

基于改进 YOLOv4 的灾后人员检测算法^①

张立国^② 李佳庆 赵嘉士 耿星硕 章玉鹏

(燕山大学电气工程学院 秦皇岛 066004)

(河北省测试计量技术及仪器重点实验室 秦皇岛 066004)

摘要 针对地震等灾后环境复杂、救援机器人对救援目标识别实时性和准确度要求较高的问题,提出一种基于改进 YOLOv4 的目标检测模型。该算法将 YOLOv4 网络中的主干特征提取网络替换成 MobileNetv1 模型以增强特征复用,同时缩减网络参数量,提高运行速度;通过 K-means++ 算法进行锚点维度聚类以适应灾后人员检测,提升算法精度。并且针对灾后人员检测数据集缺乏问题,贡献相应的数据集。实验结果表明,改进的网络与 YOLOv4 相比在保证模型精度的前提下帧率提升了约 92%,权重文件大小变为原来的 20.73%,满足了灾后救援机器人目标检测实时性和准确性的需求,对于灾后人员检测场景有一定的借鉴意义。

关键词 灾后救援;目标检测;改进 YOLOv4; Mobilenetv1; K-means++

0 引言

近些年来,全球自然灾害与意外事故频频发生,人民的生命安全受到重大威胁。普遍地,部分被困人员因为障碍的遮挡而得不到及时救援,最终失去生命^[1-3],因此,能否在最佳救援时间内找到被困人员尤为重要。

随着传统目标检测算法性能达到瓶颈,目标检测技术达到了一个稳定的水平。随着 AlexNet 网络使用卷积神经网络(convolutional neural network, CNN)大幅提升目标分类的准确度之后,卷积神经网络再次回到人们的视线中,基于卷积神经网络的目标检测技术开始飞速发展。如今,目标检测算法可以分为两类,一类是以 YOLO(you only look once)系列和 SSD(single shot multibox detector)算法为代表的一阶段检测算法,另一类是以 R-CNN(region-CNN)系列为代表的二阶段检测算法^[4-7]。最初的 YOLO

算法检测效果并不好,但随后更新的 YOLOv2^[8]、YOLOv3^[9]、YOLOv4^[10]等一系列算法,在速度和精度上都得到了巨大的提升^[11-12],使得其不同背景下行人检测中得到大规模应用。如文献[13]使用 Dense-YOLO 网络模型进行井下行人检测。文献[14]使用改进 YOLO 轻量化网络的算法检测行人。文献[15]使用改进 YOLO 网络对遮挡行人进行检测仿真。但是在地震等灾后救援场景中,算法需要部署在有限的嵌入式设备计算资源上,这对 YOLO 网络的实时性和准确性提出了巨大挑战。

本文选择目前在速度和精度上有明显优势的 YOLOv4 网络为基础,将 MobileNetv1 轻量级网络引入 YOLOv4 中,同时为了解决减少参数量而导致的准确率降低问题,使用 K-means++ 算法优化预选框以提升模型识别精度,最终得到改进的 YOLOv4 模型。新的网络模型可以大幅减少网络的参数量,提升检测速度,使其满足移动端设备存储容量小、算力低等条件。

^① 工信部生产现场信息物理系统平台重点项目(CEIEC-2020-ZM02-0131)和河北省中央引导地方科技发展专项(199477141G)资助项目。

^② 1978 年生,博士,副教授;研究方向:视觉跟踪与可重构技术,机器人应用技术,虚拟现实技术;联系人,E-mail: zlgtime@163.com。
(收稿日期:2022-06-05)

1 YOLOv4 原理

YOLOv4^[10] 是基于回归的一阶段目标检测算法。其网络结构由主干特征提取网络、特征融合网络和进行分类、回归的检测头(Head)组成。主干特征提取网络利用残差块对输入图片进行 5 次卷积,获得数据的特征信息。YOLOv4 在 DarkNet53 的基础上结合跨阶段局域网络(cross stage partial network, CSPNet)^[16] 的优点,提出了 CSPDarkNet53 结构,利用大残差块的不完全截流特性,降低了主干网络的计算瓶颈和内存成本。空间金字塔池化(spatial pyramid pooling, SPP)结构主要功能是增加网络的感受野,提高特征金字塔结构的特征提取能力。同时特征融合网络采用 SPP 和路径聚合网络(path aggregation network, PANet)的网络结构,在原有上采样的特征浅层融合基础上^[17-19],增加了下采样的方式进行深层信息融合,这样可以使不同特征层有效融合,避免直接对图片进行剪裁缩放造成数据信息丢失。

