基于改进 YOLOv5 的小目标检测算法

郭 磊[1*]，王邱龙[2]，薛 伟[2]，郭 济[3]

(1. 电子科技大学计算机科学与工程学院 成都 611731；2. 新疆大学信息科学与工程学院 乌鲁木齐 830000；
3. 西藏民族大学财经学院 陕西 咸阳 712082)

【摘要】针对目标检测中小目标误检、漏检及特征提取能力不足等问题，提出一种基于改进 YOLOv5 的小目标检测算法。该算法使用 Mosaic-8 方法进行数据增强，通过增加一个浅层特征图、调整损失函数，来增强网络对小目标的感知能力；通过修改目标框回归公式，解决训练过程中梯度消失等问题，提升了小目标的检测精度。将改进后的算法应用在密集人群情景下的防护面具佩戴检测中，实验结果表明，相较于原始 YOLOv5 算法，该算法在小目标检测上具有更强的特征提取能力和更高的检测精度。

关 键 词 数据增强； 深度学习； 小目标检测； YOLOv5
中图分类号 TP39 **文献标志码** A doi:10.12178/1001-0548.2021235

A Small Object Detection Algorithm Based on Improved YOLOv5

GUO Lei[1*], WANG Qiulong[2], XUE Wei[2], and GUO Ji[3]

(1. School of Computer Science and Engineering, University of Electronic Science and Technology of China Chengdu 611731;
2. School of Information Science and Engineering, Xinjiang University Urumqi 830000;
3. College of Finance and Economics, Xizang Minzu University Xianyang Shanxi 712082)

Abstract For object detection, one immediate problem is the insufficiency of feature extraction on small objects, which is easy to make false detection and miss the inspection on small targets. To solve the problem of small object detection, an improved detection algorithm based on YOLOv5 was proposed. The algorithm uses the method of Mosaic-8 on data augmentation. A shallow feature map is added to the YOLOv5 network and loss function is adjusted to improve the sensibility of network on small targets. The target box regression formula is modified to solve the problem of gradient disappearance in training process, which realized accurate precision on small targets. The improved algorithm is applied to mask wearing detection under crowed environment. Experimental results show that the proposal algorithm has stronger feature extraction ability and higher detection accuracy on small object detection compared to the original YOLOv5 algorithm.

Key words data augmentation; deep learning; small object detection; YOLOv5

随着人工智能理论和深度学习技术的深入研究，作为计算机视觉领域核心问题之一的目标检测技术也取得了较大进展，已被应用于人脸检测[1]、智慧医疗[2-3]、行人检测[4]、活动识别[5]等。目标检测是利用图像处理、深度学习等技术，从图像或视频中定位感兴趣的对象，通过目标分类判断输入图像中是否包含目标，用目标定位找出目标物体位置并框出目标，其任务是锁定图像中的目标，定位目标位置、确定目标类别。作为图像及视频理解的基石，目标检测是解决图片分割、目标跟踪、图像描述、事件检测和场景理解等更高层次视觉任务的基础。

小目标检测长期以来是目标检测中的重点和难点之一。由于小目标具有图像覆盖面积较少、分辨率不足、位置缺乏准确性、特征表达不充分的特点，因而相对于常规目标小目标检测更困难。目标检测中对小目标的定义通常有两种：1) 国际光学工程学会对小目标的定义，将 256×256 像素的图像中成像点小于 80 个像素点 (即目标所占的像素点数与原图总像素点数的比例小于 0.12%) 的目标定义为

收稿日期：2021－08－30；修回日期：2021－11－23
基金项目：国家重点研发计划 (2018YFC0831800)
作者简介：郭磊 (1971－)，男，博士，副教授，主要从事机器学习、视频理解、嵌入式系统方面的研究.
*通信作者：郭磊，E-mail：leiguo@uestc.edu.cn

小目标；2) 根据具体的数据集对小目标进行定义，如在 COCO 数据集[6] 中，将尺寸小于 32×32 像素的目标定义为小目标；文献 [7] 在其交通标志数据集中，将宽度占整个图像比例小于 20% 的目标定义为小目标。一般而言，常规目标特征表达充分，位置准确明了，而小目标的分辨率相对较低，特征表达会相对缺乏。

1 相关工作

传统的目标检测算法通常由人工提取目标的特征，检测精度低、效果不好。随着深度学习的发展和硬件设备算力的提升，基于深度学习的卷积神经网络 (convolutional neural network, CNN)[8] 崭露头角，人们开始利用卷积神经网络自动提取图像中的特征并将其应用在目标检测中，极大地提升了目标检测效果。目前最通用的两个方法是以 R-CNN (region-based CNN) 系列为代表的基于候选框的两阶段深度学习算法和以 YOLO(you only look once) 系列为代表的基于回归的单阶段深度学习目标检测算法。

