DOI:10.19651/j.cnki.emt.2211406

残差特征融合的小目标动态实时检测算法

冉险生　陈俊豪　苏山杰　张之云
(重庆交通大学机电与车辆工程学院 重庆 400047)

摘　要： 针对图片中小目标携带信息少、尺度变化大等检测难点，本文以 YOLOv5s 为框架，提出一种特征融合的小目标动态实时检测模型(HCD-YOLOv5s)。针对模型下采样易造成小目标信息丢失、深层网络位置信息表达不足等问题，从浅层中新增检测小目标的检测头；本文针对特征融合造成的特征混淆等问题，设计一种特征融合方式 CCAT，减少检测层位置信息与语义信息的丢失；针对检测任务与数据分布不同适应的激活函数不一致，设计 DConv 模块，分离回归任务与检测任务，实现模型的动态检测。本文在 VisDrone 数据集上对模型进行消融实验，3 个模块相互促进。选取不同输入尺寸的图片对模型进行速度与精度测试。在 YOLOv5s 的基础上 HCD-YOLOv5s 的 mAP50 提高了 10.2%，检测精度与参数量明显优于 YOLOv5m，FPS 达到 90。最后在 DOTA-v1.0 上进行实验验证，mAP50、mAP 分别提升了 1.8%与 2.0%，证明本文提出的 HCD-YOLOv5s 在小目标检测上有更佳的性能。

关键词： 小目标；特征融合；动态函数；实时检测
中图分类号： TP391　　**文献标识码：** A　　**国家标准学科分类代码：** 52060

Small target dynamic real-time detection algorithm based on residual feature fusion

Ran Xiansheng　Chen Junhao　Su Shanjie　Zhang Zhiyun
(School of Mechatronics and Vehicle Engineering, Chongqing Jiaotong University, Chongqing 400047, China)

Abstract: Aiming at the detection difficulties of small targets in pictures, such as less information and large scale changes, this paper proposes a feature fusion small target dynamic real-time detection model (HCD-YOLOv5s) based on YOLOv5s. In view of the problems that sampling under the model is easy to cause the loss of small target information and insufficient expression of deep network location information, a detection head for detecting small targets is added from the shallow layer; Aiming at the problem of feature confusion caused by feature fusion, this paper designs a feature fusion method CCAT to reduce the loss of location information and semantic information in the detection layer; In view of the inconsistency between the detection task and the activation function adapted to the different data distribution, the DConv module is designed to separate the regression task and the detection task, so as to realize the dynamic detection of the model. In this paper, the Ablation Experiment of the model is carried out on the visdrone data set, and the three modules promote each other. Select pictures with different input sizes to test the speed and accuracy of the model. On the basis of YOLOv5s, the mAP50 of HCD-YOLOv5s is increased by 10.2%, the detection accuracy and parameter quantity are significantly better than YOLOv5m, and the FPS reaches 90. Finally, the experimental verification is carried out on DOTA-v1.0, and the mAP50 and mAP are increased by 1.8% and 2.0% respectively, which proves that the HCD-YOLOv5s proposed in this paper has better performance in small target detection.

Keywords: small goals; feature fusion; dynamic function; real-time detection

0　引　　言

目标检测近年来成为智能驾驶的火热课题，特斯拉使用传统相机替换检测激光雷达，极大地减少了汽车的生产成本。目标检测算法在 PASCAL VOC、COCO 等数据集中，对中、大型目标的检测阈值为 50 的平均精度(mean average precision，mAP)可以达到 90%以上，但小目标的检测精度只有中、大型目标的 1/3。YOLOv5 算法在小目标

收稿日期：2022-09-15

数据集 VisDrone 上只能达到 34.5%的 mAP，所以想提高目标检测的精度，需要先解决小目标的检测问题。小目标检测困难主要是感受野、特征融合、数据集的影响。过大的感受野使模型关注检测背景，过小的感受野使模型只能关注局部信息；特征融合容易造成信息的混淆效应；数据聚集、尺度变化等因素容易造成样本不均衡的问题。

Bodla 等[1]通过 soft-NMS 函数，解决传统非极大值抑制 NMS(non-maximum suppression，NMS)中小目标检测框易被剔除的现象。Zhang 等[2]从小目标感受野的角度出发，主干网络中融合不同大小的空洞卷积增大感受野，从浅层中获得细粒度信息并融合深层语义信息。奉志强等[3]改进 SCAM 提高模型对小目标的关注程度；为不同的特征层添加自适应权重提高模型的融合效率。Gong 等[4]设计融合因子控制深层向浅层的信息传送，使用特征金字塔(feature pyramid network，FPN)适应微小目标的检测。Ge 等[5]利用 anchor free 与 SimOTA 方法减少检测框的数量与计算量，同时减轻样本不均衡问题；利用 double head 将分类任务与回归任务分离。Wang 等[6]设计采样算子 carefe，使上采样可以根据特征图的语义信息轻量化的进行上采样，弥补小目标下采样时的信息损失。Zhu 等[7]采用 anchor free 通过中心权重与置信度权重来对每个实例进行分配，提高小目标的检测能力。

