮度和添加噪声的预处理方式,又采用了边框增强策略[19],如图 6 所示。对边界框内检测到的对象进行颜色操作,包括改变曝光和亮度。实验[19]证明:在数据集较小时,采用边界框增强策略能实现了更多的改进,可以检测更小的对象并提高检测准确性。

图 6 边界框增强策略

Fig. 6 Bounding box augmentation strategy

3.3 检测指标

为了验证所提模型的性能,采用精确度(P)、召回率(R)、平均精度均值(mAP)来评估检测质量。同时,还使用了参数数量、浮点运算次数(FLOPs)和模型大小来检测模型的计算性能。

(1)精确度(P)。是预测为正样本中实际为正样本的数量,计算公式如下:

$$P = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{1}$$

(2)召回率(R)。是模型预测出的正样本数占所有正样本的比例,计算公式如下:

$$R = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{2}$$

其中,TP 表示预测和实际都是正样本;FP 表示预测是正样本但实际是负样本;TN 表示预测和实际

都是负样本；FN 表示预测是负样本但实际是正样本。

(3)平均精度(AP)。是衡量目标检测模型性能最常用的指标。AP 值越高，训练后的模型性能越好。AP 的计算公式如下：

$$AP = \int_0^1 P(r)\,dr \tag{3}$$

其中，$P(r)$ 表示召回率为 r 时的精确度。

(4)mAP。综合了模型在不同 IoU 阈值下的精度和召回率，通过对每个类别的 AP 值求和并取平均值得到。平均精度越高，说明模型的精确度越好，在本文的口罩检测任务中，mAP 的计算公式如下：

$$mAP = \frac{AP_M + AP_{NM} + AP_{MWI}}{3} \tag{4}$$

其中，M、NM 和 MWI 分别表示佩戴口罩、未佩戴口罩和不正确佩戴口罩的三种标签类别。

3.4　实验结果与分析

3.4.1　消融实验

为了证明 EfficientVit 模块和 Dyhead 模块对模型检测精度的提高，本文设计了消融实验。消融实验设计和结果见表 1。表 1 中，模型①代表基准模型 YOLOv5n；模型②代表在 YOLOv5n 的骨干网络部分融入 EfficientVit 模块；模型③代表在模型②中引入 Dyhead 模块，即 YOLOv5-Eff-Dyhead 模型。M、MWI 和 NM 分别代表佩戴口罩、未佩戴口罩和不正确佩戴口罩的 3 种标签类别。

表 1　消融实验设计和结果

Table 1　Ablation experiment design and results　　%

模型	模块			精确度			召回率			mAP@0.5
	YOLOv5n	EfficientVit	Dyhead	M	MWI	NM	M	MWI	NM	
①	√			93.8	96.9	88.5	92.7	93.0	80.7	92.8
②	√	√		94.5	95.5	89.9	94.7	94.1	84.2	94.0
③	√	√	√	96.2	98.2	90.9	94.9	94.2	85.4	95.3

对比模型①和模型②，发现融入 EfficientVit 注意力模块，能使得模型的检测精度 mAP@0.5 提高了 1.2 %。

对比模型②和模型③，证明引入 Dyhead 后，模型的 mAP@0.5 再次提升了 1.3 %。同时在 M、MWI 和 NM 上的准确度提升了 1.7%、2.7% 和 1.0%，召回率则提高了 0.2 %、0.1 %和 1.2 %。证明所提出的模型在复杂环境中对口罩佩戴有着高检测精度。

3.4.2　对比实验

为了进一步验证所提模型在口罩佩戴检测中的性能优势，将其与 2 类模型进行了比较。第一类是常见的轻量化模型：MobileNetV3 模型[20] 和 GhostNet 模型[21]。第二类是专注于检测精度的 CoordConv 模型[22]和 FasterNet[23]模型。