R-CNN 模型[9] 使用 CNN 提取的特征替代传统视觉特征，并采用大样本的有监督预训练与小样本微调的方式解决模型的过拟合问题，使得模型的检测性能有了较大的提升，但 R-CNN 需对每个候选区域的 CNN 特征进行大量重复计算。SPP-Net 网络[10] 能产生固定大小的输出，而与输入图像大小无关；由于输入尺寸的灵活性，使得 SPP-Net 能够提取多个尺度下的特征，且一张图片中候选区域的 CNN 特征只需要计算一次，在很大程度上能够节省计算资源。在 SPP-Net 的基础上，文献 [11-12] 先后提出 Fast R-CNN 和 Faster R-CNN 模型。从 R-CNN 模型发展到 Fast R-CNN 模型，进一步发展到 Faster R-CNN 模型，检测速度不断提高，检测精度也不断增强，但与单阶段目标检测算法在检测速度上相比，仍具有一定差距。

YOLO 系列算法和单点多盒检测器 (single shot multibox detector, SSD) 是典型的单阶段目标检测算法。文献 [13] 提出了第一个单阶段目标检测算法 YOLO，与 YOLO 最后采用全连接层提取检测结果不同，SSD[14] 使用不同尺度的特征图来做检测，并直接使用卷积提取检测结果。文献 [15-16] 在 YOLOv1 的基础上继续改进，又提出了 YOLOv2 和 YOLOv3 检测算法，其中 YOLOv2 进行了多种尝试，使用了批标准化 (batch normalization, BN) 技术，引入了锚框机制；YOLOv3 采用 darknet-53 作为骨干网络，并且使用了 3 种不同大小的锚框，在逻辑分类器中使用 sigmoid 函数把输出约束在 0～1 之间，使得 YOLOv3 拥有更快的推理速度。文献 [17] 在传统的 YOLO 基础上，加入了一些实用的技巧，提出了 YOLOv4 算法，将 BackBone 骨干网络中的 ReLU 激活函数改为 Mish 激活函数，与 ReLU 相比，Mish 函数图像更加平滑，实现了检测速度和精度的最佳权衡。从 YOLOv1 至今，YOLO 系列已经发展到了 YOLOv5，YOLOv5 融合了先前版本的优点，在检测速度和精度上都更胜一筹，在某种程度上 YOLOv5 已经成为 YOLO 系列算法中的 SOTA(State Of The Art)。

YOLOv5 是一个高性能、通用的目标检测模型，能一次性完成目标定位与目标分类两个任务，因此选择 YOLOv5 作为目标检测的基本骨架是可行的。但是为了实现一些场景下对小目标的独特性检测，就需要对 YOLOv5 的网络结构进行相应的调整和改进。

2 改进的 YOLOv5 算法

本文在 YOLOv5 网络的基础上进行改进，改进后的整体的网络结构如图 1 所示，通过新增尺寸为输入图像尺寸四分之一的特征图来提升对小目标特征的挖掘，采用多尺度反馈以引入全局上下文信息来提升对图像中小目标的识别能力。损失函数使用 CIOU[18]，从重叠面积、中心点距离、长宽比 3 个方面更好地描述目标框的回归。在原始 YOLOv5 的基础上使用 Mosaic-8 数据增强，修改目标框回归的公式，提高模型的收敛精度。下面分别从 Mosaic-8 数据增强、特征提取器、损失函数和目标框回归 4 个方面进行详细介绍。

2.1 Mosaic-8 数据增强

要获得一个表现良好的神经网络模型，往往需要大量的数据作支撑，然而获取新的数据这项工作往往需要花费大量的时间与人工成本，因此数据增强应运而生。使用数据增强技术，可以充分利用计算机来生成数据，增加数据量，如采用缩放、平移、旋转、色彩变换等方法增强数据，数据增强的好处是能够增加训练样本的数量，同时添加合适的噪声数据，能够提高模型的泛化力。

在 YOLOv5 中除了使用最基本的数据增强方法外，还使用了 Mosaic 数据增强方法，其主要思想就是将 4 张图片进行随机裁剪、缩放后，再随机排列拼接形成一张图片，实现丰富数据集的同时，