2 嵌入 MobileNetv1 和 K-means ++ 模块的改进 YOLOv4 算法

2.1 YOLOv4-MobileNetv1 整体框架

YOLOv4-MobileNetv1 具体实现是将 YOLOv4 的主干特征提取网络 CSPDarkNet53^[17] 替换成网络参数轻量化的 MobileNetv1 神经网络,降低 YOLO 网络整体的计算量和参数量。改进后的 YOLOv4 网络的整体结构如图 1 所示。其中将原 YOLOv4 网络的输入图片尺寸大小 608×608 改成 416×416 , 在 MobileNetv1 主干特征网络采用 Mish 激活函数,后面的网络仍采用 ReLU 激活函数。Mish 激活函数保证了正数可以达到任意高度,避免了由于封顶而导致的梯度饱和,对负值的轻微变化允许更好的梯度流,而不是像 ReLU 激活函数一样具有零硬边界。采用 Mish 激活函数可以允许更好的信息深入神经网络,从而得到更好的准确性和泛化能力。Mish 激活函数的计算公式如式(1)、(2)所示。

$$f(x) = x \times \tanh(\ln(1 + e^x)) \quad (1)$$

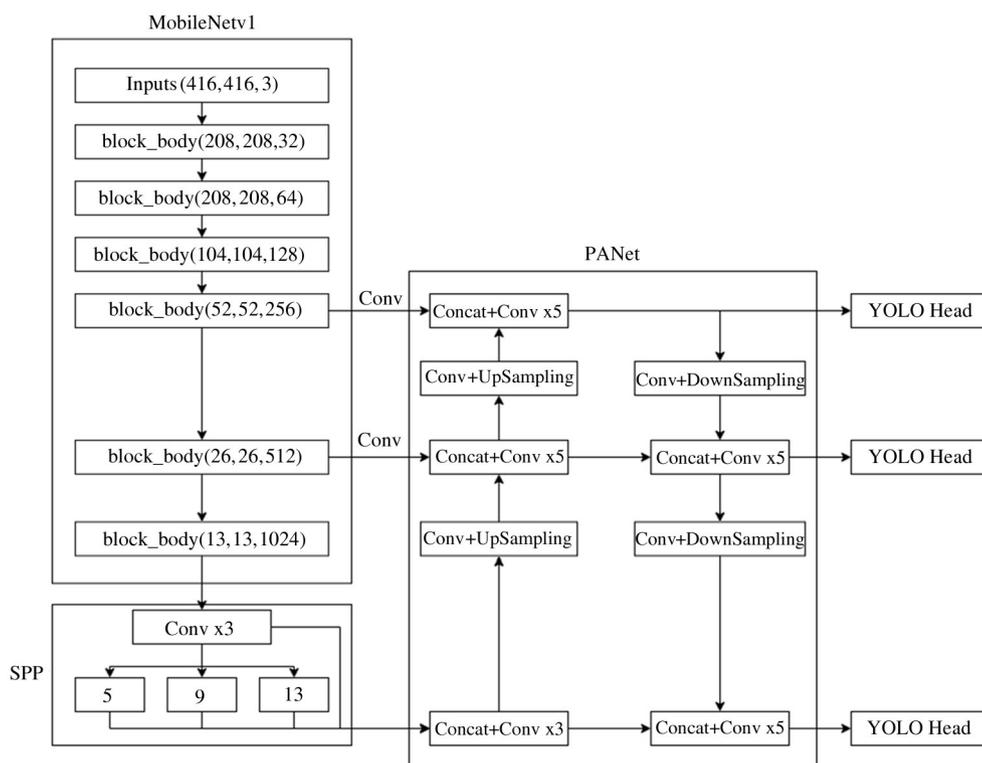


图 1 YOLOv4-MobileNetv1 整体结构图

$$\tanh(x) = \frac{\sinh(x)}{\cosh(x)} = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (2)$$