上诉文章大多数从感受野、特征融合、注意力机制、anchor free 上出发，但未考虑以下两点：第一在检测部分未考虑各通道之间数据输入的差异，使用同样的激活函数，属于静态检测。Chen 等[8]指出了不同的检测任务与数据输入适合不同的激活函数；第二特征层进行特征融合时，简单进行特征层之间的拼接，没有考虑各层之间的信息差异，容易产生融合混淆。

根据以上问题本文提出了 HCD-YOLOv5 网络结构。首先针对小目标下采样易造成信息损失的问题，从浅层引出一个检测头，检测像素在 4×4 及其以上的小目标；然后针对各通道检测任务与输入数据分布的不同，在检测头设计一种 DConv 结构，动态适应检测任务与输入数据的分布；最后根据各层携带信息的不同，在融合过程中设计 CCAT 结构，提高融合过程中小目标位置信息的传递，与大目标的语义信息的传递。

1　检测模型

YOLOv5 网络模型的检测精度高，推理速度快，能够有效的识别复杂环境下的目标、小目标。该网络一共有 5 种基础架构，分别是 n、s、m、l、x；对应模型大小依次是 1.9、7.2、21.2、46.5、86.7 M；网络的参数量为 4.5、16.5、49.0、109.1、205.7 B。随着参数量增大，模型的精度会得到提升，但需要消耗更多的网络计算资源，同时检测速率也会降低。为了满足检测速度与检测精度，本文的以 YOLOv5s 模型为基础，改进后的模型如图 1 所示，分为 backbone、neck、heads 3 个模块，backbone 进行特征提取，neck 部分进行不同特征层之间的融合，heads 部分对 neck 的输出进行检测。

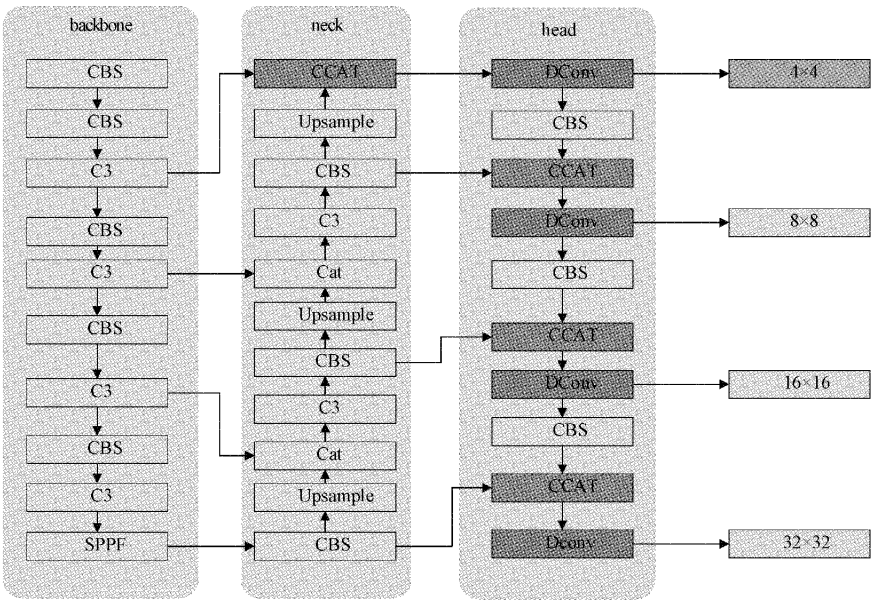


图 1　HCD-YOLOv5 结构

本文主要进行以下 3 点改进：首先针对提高小目标的检测能力，在 YOLOv5s 的基础上从浅层引出小目标检测头。其次，针对特征融合产生的特征混淆，提出一种新的特征融合形式 CCAT 模块，减小深层网络对浅层网路位置信息的影响，提高检测小目标的位置信息的能力；同时增加深层网络的语义信息，在不影响特征融合网络的参数量的情况下，提高模型检测小目标位置信息与分类信息的能力。最后，针对检测头在不同特征层的检测任务与数据分

