

Road Small Target Detection Algorithm Based on Improved YOLO v5^{*}

NIU Weihua^{*}, YIN Miaomiao

(Department of computer science, North China Electric Power University, Baoding Hebei 071003, China)

Abstract: Aiming at the problems of low detection accuracy of small targets, missed detection and false detection in the detection process, an improved YOLO v5 algorithm for small target detection in road scenes is proposed. Firstly, the ConvFocus module is designed in the YOLO v5 model. In the feature fusion part, bilinear interpolation up-sampling operation is used and 5 CBAM attention mechanism modules are inserted to reduce the loss of small target features and enhance the semantic and location information of small targets; In addition, a small target detection layer with a size of 160×160 is added, which uses the rich semantic and location information contained in the shallow feature layer to accurately locate and identify small targets. The KITTI data set is used in the experiment, which is processed and divided into the training set and the test set. The experimental results show that the mAP of the improved YOLO v5 model on the KITTI dataset reaches 96.5%, which is significantly improved compared with the original YOLO v5 algorithm, and the detection effect of small objects is better.

Key words: small target detection; YOLO v5; ConvFocus; bilinear interpolation up-sampling; attention mechanism

EEACC: 7230

doi: 10.3969/j.issn.1004-1699.2023.01.006

基于改进 YOLO v5 的道路小目标检测算法^{*}

牛为华^{*}, 殷苗苗

(华北电力大学计算机系, 河北 保定 071003)

摘要: 针对小目标检测精度低、检测过程中易出现目标漏检和误检等问题, 提出一种改进的 YOLO v5 道路小目标检测算法。首先在 YOLO v5 中设计了 ConvFocus 模块; 在特征融合部分使用双线性插值上采样操作, 并插入 5 个 CBAM 注意力机制模块, 减少小目标特征的丢失, 增强小目标语义及位置信息; 另外增加了一个大小为 160×160 的小目标检测层, 利用浅层特征层中包含的丰富的语义及位置信息精确定位识别小目标。实验使用 KITTI 数据集, 对该数据集进行处理并划分训练集和测试集。实验结果表明, 改进的 YOLO v5 模型在 KITTI 数据集上平均精度均值达到 96.5%, 与原 YOLO v5 算法相比有显著提高, 小目标检测效果更好。

关键词: 小目标检测; YOLO v5; ConvFocus; 双线性插值上采样; 注意力机制

中图分类号: TP391.4; TP391.9

文献标识码: A

文章编号: 1004-1699(2023)01-0036-09

近年来, 随着计算机技术飞速进步, 计算机在各大领域得到充分推动与发展。计算机视觉领域是计算机一大重要领域, 目标检测作为该领域之一, 可应用于缺陷检测^[1-2]、状态监测^[3]、交通监测^[4]、自动驾驶^[5]及遥感^[6]等领域, 在生活的方方面面发挥了不可或缺的作用。

目标检测在最初期使用手工方法设计检测目标, 近年来, 深度卷积网络快速发展并应用于方方面面, 基于深度学习的目标检测算法相继被提出^[7-8]。现阶段基于深度学习的目标检测算法可分为两大类^[9], 第一类是基于区域提取的目标检测算法, 也可称为两阶段目标检测算法, 代表算法有 R-CNN^[10]、Fast R-

CNN^[11]和 Faster R-CNN^[12]等, 该类算法在目标检测上分两步, 首先在图片上生成候选区域, 然后利用回归等方法对候选区域进行分类及预测; 第二类是基于回归的目标检测算法, 也可称为一阶段目标检测算法, 代表算法有 SSD^[13]、YOLO^[14]、YOLO v2^[15]、YOLO v3^[16]、YOLO v4^[17]、YOLO v5 等, 该类算法无需提前生成候选区域, 而是直接在卷积后的图像上预测目标的位置和分类。近两年, 多数学者在前人目标检测算法基础上继续研究^[18-20], 结合所用算法的优缺点以及检测目标的特点设计有效模块。如 Zhang 等^[21]结合一阶段 SSD 算法和二阶段 Faster R-CNN 算法优缺点, 设计了 RefineDet 目标检测网络; RezaTofighi 等^[22]