对比结果见表 2。由表 2 可知，与 MobileNetV3 和 GhostNet 相比，YOLOv5-Eff-Dyhead 在 mAP@0.5 的指标上分别提高了 6.9 %和 6.5 %。结果证明，在资源受限的情况下，YOLOv5-Eff-Dyhead 具备比经典的轻量级模型更高的检测精度。与 CoordConv 模型相比，在参数数量和计算浮点次数相接近的情况下，YOLOv5-Eff-Dyhead 在检测精度上表现出明显提高，具体表现为 mAP@0.5 提高了 1.8 %。而与 FasterNet 相比，YOLOv5-Eff-Dyhead 在计算浮点次数、参数数量和模型大小上分别减少了 36.2 %、23.8 %和 18.9%的同时，检测精度依旧和先进的 FasterNet 模型相同。

表 2　对比实验结果

Table 2　Comparison experiment results

模型	mAP@0.5/%	FLOPs/G	参数数量/M	模型大小/M
MobileNetV3	88.4	2.0	1.14	2.67
GhostNet	88.8	2.1	0.99	2.41
CoordConv	93.5	5.6	2.43	5.22
FasterNet	93.8	10.5	4.07	8.70
本文	95.3	6.7	3.10	7.06

实验结果证明，针对复杂环境下密集人群及多尺度口罩带来的影响，YOLOv5-Eff-Dyhead 不仅比优秀的 FasterNet 模型在 mAP@0.5 指标上达到了最高 95.3%，而且在模型计算开销中计算浮点次数和参数数量上只是略微提升。模型在实际资源受限的环境中也能有良好的表现。综上所述，YOLOv5-Eff-Dyhead 在口罩佩戴检测任务中展现出了较高的检测精度和较低的计算成本，为实际应用提供了有

价值的解决方案。

3.4.3 改进前后检测效果对比

为更直观地评价改进后的效果，在 MNWFace 数据集的测试集中进行测试。实验结果如图 7 所示。对比图 7(a)、图 7(b)可以看出，YOLOv5 在昏暗场景下存在漏检，没有将位于正中央的佩戴口罩情况检测到。对比图 7(c)、图 7(d)可以看出，YOLOv5 在昏暗密集人群的情况下存在错检，将图中的汽车错误检测为佩戴口罩的情况。对比图 7(e)、图 7(f)可以看出，YOLOv5 在曝光场景下存在漏检，没有将佩戴口罩的情况检测出来。对比图 7(g)、图 7(h)可以看出，YOLOv5 在曝光密集人群的情况下存在漏检，没有将正确佩戴口罩的情况检测出来。从图 7 的实验结果可以看到，改进后的模型的检测准确率更高，很好地解决口罩佩戴检测中错检和漏检问题。

(a) YOLOv5 昏暗场景 (b) 本文的昏暗场景

(c) YOLOv5 昏暗密集场景 (d) 本文的昏暗密集场景

(e) YOLOv5 曝光场景 (f) 本文的曝光场景

(g) YOLOv5 曝光密集场景 (h) 本文的曝光密集场景

图 7 实验结果

Fig. 7 Experimental results

4 结束语

为了提升复杂环境下的口罩检测模型的精度，本文提出了一种基于改进 YOLOv5 的新模型，引入了 EfficientVit 注意力机制到骨干网络中，提高了模型在特征提取阶段的准确性。同时，引入了 Dyhead 注意力模块到检测头中，让模型能准确地实现口罩佩戴的定位和分类检测。最终，提高了其对密集人群、高曝光和昏暗环境下的检测精度，解决了口罩佩戴检测中的漏检和错检问题。由实验结果可知，改进后的模型的 mAP@0.5 比目前优秀的高精度检测网络 FasterNet 高 1.5 %。在之后的研究中，可以扩大和融入更多光照强度和密集程度的口罩佩戴数据，进一步研究模型在不同场景和环境下对口罩佩戴情况的检测性能。

参考文献

[1] HOWARD J, HUANG A, LI Z, et al. An evidence review of face masks against COVID - 19 [J]. Proceedings of the National Academy of Sciences, 2021, 118: e2014564118.

[2] BUNDGAARD H, BUNDGAARD J S, TODSEN T, et al. Effectiveness of adding a mask recommendation to other public health measures to prevent SARS-CoV-2 infection in Danish mask wearers: A randomized controlled trial [J]. Annals of Internal Medicine, 2021, 174(3): 335-343.