布不同,提出 DConv 模块替代检测头的 C3 模块,使检测头可以根据检测任务与数据分布动态的改变激活函数的参数,优化分类任务与回归任务之间的冲突。

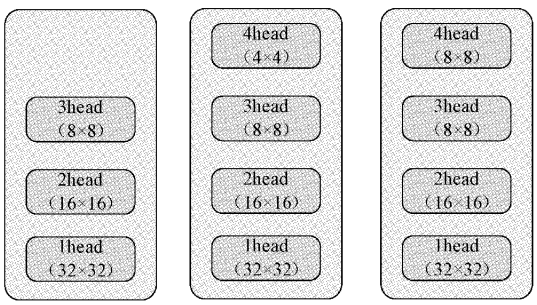
1.1 4heads

YOLOv5 的 backbone 中一共经过 5 次下采样,输出有 3 个检测头,分别用来检测像素为 8×8,16×16,32×32 以上的目标。在小目标检测中存在大量像素小于 8×8 的目标,YOLOv5 中的 3 个检测头无法检测像素小于 8×8 的目标。浅层信息包含更多的位置信息更适合检测小目标,深层网络的语义信息丰富,但缺少位置信息,随着网络的深化浅层位置信息会不断的丢失。本文从 backbone 的第 2 层中引出小目标检测头,用以检测 4×4 像素及其以上的小目标,同时包含更丰富的位置信息,来改善检测小目标的能力。为了验证浅层检测头对小目标检测有效性,使用 Visdrone 数据集作为验证,输入图片的尺寸为 640×640,结果如表 1。

表 1 小目标检测头的影响

	方法	mAP50/%	mAP/%	Params/M	FPS
A	YOLOv5	34.5	19.0	6.9	128
B	4head	39.7	22.7	7.8	113
C	3heads	41.6	23.7	7.7	113

图 2 中 a 表示 YOLOv5s 中的 3 个检测头,b 表示增加检测像素为 4×4 的小目标检测头,c 表示去掉 b 中的大目标检测头。



(a) 3浅层检测头 (b) 4浅层检测头 (c) 3深层检测头

图 2 head 对比图

表中的 A、B、C 对应图 2 中的 a、b、c。对比表 1 中 AB 组实验,B 在 A 的基础上从 backbone 的第 2 层引出小目标检测层,在 A 的基础上 mAP50 提升了 5.4%;C 去掉 B 中的大目标的 head,在 A 的基础上 mAP50 提升了 7.1%,分析认为小目标检测头的引入更适合 YOLOv5s 检测小目标。虽然 C 组实验的检测精度最高,为了保留模型对大目标的检测能力,模型选择方式 b。

1.2 CCAT

深层网络含有丰富的语义信息,浅层信息含有丰富的位置信息。Tan 等[9]设置的加权双向特征金字塔网络(bidirectional feature pyramid network,BiFPN)融合不同层之间的信息,提高模型的性能。但传统的 BiFPN 进行输入层的拼接时,采用平均分配的原则,易造成特征融合过程中引入的混叠效应。He 等[10]在主干网络中添加残差模块,避免梯度消失与网络退化,同时残差网络结构在特征提取过程中能够降低特征信息的丢失。Chen 等[11]指出尺度特征融合并非 FPN 最重要的影响因素,分而治之将不同尺度的目标检测进行拆分处理,缓解了优化问题。

对于小目标而言,位置信息更重要,受 BiFPN、残差连接的启发,所以小目标检测需要的浅层网络所占的比重高于深层网络的比重。在 YOLOv5s 中的 Neck 部分采用图 3(a)的融合方式,将浅层网络与深层网络进行 cat,浅层信息与深层信息占比相同,对小目标检测不利。本文设计了如图 3(b)的融合方式,浅层网络输入 X_i 与深层网络输入 X_{i-1} 进行融合,融合后的输出 X_{i-2} 与 X_i 进行残差连接,保留浅层网络的位置信息,增强对小目标的检测效果。

$$X_o = f(X_i, X_{i+2}) = f(f(X_i, X_{i+1}), X_i) \tag{1}$$

式中:X_i 表示浅层的模型输入,X_{i+1} 表示深层的模型输入,X_o 表示特征融合后的模型输出。为了证明 CCAT 的有效性,在 Visdrone 数据集做了以下对比实验,结果如表 2。

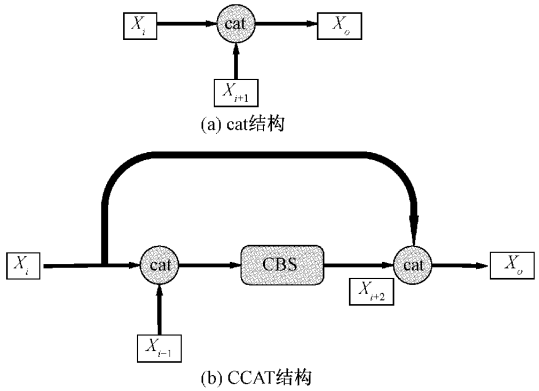


图 3 特征融合结构

表 2 CCAT 验证实验

	方法	mAP50/%	mAP/%	Params/M	FPS
A	YOLOv5s	34.5	19.0	6.9	128
B	YOLOv5s+CCAT	33.9	18.3	7.2	119
C	4head	39.7	22.7	7.8	113
D	4head+CCAT	42.3	24.3	7.9	104
E	3heads	41.6	23.7	7.7	113
F	3heads+CCAT	42.5	24.2	7.8	109