式中 x 为输入数据。

SPP 结构^[18]分别使用 4 个不同大小的卷积核对特征提取网络提取到的特征进行最大池化处理,卷积核大小分别为 1×1 、 5×5 、 9×9 、 13×13 ,将 4 个池化处理后的不同尺度特征图进行拼接组成新的特征层。在 PANet 结构^[19]对 MobileNetv1 主干特征提取网络特征提取的结果和 SPP 网络特征提取结果做进一步的信息提取。进行特征提取的特征图分别为 52×52 、 26×26 、 13×13 。PANet 网络加强了对低层特征信息的利用率,通过不同层的自适应池化,将特征网格与全部特征层直接融合在一起,在下采样的过程中将融合的结果传入检测头部分进行最终的回归和分类^[20]。最后,YOLOv4-Head 采用 PANet 结构的 3 个特征层进行预测,分别进行卷积核大小为 3×3 与 1×1 的卷积操作,判别特征层的 3 个先验框,判断网格区域内目标及种类,最终获得预测框^[21]。

2.2 MobileNetv1 网络结构

通常搭建更深更复杂的神经网络可以提升准确率,但其在推理速度和模型大小上却不一定是高效的。MobileNetv1 是基于深度可分离卷积构建的轻量级网络模型^[22],并定义 2 个超量化参数 α 和 β 来降低模型的宽度和输入图像的分辨率。 α 为宽度数乘用来控制卷积层卷积核数量从而显性地将卷积的输出通道数下降 α 倍, β 为分辨率数乘用来隐性地控制输入图像大小,通过这 2 个参数来控制模型的大小。

为了使网络更好地进行融合,本文对 MobileNetv1 模型进行了修改。首先将输入图片尺寸从

224×224 替换成 416×416 ,这样可以通过优化后续的 K-means ++ 算法提高检测精度;其次裁剪部分了卷积层和全连接层,这样可以使得网络更加轻量化,提高运行速度。裁剪后的 MobileNetv1 网络结构如表 1 所示,其中 t 是拓展因子, c 是输出特征矩阵深度 channel, n 是 block_body 的重复次数, s 是步距。

表 1 裁剪后 MobileNetv1 网络结构

Input	Operator	t	c	n	s
$416^2 \times 3$	Conv2d	-	32	1	2
$208^2 \times 32$	block_body	1	64	1	1
$208^2 \times 64$	block_body	6	128	2	2
$104^2 \times 128$	block_body	6	256	2	2
$52^2 \times 256$	block_body	6	512	5	2
$26^2 \times 512$	block_body	6	1024	1	2
$13^2 \times 1024$	block_body	6	-	1	1

本文中 MobileNetv1 的平均池化采用的步长为 1,卷积核尺寸为 7×7 。由于采取的步长为 1,所以输出的特征图不变,仍然为 $13 \times 13 \times 1024$ 。

MobileNetv1 网络采用了深度可分离卷积作为基础的特征提取单元^[23]。深度可分离卷积^[24]将卷积过程拆分为逐通道卷积与逐点 1×1 卷积 2 步,这种分解过程能大量减少冗余计算量。相比传统卷积,逐通道卷积的卷积核通道为 1,输入特征矩阵的维度同卷积核个数以及输出特征矩阵的维度相同。逐点卷积同普通卷积基本一致,唯一区别在于逐点卷积的卷积核大小为 1×1 。逐通道卷积和逐点卷积共同组成了深度可分离卷积,理论上使用深度可分离卷积的参数计算量为传统卷积的 $1/9$ 。深度可分离模块的具体结构如图 2 所示。

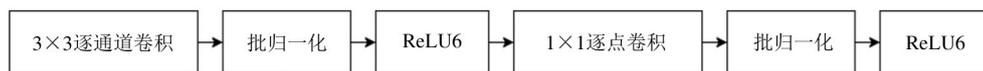


图 2 深度可分离卷积结构图

MobileNetv1 网络使用 ReLU6 激活函数是为了满足移动端设备的需求。移动端通常使用 Float16 或者 Int8 等较低精度的模型,这种模型的精度范围无法覆盖到正无穷,若使用 ReLU 激活函数会带来