增加了小样本目标，提升网络的训练速度。在进行归一化操作时会一次性计算 4 张图片的数据，因此模型对内存的需求降低。Mosaic 数据增强的流程如图 2 所示。

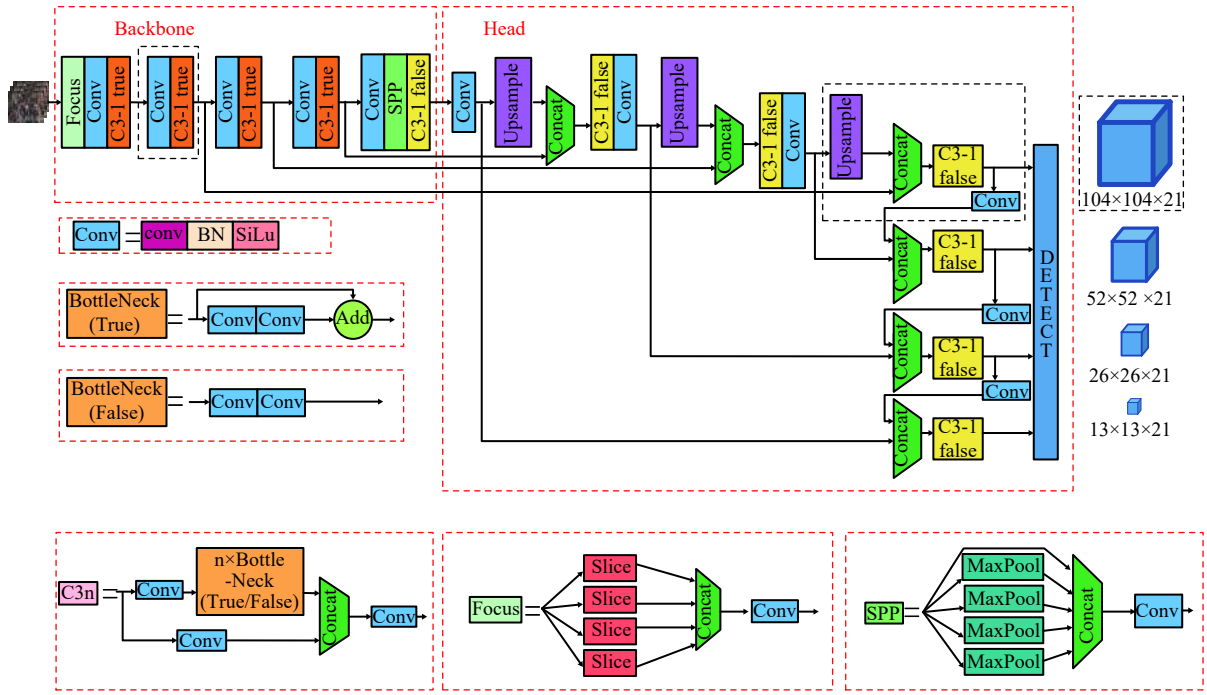


图 1　整体网络结构

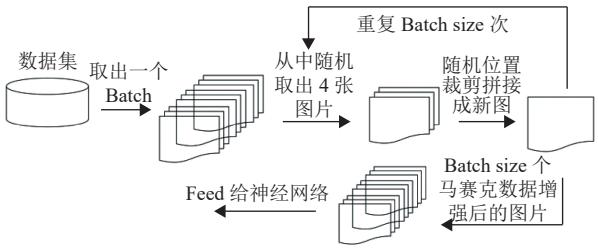


图 2　Mosaic 数据增强流程

本文受 Mosaic 思想的启发，采用 Mosaic 方法的增强版——Mosaic-8，即采用 8 张图片随机裁剪、随机排列、随机缩放，然后组合成一张图片，以此来增加样本的数据量，同时合理引入一些随机噪声，增强网络模型对图像中小目标样本的区分力，提升模型的泛化力，其细节如图 3 所示。

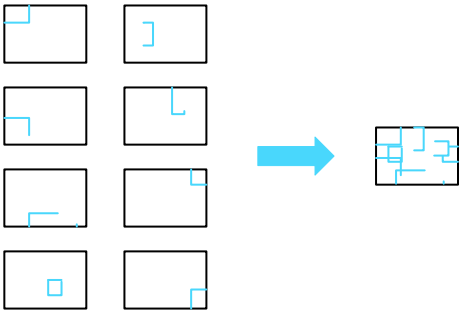


图 3　Mosaic-8 数据增强细节

2.2　特征提取器

在原始 YOLOv5 骨干网络中，使用 3 种不同尺寸的特征图来检测不同大小的目标，如图 4 所示，该网络将原始输入图像通过 8 倍下采样、16 倍下采样、32 倍下采样得到 3 种不同尺寸大小的特征图，将其输入到特征融合网络中。根据特征金字塔网络 (feature pyramid network, FPN)[19] 的思想可以看出，经过深层次卷积后的特征图虽然拥有丰富的语义信息，但在多次卷积的过程中会丢失掉目标的一些位置信息，不利于小目标的检测；而浅层卷积后得到的特征图语义信息虽然不够丰富，但目标的位置信息却比较精确。