对比 AB 组实验,B 在 YOLOv5s 的 3 个 head 添加 CCAT 模块,B 组的 mAP50 降低了 0.6%;对照 CD 组实验,在 4 个检测头的 head 添加 CCAT 模块,D 组的 mAP50

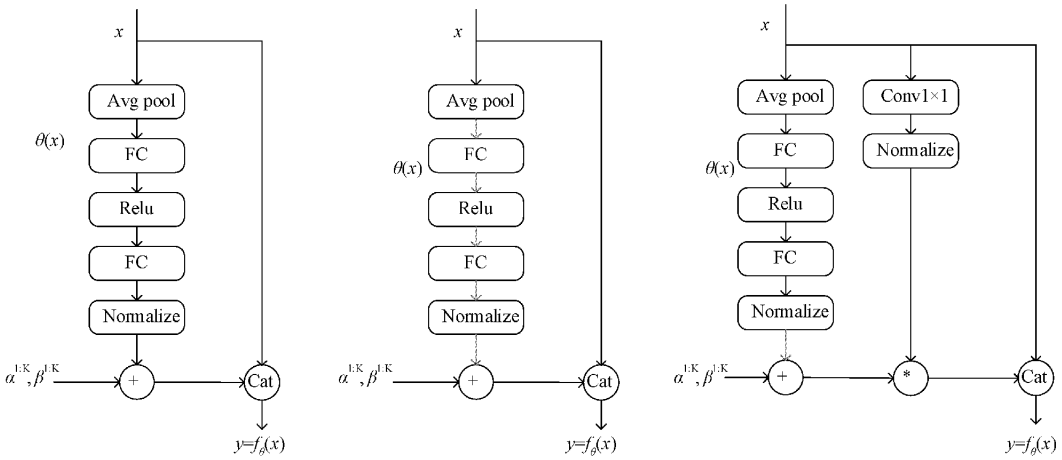
提升了 2.6%；对照 EF 组实验，去掉 4 个 head 中的大目标检测模块，在其余 3 个 head 中添加 CCAT 模块，模型得到了 0.9%的提升。通过以上 3 组实验分析认为 CCAT 模块适合用于小目标检测层，在相同网络结构中只利用 CCAT 模块替代 cat 模块后，模型的检测速度有小幅下降。同时对比 ACE 组实验，可以得到添加小目标检测层对小目标检测的效果更好。

1.3　DConv

检测头最后输出的特征层的维度为类别个数＋1 个目标置信度＋4 个位置信息，分类与回归任务的冲突是一种常见问题。Yolox 使用 double head 使用不同的卷积对分类与定位进行分离操作。面对分类与回归使用的激活函数的差异，Chen 等[8]使用动态激活函数根据输入数据的分布实现激活函数的自适应。Dai 等[12]使用 dyrelu 激活函数在性能上得到较大提升。

激活函数 RELU 应用很广泛，近几年成功的网络都采用网络都用到 RELU 作为激活函数，但无论是 RELU 还是其变体 Leaky-RELU、PRELU 都是静态的，输入数据的变化不会改变 RELU 的参数，不同的检测任务适应 RELU 的参数是不同的，Chen 等[8]认为这是不合适的，在 RELU 的基础上发表了 DY-RELU，使得 RELU 的参数随输入数据的变化而改变。

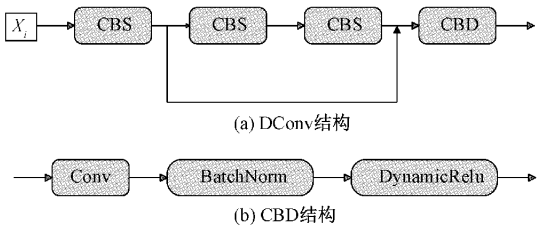
DY-RELU 有如图 4 中的 a、b、c 三种形式。a 表示空间、通道共享，输出 2K 个参数；b 表示空间共享，通道不共享，输出 2KC 个参数；c 表示空间、通道不共享，参数量极大。所以 a、b 适合用于主干网络中，b、c 适合用在输出头。Dai 等[12]指出 b 结构更适合用于图像分类中，c 结构更适合用于关键点的检测任务中。