精度损失,而 ReLU6 激活函数将上限设为 6,这样可以很好地满足移动设备的需求。

2.3 锚框优化

为了增强模型对跨尺度目标的预测能力,本文

采用 K-means ++ 算法对图像特征层上的锚框进行重新聚类,得到 9 个新的锚框。相较于 K-means 算法需要人为地给定初始聚类中心、初始聚类中心对最终聚类结果影响较大的缺点,本文使用的 K-means ++ 算法只需要给出第 1 个初始聚类中心,并且提高了后续聚类中心选取的效率。K-means ++ 算法首先从数据集中任意选取一个样本点作为初始聚类中心,然后计算所有样本点到已有聚类中心的距离并使用加权概率分布选择一个新的样本点作为新的聚类中心,不断重复选取,直到选择好 9 个锚框为止。K-means ++ 算法解决了传统 K-means 算法对初始聚类中心敏感的问题,主要改进在于第 1 步选取初始聚类中心时不再是一次性随机选取完,而是遵循初始聚类中心之间的距离尽可能远的原则去逐次选取聚类中心。K-means ++ 算法计算样本点到聚类中心的距离为欧氏距离,其计算公式如式(3)所示。

$$d(x_i, x_j) = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \cdots + (x_{ip} - x_{jp})^2} \quad (3)$$

式中 x_i, x_j 代表 2 个样本点,其中每个样本点都有 p 个属性。

K-means ++ 算法最终聚类结果的评估标准为误差平方和,其公式如式(4)、(5)所示。

$$SSE = \sum_{i=1}^k \sum_{x \in E_i} dist(e_i, x)^2 \quad (4)$$

$$e_i = \frac{1}{n_i} \sum_{x \in E_i} x \quad (5)$$

式中, SSE 表示误差平方和,在相同迭代次数的条件下, SSE 的值越小则说明算法损失越小; e_i 为 k 个聚类中心的第 i 个聚类中心点; $dist$ 表示每个点到它所属簇的中心点的距离。

最终以重叠度(intersection over union, IOU)为标准^[25],在一定的阈值下,IOU 的值越大,生成的锚点到聚类中心的距离就越小^[26],预测结果越精确。最终得出适合灾后人员数据集的锚框,其宽和高分别是(142, 110)、(192, 243)、(459, 401)、(36, 75)、(76, 55)、(72, 146)、(12, 16)、(19, 36)、(40, 28)。

3 实验结果与分析

3.1 数据集制作及数据预处理

由于目前的灾后人员图像数据集较少,数据集图像质量也不高,因此本文制作了一个灾后人员检测数据集。数据集包含 2000 张不同场景下的灾后人员的图片和高度遮挡的图片。在 VOC2007 + VOC2012 数据集中,训练集有 16 551 张图片,测试集有 4952 张图片。其图片具有丰富的前景和背景,同时具有大量高度遮挡的人群场景。实验时使用 VOC2007 + VOC2012 公开数据集和自制灾后人员数据集进行训练和验证,对 VOC2007 + VOC2012 公开数据集进行数据预处理,使用 Mosaic 和 Label smooth 2 种数据增强方法。Mosaic 数据增强可以有效提高数据集的丰富度和检测的鲁棒性,Label smooth 数据增强可以提升精度。

3.2 评价指标

本文使用精准率 P (precision)、召回率 R (recall)和平均精度均值 mAP (mean average precision)3 个指标检测网络模型的识别效果,同时使用画面每秒传输帧数 fps(frames per second)作为模型检测速度评估的指标^[27]。精准率和召回率的计算公式如式(6)、(7)所示。

$$P = \frac{TP}{TP + FP} \quad (6)$$

$$R = \frac{TP}{TP + FN} \quad (7)$$

式(6)~(7)中, TP (true positives)表示将正样本的目标预测为正值, FN (true negatives)表示将负样本的目标预测为负值, FP (false positives)表示将正样本的目标预测为正值。