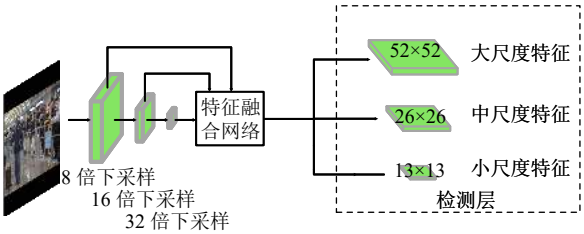


图 4　原始 YOLOv5 特征提取模型

在密集人群的条件下，大部分人脸检测目标占整幅图像的比例较小。因此，本文在 YOLOv5 骨干网络的基础上对原始输入图片增加一个 4 倍下采样的过程，如图 5 所示。原始图片经过 4 倍下采样后送入到特征融合网络得到新尺寸的特征图，该特

征图感受野较小，位置信息相对丰富，可以提升检测小目标的检测效果。

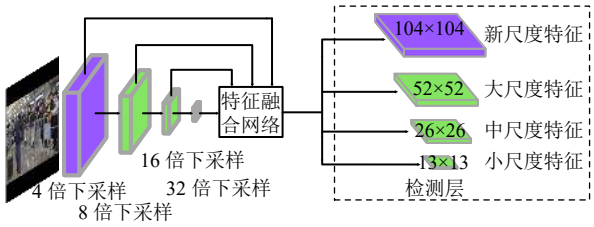


图 5 改进后的特征提取模型

在卷积神经网络中，经过不同的卷积层得到的特征图包含不同的目标特征信息。浅层卷积后得到的特征图分辨率较高，目标位置信息相对丰富，但语义信息不明显；深层卷积后得到的特征图分辨率低，语义信息丰富，但丢失了较多的目标位置信息。因此，浅层特征图能区分较为简单的目标，深层特征图能区分复杂的目标，将浅层特征图与深层特征图进行信息融合更有利于目标的区分。如图 6 所示，将特征金字塔网络与路径聚合网络 (path aggregation network, PAN)[20] 相结合，特征金字塔网络自顶向下传递深层次语义特征，路径聚合网络自底向上传递目标的位置信息，通过自顶向下和自底向上的特征信息融合有利于模型更好地学习特征，增强模型对小目标的敏感度。

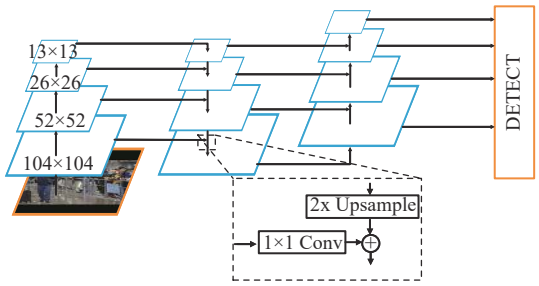


图 6 改进后的特征融合网络

2.3 损失函数

原始 YOLOv5 损失函数如式 (1) 所示，由定位损失、置信度损失和类别损失 3 部分构成。其中置信度损失和类别损失采用二元交叉熵损失函数进行计算：

$$\mathrm{Loss}_{\mathrm{object}} = \mathrm{Loss}_{\mathrm{loc}} + \mathrm{Loss}_{\mathrm{conf}} + \mathrm{Loss}_{\mathrm{class}} \tag{1}$$

$$\mathrm{Loss}_{\mathrm{loc}} = 1 - \mathrm{GIoU}$$

$$\mathrm{Loss}_{\mathrm{conf}} = -\sum_{i=0}^{K\times K} I_{ij}^{\mathrm{obj}} \left[\hat{C}_i^j \log C_i^j + \left(1-\hat{C}_i^j\right)\log\left(1-C_i^j\right)\right] - \lambda_{\mathrm{noobj}} \sum_{i=0}^{K\times K}\sum_{j=0}^{M} I_{ij}^{\mathrm{noobj}} \left[\hat{C}_i^j \log C_i^j + \left(1-\hat{C}_i^j\right)\log\left(1-C_i^j\right)\right]$$

$$\mathrm{Loss}_{\mathrm{class}} = -\sum_{i=0}^{K\times K} I_{ij}^{\mathrm{obj}} \sum_{c\in\mathrm{classes}} \left[\hat{P}_i^j \log P_i^j + \left(1-\hat{P}_i^j\right)\log\left(1-P_i^j\right)\right]$$