项目来源: 中央高校基本科研业务费专项资助项目 (2017MS156)

收稿日期: 2022-01-11 修改日期: 2022-04-22

针对目标检测中 IoU 存在的缺点,设计 GIoU 指标,解决了梯度回传缺失及两框重合度大小的精确计算问题;Joseph 等^[23]针对人类具有识别未知物体的本能,设计了一种开放世界目标检测算法。

尽管目标检测算法已很成熟,在计算机视觉领域取得了一定成绩,但在小目标检测方面仍存在一定问题与困难。现阶段,小目标检测多以主流目标检测算法为基础,从特征提取、特征融合、注意力机制及检测头等^[7]方面进行改进以提升小目标检测能力。如顾恭等^[24]设计了 Vehicle-YOLO 模型,该模型主要在以 YOLO v3 为基础的网络上更改残差网络来增强小目标特征提取能力;Mahto 等^[25]设计了 Refined YOLO v4 模型,以 YOLO v4 网络为基础,从锚框、后处理算法和注意力三方面进行改进提升小目标检测。

小目标对象分辨率低,像素占比小,多数算法在使用卷积网络进行特征提取时,所提信息有限,不足以对小目标进行表达^[26]。随着卷积神经网络的多次下采样,特征图尺寸不断减小,小目标的特征难以提取,导致小目标检测过程中出现严重的误检、漏检^[27]。

为降低小目标漏检、误检率,提高小目标检测平均精度,本文以 YOLO v5 算法模型为基础,提出了一种应用于道路场景小目标检测的改进 YOLO v5

算法,相比 Vehicle-YOLO、Refined YOLO v4 等模型具有更高的检测速度及平均精度。本文算法引入双线性插值法进行上采样操作以缓解多数算法中存在的特征损失严重问题;另外对 YOLO v5 算法网络结构进行改进,将原始的三层检测层扩展到四层,设计了一个 ConvFocus 模块替换原模型的 Focus 模块,并在特征融合部分插入 5 个 CBAM^[28]注意力机制模块,改善了多数算法存在的主干网络特征提取能力不足、融合网络有用信息提取不聚焦等缺陷。本文对 KITTI^[29]数据集进行处理,并在处理后的数据集上将原算法、改进算法以及其他目标、小目标检测算法做实验对比,由结果可得,改进的 YOLO v5 算法在保证原模型较好检测速度的同时,对于小目标检测评价指标更优,检测各个类别的平均精度和平均精度均值提升明显。

1 原始 YOLO v5 算法

YOLO v5 由 Ultralytics LLC 公司提出,是一个基于 YOLO v4 改进的轻量型版本,该网络相较于 YOLO 之前版本,检测性能有进一步的提升。对于目标检测,YOLO v5 在检测精度及速度上都有一定的提升,并满足轻量型模型特点,其网络结构分为四部分:Input、Backbone、Neck、Prediction,模型网络结构如图 1 所示。