平均准确率(AP)和平均准确率均值(mAP)的计算公式如式(8)、(9)所示,式中, N 为类别数。

$$AP = \int_0^1 P(R) dR \quad (8)$$

$$mAP = \frac{\sum_{i=1}^N AP_i}{N} \quad (9)$$

3.3 不同算法检测性能对比

为验证模型的可靠性,在同一设备的情况下,使

用不同的改进算法进行对比实验,不同算法结果如图3~6所示。

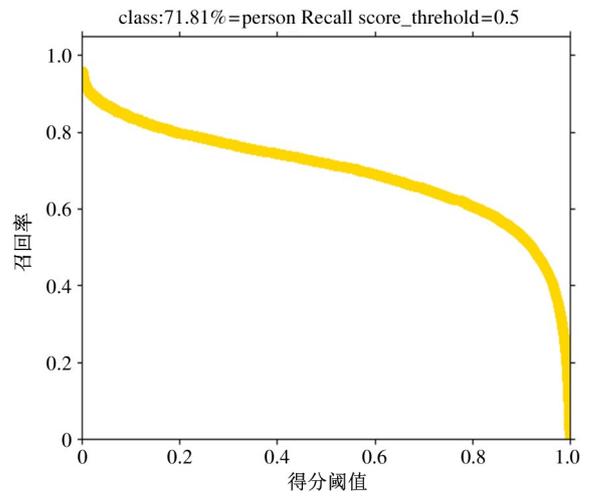
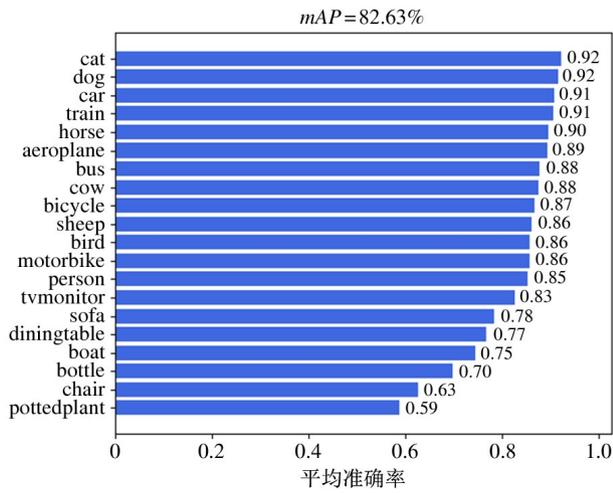


图3 MobileNetv1 P-R 曲线

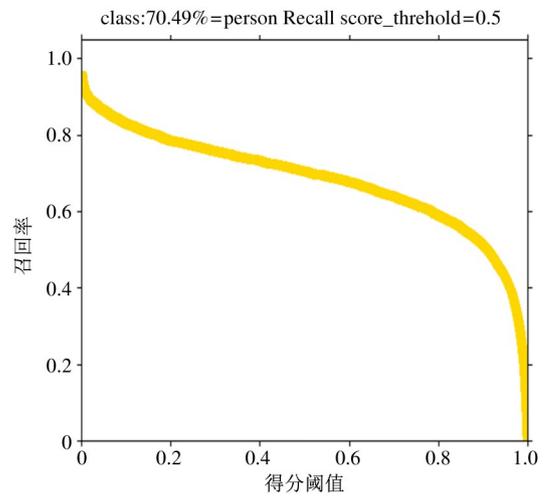
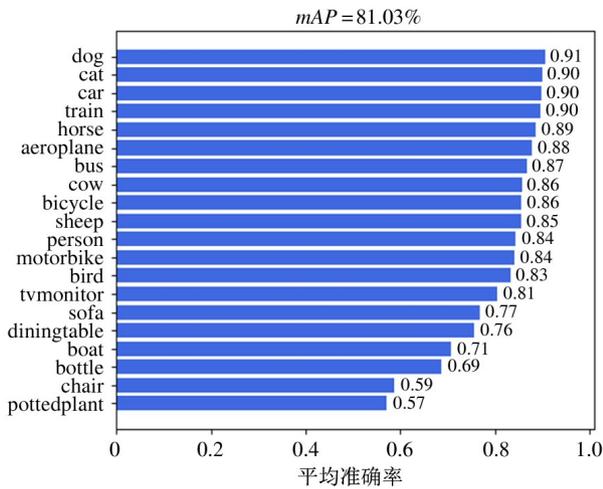