式中，K 表示网络最后输出的特征图划分为 $K\times K$ 个格子；M 表示每个格子对应的锚框的个数；I_{ij}^{obj} 表示有目标的锚框；I_{ij}^{noobj} 表示没有目标的锚框；λ_{noobj} 表示没有目标锚框的置信度损失权重系数。如图 7 所示，黑色框是真实框，记为 GT，灰色框是预测框，记为 P，外框是同时包裹真实框和预测框的最小框，记为 C。其中 c 是外框对角线的长度，d 是真实框中心点与预测框中心点的长度。

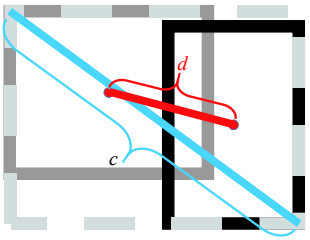


图 7 预测框 P 与真实框 GT

原始 YOLOv5 中使用 GIoU[21] 来计算定位损失：

$$\mathrm{GIoU} = \mathrm{IoU} - \frac{|\mathrm{C}-\mathrm{GTUP}|}{|\mathrm{C}|} = \frac{|\mathrm{PIGT}|}{|\mathrm{PUGT}|} - \frac{|\mathrm{C}-\mathrm{GTUP}|}{|\mathrm{C}|} \tag{2}$$

与原始 IoU 不同，GIoU 不仅关注真实框与预测框之间的重叠面积，还关注其他的非重叠区域，因此 GIoU 相较于原始 IoU 能更好的反应两者之间的重合度，但 GIoU 始终只考虑真实框与预测框之间的重叠率这一个因素，不能很好地描述目标框的回归问题。如图 8 所示，当预测框在真实框内部时，且预测框的大小相同时，此时 GIoU 会退化为 IoU，无法区分各个预测框之间的位置关系。

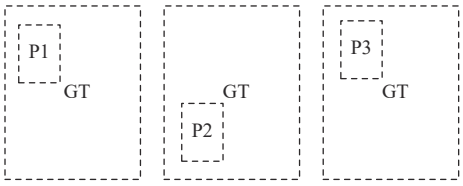


图 8 GIoU 退化为 IoU 示例

本文选择 CIoU 替代 GIoU 作为目标框回归的损失函数，其计算式为：

$$\mathrm{CIoU} = \mathrm{IoU} - \frac{\rho^2(b,b^{gt})}{c^2} - \alpha v \tag{3}$$

$$\alpha = \frac{v}{(1-\mathrm{IoU})+v} \tag{4}$$

$$v=\frac{4}{\pi^2}\left(\arctan\frac{w^{gt}}{h^{gt}}-\arctan\frac{w}{h}\right)^2 \tag{5}$$

式中，α是一个平衡参数，不参与梯度计算；v是用来衡量长宽比一致性的参数。CIoU 综合考虑了真实框与预测框之间的重叠率、中心点距离、长宽比，使得目标框回归过程中更加稳定，收敛的精度更高。

2.4　目标框回归

目标框回归的目的是要寻找某种映射关系，使得候选目标框 (region proposal) 的映射无限接近于真实目标框 (ground truth)。对真实目标框的预测是采用相对位置的方式回归出目标框相对于某个网格左上角的相对坐标。先验框与预测框的关系如图 9 所示，其中，虚线框表示先验框，实线框表示预测框。预测框通过先验框平移缩放得到。将原始图片根据特征图尺寸划分成$S\times S$个网格单元，每个网格单元会预测 3 个预测框，每个预测框包含 4 个坐标信息和 1 个置信度信息。当真实框中某个目标中心坐标落在某个网格中时，就由该网格预测这个目标。