(a) DYRELU-A结构　(b) DYRELU-B结构　(c) DYRELU-C结构

图 4　dynamic relu 结构

传统的 YOLO 系列算法采用一阶段方法在检测头部分使用相同的 RELU 激活函数同时计算分类与回归问题。本文提出了如图 5 的 DConv 结构，该结构用于 YOLOv5s 的检测头部位，替代原本的 C3 模块。结构的末尾使用 CBD 中，动态的改变 a，b 参数，达到适应该通道检测任务与数据分布的结果。



(a) DConv结构

(b) CBD结构

图 5　DConv 结构

2　实验及其相关计算

本文使用 windows10 系统，实验环境为 Python-3.9.7，pytorch1.8.0，cuda10.1，所有模型在 Tesla V100 上运行，在相同参数下（不一定最优）进行训练、验证、测试。Warmup epoch 设为 3，初始学习率为 0.01，使用余弦退火衰减策略更新学习率，epoch 设为 300，batch_size 设为 32，输入图片的尺寸设为 640×640。

2.1　相关计算及其指标

目标检测中常用的指标有平均精准率（average precision，AP）。AP 的计算需要计算检测目标的精确度 P（Precision）与召回率（Recall）。计算公式分别如下：

$$Precision = \frac{TP}{TP + FP} \times 100\% \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \times 100\% \tag{3}$$

式（2）、（3）中 TP 表示模型正确检测出来的样本个数，FP 表示模型的误检个数，FN 表示模型的漏检目标个数。AP 表示对精确度 P 对召回率在 0～1 上进行积分，计算公式如：

$$AP = \int_0^1 P(R)\mathrm{d}(R) \tag{4}$$

多目标检测常用参数为 mAP，mAP 表示平均精确度，

计算公式如:

$$mAP = \sum_{i=0}^{n} AP_i \tag{5}$$

式(5)中 N 表示种类个数,mAP50 表示交并比(intersection over union,IoU)阈值为 0.5 时,所有目标类别的平均精度;mAP 表示从 0.5～0.95 中以步长为 0.05 取出的 10 个阈值,10 个阈值下平均精确度的平均值。画面每秒传输帧数(frames per second,FPS)表示模型每秒检测图片的张数,其值越高表示模型的实时性能越好,常用来衡量模型的实时性。Params 表示模型的参数值,衡量模型的大小。

2.2 Visdrone 实验分析

本文为了证明 HCD-YOLOv5s 对 YOLOv5s 模型的提升,使用 Visdrone 无人机数据集训练 HCD-YOLOv5s 模型,Visdrone 数据集是由天津大学 AISKYEYE 实验室通过无人机摄像头捕获收集,共包含 261 908 帧和 10 209 张静态图片,一共有 12 种预测类别(ignored regions,pedestri-an,people,bicycle,car,van,truck,tricycle,awning-tricycle,bus,motor,other)。数据集中存在以下难点:大量物体;部分目标过小;数据分布不均衡;目标遮蔽严重。将训练集中的 6 471 张图片按 8∶2 的比例划分为本次实验的训练集与验证集。

为了验证 heads、CCAT、DConv 3 个模块之间的关系以及有效性。本文进行消融实验评估不同模块在相同实验条件下对目标检测算法性能的影响。选择 Ultralytics6.0 版本的 YOLOv5s 作为基准模型,输入图像的尺寸为 640×640,训练 300 个 epoch 后的结果如表 3 所示。

表 3 消融实验

	4heads	CCAT	DConv	mAP50/%	mAP/%	Params/M	FPS
A				34.5	19.0	6.9	128
B	√			39.7(+5.2)	22.7(+3.7)	7.8	113
C		√		33.9(−0.6)	18.3(−0.7)	7.2	119
D			√	34.9(+0.4)	19.2(+0.2)	11.2	94
E	√	√		42.3(+7.8)	24.3(+5.3)	7.9	104
F	√		√	43.4(+8.9)	25.3(+6.3)	12.7	93
G	√	√	√	44.7(+10.2)	26.4(+7.4)	13.0	90

对比实验 A 与 B 的结果,mAP50 增加了 5.2%,网络参数只增加了 0.8 M,这说明增加浅层检测头有利于小目标的检测。

对比实验 A、D、E 的结果,简单地在模型的检测头部分增加 DConv 模块,mAP50 只增加了 0.4%,E 在 D 的检测头部分增加 DConv 模块,模型获得大幅提升,mAP50 相对于 A 增加了 8.9%,表明 DConv 对小目标检测层的效果更加明显。