图4 MobileNetv2 P-R 曲线

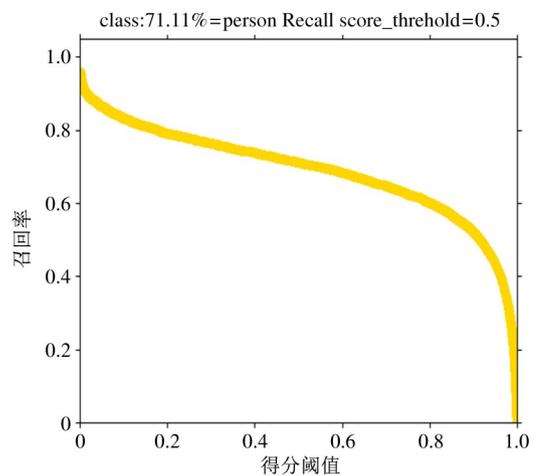
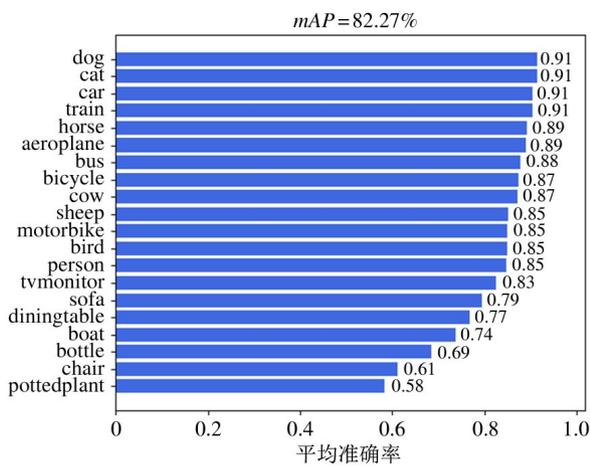


图5 MobileNetv3 P-R 曲线

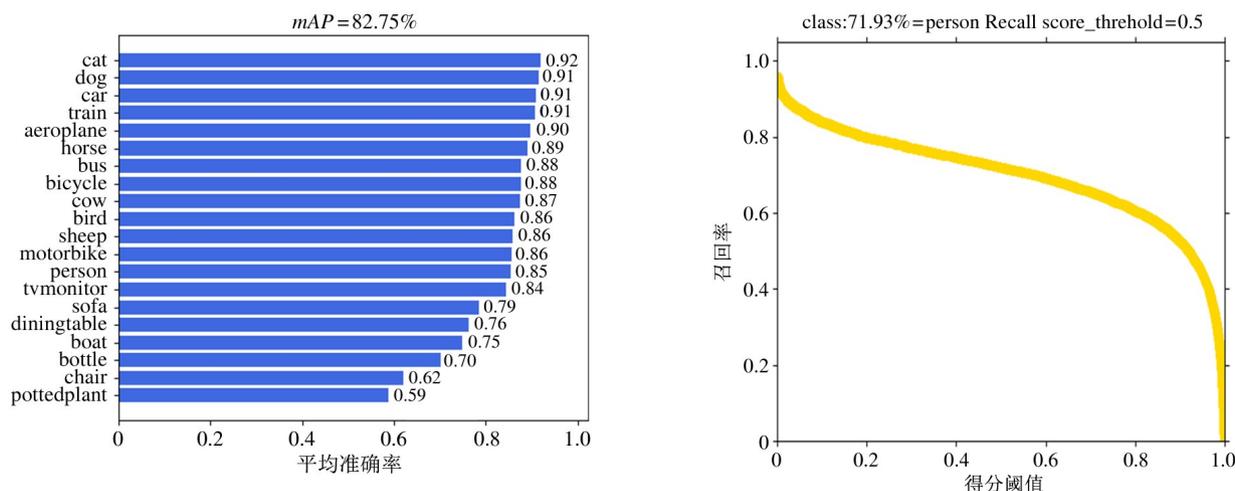


图6 DenseNet121 P-R 曲线

为了检测改进后算法的可行性,本文将所有尝试改进的网络模型和原 YOLOv4 网络模型从检测精度和运行速度两方面进行对比,同时还和国内外一些目标检测的先进模型进行对比,采用平均准确率均值(mAP)检验 8 个网络模型的精度,采用画面每秒传输帧数(FPS)比较 8 个模型的检测速度,对比结果见表 2。