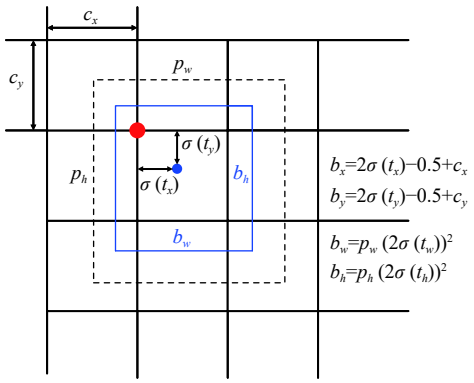


图 9　目标框回归原理图

目标框的坐标预测计算公式为：

$$b_x=2\sigma(t_x)-0.5+c_x \tag{6}$$

$$b_y=2\sigma(t_y)-0.5+c_y \tag{7}$$

$$b_w=p_w(2\sigma(t_w))^2 \tag{8}$$

$$b_h=p_h(2\sigma(t_h))^2 \tag{9}$$

$$P_r(\text{object})\times\text{IOU}(b,\text{object})=\sigma(t_o) \tag{10}$$

式中，t_x、t_y、t_w、t_h为偏移；σ表示 Sigmoid 激活函数，用于将网络预测值t_x、t_y、t_w、t_h映射到[0,1]之间；c_x、c_y是单元网格中相对于图片左上角的偏移量；p_w、p_h是先验框宽高；b_x、b_y和宽高b_w、b_h为预测目标框的中心坐标；$\sigma(t_o)$是预测框的置信度，由预测框的概率和预测框与真实框的 IoU 值相乘得到。对$\sigma(t_o)$设定阈值，过滤掉置信度较低的预测框，然后再对剩下的预测框用非极大值抑制算法 (non-maximum suppression, NMS)[22] 得到最终的预测框。

在最小的特征图上，由于其感受野最大，故应该用其来检测大目标，所以大尺度的特征图应该应用小尺寸的先验框，小尺寸的特征图应该应用大尺度的先验框来进行预测框的回归。本文采用 4 尺度检测结构，4 个尺度的特征图大小与先验框尺寸的对应关系如表 1 所示。

表 1　特征图大小与先验框尺寸对应关系

特征图大小	先验框尺寸		
13×13	[116,90]	[156,198]	[373,326]
26×26	[30,61]	[62,45]	[59,119]
52×52	[10,13]	[16,30]	[33,23]
104×104	[5,6]	[8,14]	[15,11]

3　实验与结果分析

将改进后的算法应用在密集人群的防护面具佩戴场景下，并与文献 [23] 提出的算法、AIZOO 算法 (https://github.com/aky15/AIZOO_torch) 和原始 YOLOv5 算法进行对比实验。防护面具主要包括医疗口罩、电焊面具、钢化玻璃面罩等，本文主要以医疗口罩 (以下简称口罩) 为研究对象进行识别。由于密集人群条件下往往人物众多，且容易出现人与人之间相互遮挡的现象，检测难度大，且单个人员的口罩占整幅图像的比例远远小于 20%，因此可以将其作为小目标对待。

3.1　数据集

本文数据集来源于 WIDER FACE、MAPA (Masked Faces) 这两个公开数据集和网络，从中手动筛选出密集人群场景下的佩戴口罩和未佩戴口罩的图片，最终得到训练集 4000 张，测试集 1320 张，共计 5320 张。

利用标记软件 LabelImg 对数据集进行 YOLO 格式的标注，共有两个标记类别，分别是 bad(未佩戴口罩) 和 good(佩戴口罩)。标注完成后，每一张图片都对应着一个与该图片名称相同的 txt 文件，txt 文件中的每一行表示一个标记实例，共 5 列，从左到右分别表示标签类别、标记框中心横坐标与图片宽度的比值、标记框中心纵坐标与图片高度的比值、标记框宽度与图片宽度的比值、标记框高度与图片高度的比值。

3.2 实验环境与模型训练

实验环境使用 Ubuntu20.04 操作系统，使用 GeForce GTX 1080Ti 显卡进行运算，显存大小为 11 GB，CPU 配置为 Intel(R) Xeon(R) CPU E5-2620 v3 @ 2.40GHz，CUDA 版本为 11.4.0，Pytorch 版本为 1.9.0，Python 语言环境为 3.7.4。

本实验总迭代次数为 140 次，迭代批量大小设置为 32，优化器选择 SGD。模型训练时学习率使用 Warmup[24] 训练预热，减缓模型在初始阶段对小批量数据的过拟合现象，避免模型振荡以便保证模型深层次的稳定性。在 Warmup 阶段，偏置层的学习率由 0.1 下降至 0.01，其他的参数学习率由 0 增加至 0.01，Warmup 结束之后，采用余弦退火学习算法[25] 对学习率进行更新。

3.3 评估指标与实验结果分析

本文评估指标采用平均精度 (average precision, AP)、平均精度均值 (mean AP, mAP) 以及每秒检测图片的帧数 (frames per second, FPS) 这 3 种在目标检测算法中较为常见的评价指标来评估本文算法的性能。平均精度与精确率 (precision) 和召回率 (recall) 有关，精确率是指预测数据集中预测正确的正样本个数除以被模型预测为正样本的个数；召回率是指预测数据集中预测正确的正样本个数除以实际为正样本的个数。上述衡量指标的计算公式分别为：

$$\mathrm{AP}=\int_0^1 P\mathrm{d}R \tag{11}$$

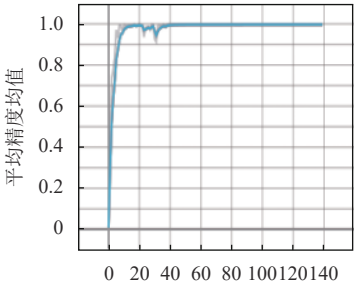
$$\mathrm{mAP}=\frac{\sum_{i=1}^{N}\mathrm{AP}_i}{N} \tag{12}$$

$$\mathrm{Precision}=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}=\frac{\mathrm{TP}}{\text{all detctions}} \tag{13}$$