[3] LEUNG N H L, CHU D K W, SHIU E Y C, et al. Respiratory virus shedding in exhaled breath and efficacy of face masks [J]. Nature Medicine, 2020, 26(5): 676-680.

[4] BROOKS J T, BUTLER J C, REDFIELD R R. Universal masking to prevent SARS-CoV-2 transmission: The time is now [J]. JAMA, 2020, 324(7): 635.

[5] CHENG Yafang, MA Nan, WITT C, et al. Face masks effectively limit the probability of SARS-CoV-2 transmission [J]. Science, 2021, 372(6549): 1439-1443.

[6] LIU Wei, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector [M]//LEIBE B, MATAS J, SEBE N, et al. Computer Vision. Lecture Notes in Computer Science. Cham: Springer, 2016, 9905: 21-37.

[7] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection [C]// Proceedings of 2017 IEEE International Conference on Computer Vision (ICCV). Piscataway, NJ: IEEE, 2017: 2999-3007.

[8] HE W, LI H. Mask recognition based on improved YOLOv5 target detection algorithm [C]//Proceedings of 2022 International Conference on Computing, Communication, Perception and Quantum Technology (CCPQT). Piscataway, NJ: IEEE, 2022: 361-366.

[9] XU Sheng, GUO Zhanyu, LIU Yuchi, et al. An improved lightweight YOLOv5 model based on attention mechanism for face mask detection [M]//PIMENIDIS E, ANGELOV P, JAYNE C, et al. Artificial Neural Networks and Machine Learning. Cham: Springer, 2022: 531-543.

[10] REN Shaoqing, HE Kaiming, 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.

[11] HANNA M A, ALHARBI A H, ALGHAMDI N S. A framework for mask-wearing recognition in complex scenes for different face sizes[J]. Intelligent Automation & Soft Computing, 2022, 32(2): 1153-1165.

[12] 李泽琛, 李恒超, 胡文帅, 等. 多尺度注意力学习的 Faster R-CNN 口罩人脸检测模型[J]. 西南交通大学学报, 2021, 56(5): 1002-1010.

[13] LIU X Y, PENG H W, ZHENG N X, et al. EfficientViT: Memory efficient vision Transformer with cascaded group attention[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2023: 14420-14430.

[14] DAI Xiyang, CHEN Yinpeng, XIAO Bin, et al. Dynamic head: Unifying object detection heads with attentions[C]//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2021: 7369-7378.

[15] DAI Jifeng, QI Haozhi, XIONG Yuwen, et al. Deformable Convolutional Networks [C]//Proceedings of 2017 IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE, 2017: 764-773.

[16] KAGGLE. COVID-19 medical face mask detection dataset | Kaggle[EB/OL]. [2023-05-31]. https://www.kaggle.com/datasets/mloey1/medical-face-mask-detection-dataset.

[17] KAGGLE. Face mask detection[EB/OL]. [2023-05-31]. https://www.kaggle.com/datasets/andrewmvd/face-mask-detection.

[18] CABANI A, HAMMOUDI K, BENHABILES H, et al. MaskedFace-Net: A dataset of correctly/incorrectly masked face images in the context of COVID-19[J]. Smart Health, 2021, 19: 100144.

[19] ZOPH B, CUBUK E D, GHIASI G, et al. Learning data augmentation strategies for object detection[M]//VEDALDI A, BISCHOF H, BROX T, et al. Computer Vision. Cham: Springer, 2020: 566-583.

[20] HOWARD A, SANDLER M, CHEN B, et al. Searching for MobileNetV3[C]//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Piscataway, NJ: IEEE, 2019: 1314-1324.

[21] HAN Kai, WANG Yunhe, TIAN Qi, et al. GhostNet: More features from cheap operations[J]. arXiv preprint arXiv, 1911.11907, 2020.

[22] HOU Qibin, ZHOU Daquan, FENG Jiashi. Coordinate attention for efficient mobile network design[C]//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2021: 13708-13717.

[23] CHEN Jierun, KAO S H, HE Hao, et al. Run, Don't walk: Chasing higher FLOPS for faster neural networks [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2023: 12021-12031.