对比实验 A、C、F 的结果,C 在 A 的基础上添加 CCAT 模块,模型的 mAP50 在 A 的基础上减少了 0.6%,F 在 C 的基础上增加 CCAT 模块,模型 F 的 mAP50 在 A 的基础上增加了 7.8%,这表明 CCAT 模块适合用于小目标检测层。

对比 E、F、G 3 组实验,G 在 E、F 的基础上 mAP50 分别增加了 1.3%、2.4%,这表明小目标检测头、CCAT、DConv 模块之间相互促进,最后 G 在 A 的基础上 mAP50、mAP 分别提升了 10.2%、7.4%,虽然模型的 FPS 有所下降,仍然能达到 90 s/帧,HCD-YOLOv5s 算法精度得到较大提升,同时也保证了算法的实时性。

本实验模块应用于 YOLOv5m 时,HCD-YOLOv5m 的 mAP50 能达到 48%,YOLOv5m 的 mAP50 为 40%,模型精度仍然提高了 8%。

本实验检测对象有 12 类,其中 ignored regions,other 两类的 AP 值较小,拉低了 HCD-YOLOv5s 的 mAP 值,去掉这两类检测对象后,HCD-YOLOv5s 在 1 440×1 440 的输入尺寸下 mAP50 分别可以达到 61.5%,优于奉志强[13]、陈旭[13]等改进后的 YOLOv5s 的检测效果。综上证明了改进模型的有效性。

2.3 不同输入尺寸对精度的影响

不同的尺寸图片的输入,对模型的检测精度与检测速度影响较大,本文以 YOLOv5s 与 HCD-YOLOv5s 为基础,探究不同尺寸图片对改进模型的速度与精度的影响,实验结果如图 6 所示。

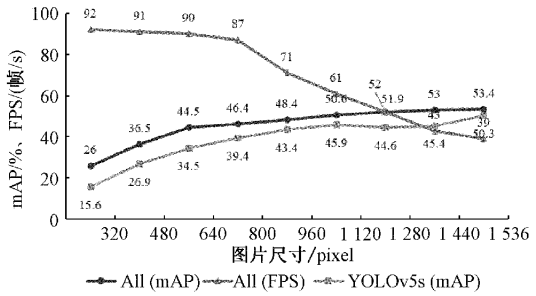


图 6 尺寸对比图

图 6 中可以看出,在不同的图像输入尺寸的条件下,HCD-YOLOv5s 的 mAP50 相比于 YOLOv5s 都得到较大提升。当图片的尺寸越来越大时,HCD-YOLOv5s 与

YOLOv5s 的 mAP50 都得到了较大提升，这是由于输入的尺寸越大，图片中的小目标越少，所以检测精度得到提升。HCD-YOLOv5s 的检测速度随图片尺寸的增大而减少，当图片的尺寸小于 800 时，模型的检测速度较少较为平缓，能稳定在 85 以上。为了同时满足精度要求与速度要求，模型选择 640×640 的尺寸作为输入更为合理。

2.4　对比实验

为了验证改进后的算法对比其他算法在小目标检测任务中的优越性，本文与各种先进的目标检测算法在 Visdrone 数据集上进行比较，输入图片的尺寸为 640×640，评价指标为 mAP50 与 mAP。

表 4 数据显示 HCD-YOLOv5s 的检测精度明显高于其他主流算法，比同阶段算法中的最高的 Carafe 高出 4.5%；mAP50 比高出 yoloxs 算法 9.0%；在同一算法中比 YOLOv5m 高出 4.7%。同时检测速度达到 90FPS 满足实时检测的要求，证明出了该算法在检测小目标的优越性。

表 4　不同检测算法的对比

方法	mAP50/%	mAP/%
YOLOv5s	34.5	19.0
YOLOv5m	40.0	23.1
ATSS[14]	22.1	12.7
RetinaNet[15]	11.0	6.7
Faster-RCNN[16]	33.2	17
Carafe[17]	40.2	24.5
Autoassign[7]	36.1	20.6
Yoloxs	35.7	20.2
HCD-YOLOv5s	44.7	26.4

2.5　检测效果对比

小目标检测中常出现遮蔽、拥挤、光线、模糊等 4 种情况，会严重影响小目标的检测精度。为了验证改进算法在实际小目标应用场景中的效果，本文随机从 Visdrone 中抽取 500 张图片，输入尺寸为 640×640，对图片进行可视化。左侧为改进后的 HCD-YOLOv5s 算法，右侧为 YOLOv5s 算法。

图 7 中第 1 行图片表示改进后的算法在遮蔽情况下的检测效果，第 2 行图片表示高空位置的小目标检测效果，第 3 行图片表示模糊环境下的检测效果，第 4 行图片表示光线变化环境下的检测效果。对比以上 4 种情况，改进后的算法对小目标在复杂场景下的检测效果有不少的提升。