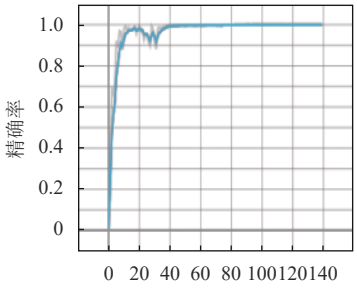
$$\mathrm{Recall}=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}=\frac{\mathrm{TP}}{\text{all ground truths}} \tag{14}$$

式中，AP 值是指 P-R 曲线面积，本文采用插值计算的方法来计算式 (11) 中的积分；mAP 的值是通过所有类别的 AP 求均值得到；N 表示检测的类别总数，本实验中 N=2，mAP 的值越大，表示算法检测效果越好，识别精度越高；TP、FP 和 FN 分别表示正确检测框、误检框和漏检框的数量。

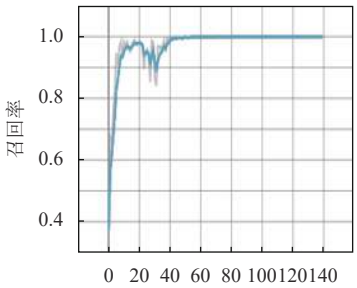
在训练 140 个迭代周期过程中，平均精度均值、精确率和召回率的变化曲线如图 10 所示。可以看出，模型在训练的过程中，在 Warmup 阶段结束后的几个迭代周期中，平均精度均值、精确率和召回率有些许下降，随后随着余弦退火算法对学习率的调整，模型逐渐达到收敛状态。



a. mAP 变化曲线



b. Precision 变化曲线



c. Recall 变化曲线

图 10 模型在数据集上的训练过程

为进一步验证本文算法的有效性，将本文算法与文献 [23] 方法、AIZOO 方法、原始 YOLOv5 算法在同一测试集上进行测试，各项性能指标比较结果如表 2 所示。

表 2 不同算法性能对比结果

算法	AP/%		mAP/%	Times/s	FPS
	bad	good			
文献[23]	83.53	84.17	83.85	0.028	35.3
AIZOO	87.36	86.88	87.12	0.021	47.6
YOLOv5	89.49	91.16	90.33	0.024	41.6
本文算法	93.21	96.54	94.88	0.033	30.3

从表 2 中可以看出，相较于文献 [23] 方法、AIZOO 方法和原始 YOLOv5 算法，本文算法在密集人群场景下对口罩这个小目标的检测表现效果更

好，mAP 值可以达到 94.88%，在原始 YOLOv5 的基础上，bad 和 good 类别的 AP 值分别提高了 3.72% 和 5.38%，mAP 值提高了 4.55%。本文算法在检测速率上不及其他算法，FPS 为 30.3，与原始 YOLOv5 相比，FPS 下降了 11.3，检测单张图片的时间增加了 9 ms，由于实时检测一般要求检测帧率大于 25 帧/s，故本文算法仍能满足实时性要求。本文算法与文献 [23] 方法、AIZOO 方法、原始 YOLOv5 算法进行对比的检测效果如图 12 所示。

从图 11 中可以看出，文献 [23] 方法在小目标异常角度、人脸区域有遮挡的条件下表现较差；AIZOO 方法在检测效果上整体表现稍好于文献 [23] 方法，单帧检测时间最少，FPS 最高；原始 YOLOv5 算法相较于前两种方法表现相对较好，但在一些小目标和存在遮挡条件下仍存在误判或者漏检的情况；与其他算法相比，本文算法在密集人群口罩佩戴检测效果中表现突出，检测精度有明显上升，误检、漏检现象明显减少，对小目标异常角度、人脸区域存在遮挡的鲁棒性明显提升。

a. 文献 [23] 方法

b. AIZOO

c. 原始 YOLOv5

d. 本文算法

图 11　检测效果对比图

4　结 束 语

本文在原有 YOLOv5 算法的基础上，分别从 Mosaic 数据增强、特征提取器、损失函数和目标框回归 4 个方面进行改进，有效地增强了 YOLOv5 网络模型对小目标物体的检测精度，改进后的算法检测速率相较于原始 YOLOv5 算法有所降低，但仍能满足实时性要求，可以直接应用在医学图像、遥感图像分析和红外图像中的小目标检测等实际场景中。

参 考 文 献

[1] NAJIBI M, SAMANGOUEI P, CHELLAPPA R, et al. Ssh: Single stage headless face detector[C]//Proceedings of the IEEE International Conference on Computer Vision. [S.l.]: IEEE, 2017: 4875-4884.