表 2 不同算法准度和精度对比

网络模型	fps	$mAP/\%$
YOLOv4-MobileNetv1	72.61	82.63
YOLOv4-MobileNetv2	58.63	81.03
YOLOv4-MobileNetv3	52.11	82.27
YOLOv4-DenseNet121	31.66	82.75
YOLOv4	37.82	89.88
CenterNet-Hourglass104	1.43	68.74
YOLOv4-tiny	126.52	71.49
YOLOv3	77.45	63.40

从表 2 可知, YOLOv4 的平均检测精确度达到 89.99%, 检测速度达到 37.8 fps。而对于其余先进算法, 精度虽然都有所下降, 但 MobileNet 系列在检测速度上得到了较大提升。其中 YOLOv4-MobileNetv1 网络模型效果最好, 检测速度达到了 72.61 fps, 较原网络提升了 91.98%。由于引入了深度可分离卷积操作所以其精度略低于 YOLOv4 模型, 但仍高于其余模型。

3.4 自制数据集实验结果

自然灾害发生后, 会出现建筑物损毁、人员掩埋、城市内涝等现象, 导致被困人员特征不明显、难以被识别等问题。而基于深度学习的目标检测算法需要大量带标签的数据进行训练, 否则会出现过拟合、泛化能力较差等问题。目前的公开数据集还没有针对灾后人员检测的, 因此为了精细化实时检测, 本文搜集了大量灾后图片并制作成数据集, 内容包括城市废墟、洪水内涝、野外山林等。

从测试集中挑选部分图片进行结果展示, 如图 7 所示, 本文提出的改进 YOLOv4-MobileNetv1 网络模型在有不同程度遮挡的复杂环境下均检测到了目标, 取得了较好表现。

从测试集结果图 7(a) 和 (b) 中可以看到, 当地震等自然灾害发生在城市中, 会造成人员被废墟掩埋的情况。这时由于遮挡过多或者目标距离太近可能出现漏检现象, 但大部分救援对象是可以被准确检测出来的。从测试集结果图 7(c) 和 (d) 可以看出, 当救援背景发生在野外时, 会出现被救人员在图片中占比过小、颜色同背景较为接近的情况, 本文提出的改进网络可以对不同尺度、颜色相近的目标进行识别。从测试结果来看, 本文提出的改进 YOLOv4 算法对尺度变换、背景颜色干扰、部分遮挡等问题均有良好的检测效果, 但对于遮挡情况过于严重、人员距离过近的情况其检测精度仍待提高。



图 7 部分测试集结果

4 结论

本文提出了一种改进的 YOLOv4 的轻量化灾后人员检测算法,将 MobileNetV1 轻量级网络替换原 YOLOv4 主干特征提取网络,进而得到 MobileNetV1-YOLOv4 模型;采用 K-means++ 聚类对灾后人员目标样本进行锚定框大小的优化,提升模型最终识别度。实验表明该方法平均每秒检测约 92 张图像,在 VOC2007 公开数据集上的平均准确率均值达到了 82.17%,单独检测受灾人员的成功率达到 71.81%,同时对不同环境有良好的抵抗能力和鲁棒性。但是本文方法受限于人员太密集、障碍物遮挡等问题,导致实际检测中仍存在少量的漏检与错检,下一阶段工作将针对此问题进行更深入的研究。

参考文献

- [1] 孙晓普,杨国强. 危险环境下救援机器人技术发展现状与趋势[J]. 科技与创新,2019,11:1-4.
- [2] 金冶纯. 人工智能机器人在防灾中的应用对策研究[J]. 研发应用,2020(22):21-52.
- [3] 贾硕,张文昌,吴航,等. 救援机器人研究现状及其发

展趋势[J]. 医疗卫生装备,2019,40(8):90-95,100.

- [4] 刘文汇,巢渊,唐寒冰,等. 移动机器人视觉目标检测与跟踪方法研究综述[J]. 电光与控制,2022,29(4):59-67,88.
- [5] 叶伯生,谢鹏,张文彬. 基于随机遗传算法优化 BP 神经网络的工业机器人整机性能评估模型[J]. 中南大学学报(自然科学版),2021,52(9):3204-3211.
- [6] ZHANG J, MENG Y, CHEN Z. A small target detection method based on deep learning with considerate feature and effectively expanded sample size[J]. IEEE Access, 2021,9:96559-96572.
- [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.
- [8] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE, 2017:6517-6525.
- [9] REDMON J, FARHADI A. YOLOv3: an incremental improvement[EB/OL]. (2018-04-08) [2022-06-09]. <http://arxiv.org/pdf/1804.02767.pdf>.
- [10] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection[EB/OL]. (2020-04-23) [2022-06-09]. <http://arxiv.org/pdf/2004.10934.pdf>.
- [11] 赵永强,饶元,董世鹏,等. 深度学习目标检测方法综