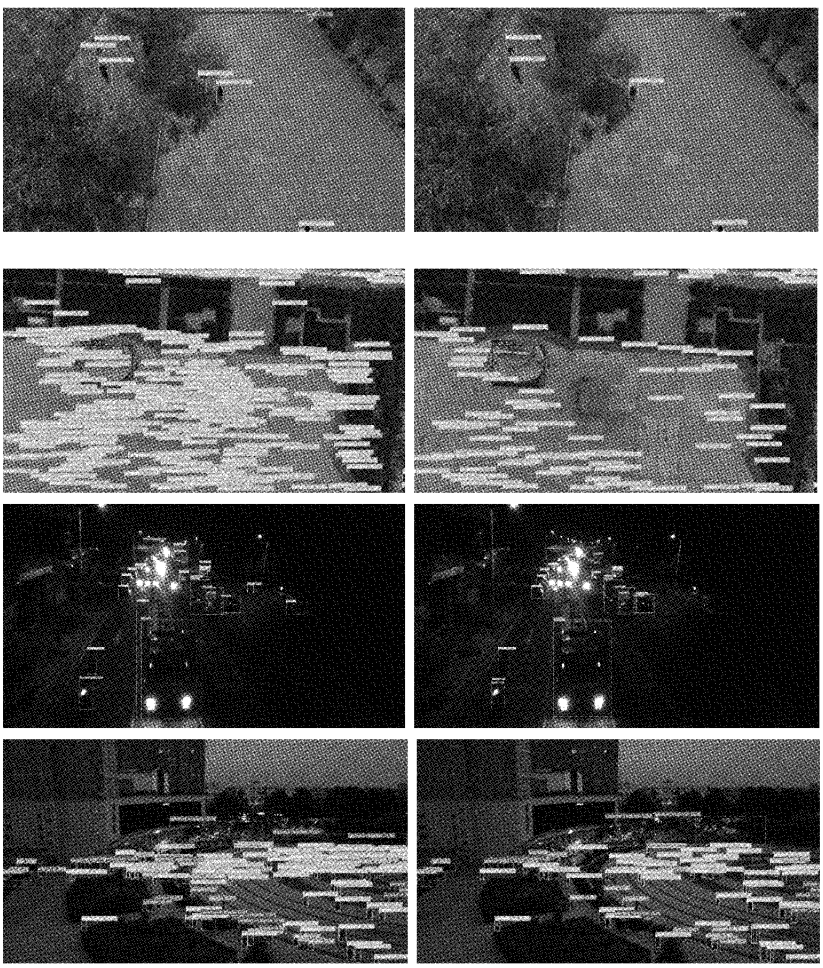


图 7　检测效果对比

2.6 Dota 验证效果

为了进一步验证 HCD-YOLOv5s 算法改进的有效性,在小目标数据集上再次进行验证,DOTA-v1.0 中的分布如表 5 所示,含有大量的 10～50 个像素的目标,因此适合做本模型的验证数据集,验证 HCD-YOLOv5s6 在小目标数据集上的效果。

表 5 Dota 的数据分布

数据集	10～50 像素	50～300 像素	超过 300 像素
DOTA-v1.0	0.57	0.41	0.02

DOTA-v1.0 是武汉大学从 Google Earth、JL-1 卫星拍摄,中国资源卫星数据和应用中心的 GF-2 卫星拍摄上收集的数据,其中包含来自不同传感器和平台的 2 806 幅航拍图像。每个图像的尺寸在约 800×800 到 4 000×4 000 像素的范围内,并且包含呈现各种尺度、方向和形状的物体。其中一共包含了 15 类物体(即飞机,轮船,储罐,棒球场,网球场,篮球场,地面跑道,港口,桥梁,大型车辆,小型车辆,直升机,环形交叉路口,足球场和篮球场),DOTA 数据集的尺寸太大,需要先将图片分割,之后再输入模型进行训练。分割之后的图片为 21 046 张,训练集为 15 749 张,测试集 5 297 张。

表 6 表示改进后的模块在 DOTA-v1.0 数据集的表现,对比 AB 组 mAP 增加了 0.9%,表明了浅层检测头的有效性。对比 BC 组实验,在 4 个检测上的基础上添加 CCAT 模块,模型的 mAP 增加了 0.2%,再次证明了 CCAT 模块更适合小目标检测层的特征融合。最后对比 HCD-YOLOv5s 与 A、B、C 3 组实验,证明了 3 个模块的之间相互促进作用,最后 HCD-YOLOv5s 在 YOLOv5s 的基础上 mAP50、mAP 分布提升了 1.8%、2.0%,表明了 HCD-YOLOv5s 对小目标有更好的识别能力。