[2] LIU Y, MA Z, LIU X, et al. Privacy-preserving object detection for medical images with faster R-CNN[J]. IEEE Transactions on Information Forensics and Security, 2019, PP(99): 1.

[3] JAEGER P F, KOHL S A A, BICKELHAUPT S, et al. Retina U-Net: Embarrassingly simple exploitation of segmentation supervision for medical object detection[C]//Machine Learning for Health Workshop. [S.l.]: PMLR, 2020: 171-183.

[4] ZHANG L, LIN L, LIANG X, et al. Is faster R-CNN doing well for pedestrian detection?[C]//European Conference on Computer Vision. Cham: Springer, 2016: 443-457.

[5] RAGHUNANDAN A, RAGHAV P, ARADHYA H V R. Object detection algorithms for video surveillance applications[C]//2018 International Conference on Communication and Signal Processing (ICCSP). [S.l.]: IEEE, 2018: 0563-0568.

[6] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft coco: Common objects in context[C]//European Conference on Computer Vision. Cham: Springer, 2014: 740-755.

[7] ZHU Z, LIANG D, ZHANG S, et al. Traffic-sign detection and classification in the wild[C]//Proceedings of the IEEE Conference on Computer Vision And Pattern Recognition. [S.l.]: IEEE, 2016: 2110-2118.

[8] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

[9] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2014: 580-587.

[10] HE K, ZHANG X, REN S, 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.

[11] GIRSHICK R. Fast R-CNN[C]//Proceedings of the IEEE International Conference on Computer Vision. [S.l.]: IEEE, 2015: 1440-1448.

[12] REN S, HE K, 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, 2016, 39(6): 1137-1149.

[13] REDMON J, DIVVALA S, GIRSHICk R, et al. You only look once: Unified, real-time object detection[C]//

Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2016: 779-788.

[14] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]//European Conference on Computer Vision. Cham: Springer, 2016: 21-37.

[15] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2017: 7263-7271.

[16] REDMON J, FARHADI A. Yolov3: An incremental improvement[EB/OL]. [2021-03-25]. https://arxiv.org/pdf/1804.02767.pdf.

[17] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. Yolov4: Optimal speed and accuracy of object detection[EB/OL]. [2021-04-15]. https://arxiv.org/abs/2004.10934.

[18] ZHENG Zhaohui, WANG Ping , LIU Wei, et al. Distance-IoU loss: Faster and better learning for bounding box regression[C]//Proceedings of the AAAI Conference on Artificial Intelligence. [S.l.]: AAAI, 2019, DOI: 10.1609/aaai.v34i07.6999.

[19] LIN T Y , DOLLAR P , GIRSHICK R , et al. Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE Computer Society, 2017, DOI: 10.1109/CVPR.2017.106.

[20] LIU S , QI L , QIN H , et al. Path aggregation network for instance segmentation[EB/OL]. [2020-11-12]. https://arxiv.org/pdf/1803.01534.pdf.

[21] REZATOFIGHI H, TSOI N, GWAK J Y, et al. Generalized intersection over union: A metric and a loss for bounding box regression[EB/OL]. [2020-11-15]. https://arxiv.org/pdf/1902.09630.pdf.

[22] NEUBECK A, GOOL L V. Efficient non-maximum suppression[C]//International Conference on Pattern Recognition. Hongkong, China: IEEE Computer Society, 2006, DOI: 10.1109/ICPR.2006.479.

[23] 肖俊杰. 基于 YOLOv3 和 YCrCb 的人脸口罩检测与规范佩戴识别[J]. 软件, 2020, 41(7): 164-169.
XIAO J J. Masked face detection and standard wearing mask recognition[J]. Computer Engineering & Software, 2020, 41(7): 164-169.

[24] XIONG R, YANG Y, HE D, et al. On layer normalization in the transformer architecture[EB/OL]. [2021-10-12]. https://arxiv.org/abs/2002.04745v1.

[25] LOSHCHILOV I, HUTTER F. SGDR: Stochastic gradient descent with warm restarts[EB/OL]. [2021-05-25]. https://arxiv.org/pdf/1608.03983.pdf.

编 辑 叶 芳