- 述[J]. 中国图象图形学报, 2020, 25(4): 629-654.
- [12] 张慧, 王坤峰, 王飞跃. 深度学习在目标视觉检测中的应用进展与展望[J]. 自动化学报, 2017, 43(8): 1289-1305.
- [13] 张明臻. 基于 Dense-YOLO 网络的井下行人检测模型[J]. 工矿自动化, 2022, 48(3): 86-90.
- [14] 常青, 韩文, 王清华, 等. 改进 YOLO 轻量化网络的行人检测算法[J]. 光学技术, 2022, 48(1): 80-85.
- [15] 向南, 王璐, 贾崇柳, 等. 改进 YOLO 的遮挡行人检测仿真[J]. 系统仿真学报, 2023, 35(2): 286-299.
- [16] WANG C Y, LIAO H Y M, WU Y H, et al. CSPNet: a new backbone that can enhance learning capability of CNN[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Seattle: IEEE, 2020: 390-391.
- [17] 张明路, 郭策. 改进的轻量化 YOLO4 用于电子元器件检测[J]. 电子测量与仪器学报, 2021, 35(10): 17-23.
- [18] 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.
- [19] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 8759-8768.
- [20] 万浪, 凌毓涛. 基于 MobileNetV3-YOLOv4 的车型识别[J]. 软件导刊, 2021, 20(12): 173-178.
- [21] 李若熙, 吕潇, 等. 改进 YOLOv4 算法井下人员检测的研究[J]. 矿业研究与开发, 2021, 31(11): 179-185.
- [22] HOWARD A G, ZHU M, CHEN B, et al. Mobilenets: efficient convolutional neural networks for mobile vision applications[EB/OL]. (2017-04-17) [2022-06-09]. <http://arxiv.org/pdf/1704.04861.pdf>.
- [23] SANDLER M, HOWARD A, ZHU M, et al. Mobilenetv2: inverted residuals and linear bottlenecks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 4510-4520.
- [24] HOWARD A, SANDLER M, CHU G, et al. Searching for mobilenetv3[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019: 1314-1324.
- [25] 杨俊闯, 赵超. K-Means 聚类算法研究综述[J]. 计算机工程与应用, 2019, 55(23): 7-14.
- [26] 胡亮, 何小海, 卿艮波, 等. 基于 YOLOv4 目标检测算法的轻量化网络设计[J]. 2022, 12(1): 164-167.
- [27] CAI Z, VASCONCELOS N. Cascade R-CNN: delving into high quality object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 6154-6162.

Post-disaster personnel detection algorithm based on improved YOLOv4

ZHANG Ligu, LI Jiaqing, ZHAO Jiashi, GENG Xingshuo, ZHANG Yupeng
(School of Electrical Engineering, Yanshan University, Qinhuangdao 066004)

(Hebei Key Laboratory of Testing and Metrology Technology and Instruments, Qinhuangdao 066004)

Abstract

A target detection model based on improved YOLOv4 is proposed for the complex post-disaster environments such as earthquakes and rescue robots with high requirements for real-time and accurate recognition of rescue targets. This algorithm replaces the backbone feature extraction network in the YOLOv4 network with the MobileNetV1 model to enhance feature reuse, while reducing the number of network parameters and improving operational speed. Anchor dimension clustering by K-means ++ algorithm to adapt to post-disaster personnel detection and improve algorithm accuracy. And it contributes to addressing the problem of missing datasets for post-disaster personnel detection. Experimental results show that the improved network improves the frame rate by about 92% compared with YOLOv4 in guaranteed model accuracy, and the weight file size becomes 20.73% of the original. The demand for real-time and accurate target detection of post-disaster rescue robots is satisfied, and this study has certain implications for post-disaster personnel detection scenarios.

Key words: post-disaster relief, object detection, improved YOLOv4, MobileNetV1, K-means ++