表 6 DOTA 的检测效果

实验组	方法	mAP50/%	mAP/%	Params/M	FPS
A	YOLOv5s	71.5	46.1	6.9	128
B	heads	72.4	47.2	7.7	109
C	4heads+CCAT	72.6	47.0	7.8	96
HCD-YOLOv5s	4heads+DConv+CCAT	73.3	48.1	13	90

3 结 论

针对小目标下采样的位置信息丢失、以及任务分布不同本文提出了 HCD-YOLOv5s 检测算法。首先从浅层网络中增加一个检测头,用以检测小目标;然后针对位置信息对小目标更重要,语义信息对大目标检测更重要,同时针对特征融合时造成的特征混淆,设计了 CCAT 模块,在 Visdrone 上得到了 CCAT 模块对小目标检测层的增益更大。最后针对检测任务的不同以及数据分布的不同适应的激活函数不相同,结合 dynamic relu 设计了 DConv 模块。

在 Visdrone 无人机小目标数据集上进行消融实验,得到 3 个改进点对模型的 mAP50 与 mAP 都有不同程度的提升,3 个改进点的作用是相互独立的,一起使用给模型带来 10.2%的增益,由于模型参数的增加,HCD-YOLOv5s 的检测速度有所降低,FPS 仍然可以达到 90,满足实时检测的要求。然后将主流的检测模型与改进模型进行比较,结果表明本模型更适合检测小目标。HCD-YOLOv5s 在去除影响精度的 ignored regions、other 两类后,在输入为 1 440×1 440 的图片下 mAP50 可以达到 61.5%,高于其他文献对 YOLOv5s 的改进。

将 HCD-YOLOv5s 应用在航空数据集 dota1.0 上,模型的 mAP50 与 mAP 同样得到 1.8%、2.0%的提升,再次证明了模型更适合对小目标的检测。

最后需要指出在 DConv 模块中含有 FC 模块使模型的参数量与计算量增大,因此可以进一步考虑将此模块进行轻量化;另外本文未将此模型应用于其他检测模型,可以继续考虑这 3 个改进点对其他模型的影响。

参考文献

[1] BODLA N, SINGH B, CHELLAPPA R, et al. Soft-NMS—improving object detection with one line of code[C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 5561-5569.

[2] ZHANG H, ZU K, LU J, et al. Epsanet: An efficient pyramid split attention block on convolutional neural network[J]. ArXiv Preprint, 2021, ArXiv: 2105.14447.

[3] 奉志强,谢志军,包正伟,等. 基于改进 YOLOv5 的无人机实时密集小目标检测算法[J]. 航空学报,2022:1-15,DOI: 10.7527/S1000-6893.2022.27106.

[4] GONG Y, YU X, DING Y, et al. Effective fusion factor in FPN for tiny object detection[C]. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2021: 1160-1168.

[5] GE Z, LIU S, WANG F, et al. YOLOx: Exceeding yolo series in 2021 [J]. ArXiv Preprint, 2021, ArXiv:2107.08430.

[6] WANG J, CHEN K, XU R, et al. Carafe: Content-aware reassembly of features[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 3007-3016.

[7] ZHU B, WANG J, JIANG Z, et al. Autoassign: Differentiable label assignment for dense object detection [J]. ArXiv Preprint, 2020, ArXiv: 2007.03496, DOI: 10.48550/arXiv.2007.03496.

[8] CHEN Y, DAI X, LIU M, et al. Dynamic relu[C]. European Conference on Computer Vision, Springer, Cham, 2020: 351-367.

[9] TAN M, PANG R, LE Q V. Efficientdet: Scalable and efficient object detection[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 10781-10790.

[10] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.

[11] CHEN Q, WANG Y, YANG T, et al. You only look one-level feature[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 13039-13048.

[12] DAI X, CHEN Y, XIAO B, et al. Dynamic head: Unifying object detection heads with attentions[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 7373-7382.

[13] 陈旭，彭冬亮，谷雨. 基于改进 YOLOv5s 的无人机图像实时目标检测[J]. 光电工程，2022，49(3)：13.

[14] ZHANG S, CHI C, YAO Y, et al. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 9759-9768.

[15] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 2980-2988.

[16] 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, 2017, 39(6), DOI: 10.1109/TPAMI.2016.2577031.

[17] WANG J, CHEN K, XU R, et al. Carafe: Content-aware reassembly of features[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 3007-3016.

作者简介

陈俊豪，硕士研究生，主要研究方向为深度学习、图像识别。

E-mail: 1471272619@qq.com

冉险生，副教授，研究生导师，主要研究方向为智能车辆技术、车辆动力学。