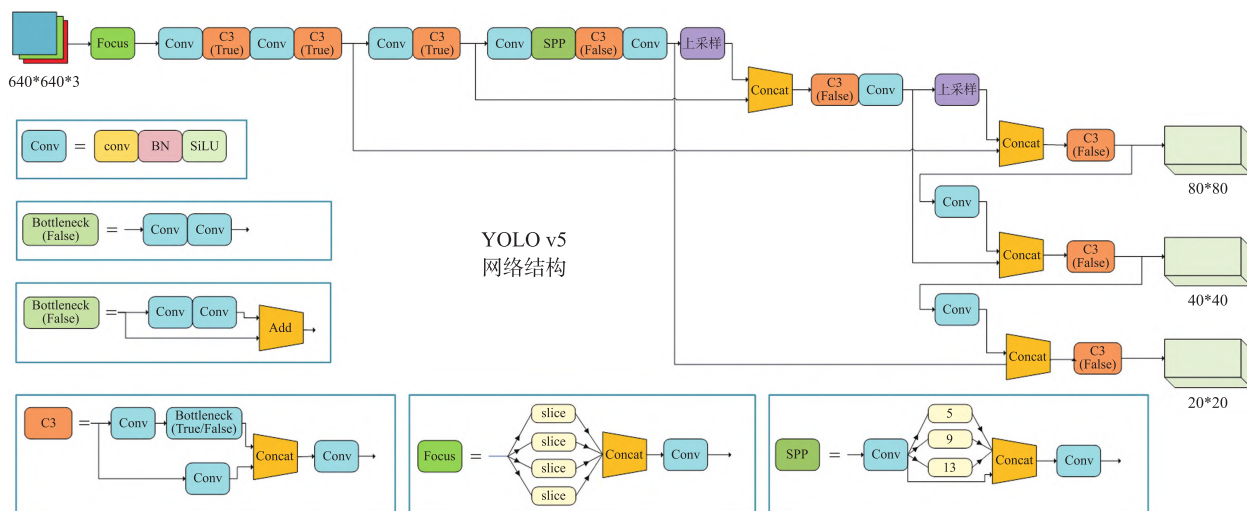


图 1 YOLO v5 网络结构

Input: Input 端包含 Mosaic 数据增强、自适应锚框计算以及自适应图片缩放三部分。Mosaic 数据增强技术采用随机裁剪、随机缩放、随机排列方法对四幅图像进行拼接,有利于小目标的检测。对于不同数据集,YOLO v5 算法将根据自适应锚框计算该数据集所对锚框。原模型中锚框效果已达到很好的效果,故本文使用原模型锚框。自适应图片缩放操作

将输入的不同尺寸图像统一缩放至同一尺寸,本文图像缩放尺寸至 640×640。

Backbone: Backbone 部分由 Focus 结构和特征提取结构组成,图像经过 Focus 结构后,像素尺寸缩减一半。Focus 结构主要对图像进行切片,当输入图像为 640×640×3 时,经过 Focus 的切片、concat 以及卷积操作,输出 320×320×32 的特征映射。Back-

bone 部分的特征提取结构由 C3、Conv 和 SPP 模块组成,其中,C3 模块是由 Conv 和若干 Bottleneck 结构组成的残差结构,在保证模型轻量化的前提下可增强特征提取能力。Conv 模块相当于 2 倍下采样操作,经过一次 Conv 模块,特征图通道数增加一倍。SPP 模块先使用 Conv 模块提取特征,再经过卷积核尺寸分别为 5,9,13 的三个最大池化操作,连同输入一起进行 concat 连接,最后再经过一个 Conv 模块将输出恢复至输入尺寸。

Neck:Neck 部分由 FPN+PAN 结构组成,融合 Backbone 部分提取的特征。FPN 和 PAN 是两个方向相反的特征金字塔网络,分别从相反方向进行层间特征融合。其中,FPN 自上而下,使用上采样操作逐层扩充特征图尺寸;PAN 自下而上,通过卷积操作缩小特征图尺寸。FPN 与 PAN 结构结合后,网

络的特征融合能力进一步加强。

Prediction:Prediction 包括 Bounding box 损失和 NMS 非极大值抑制算法。在目标检测的预测阶段,经过 NMS 算法后,冗余预测框被去除,留下最优目标框。

2 YOLO v5 网络模型改进

本文针对小目标检测存在的问题,分别从四个方面对 YOLO v5 算法进行改进。首先使用双线性插值法上采样操作扩大特征图,其次改进网络结构,将 Focus 模块替换为设计的 ConvFocus 模块,增加一个小目标检测层,并在特征融合部分共插入 5 个 CBAM 注意力机制模块,从通道和空间两个方面聚焦有用特征,加强网络特征提取能力。改进的 YOLO v5 算法网络结构如图 2 所示。

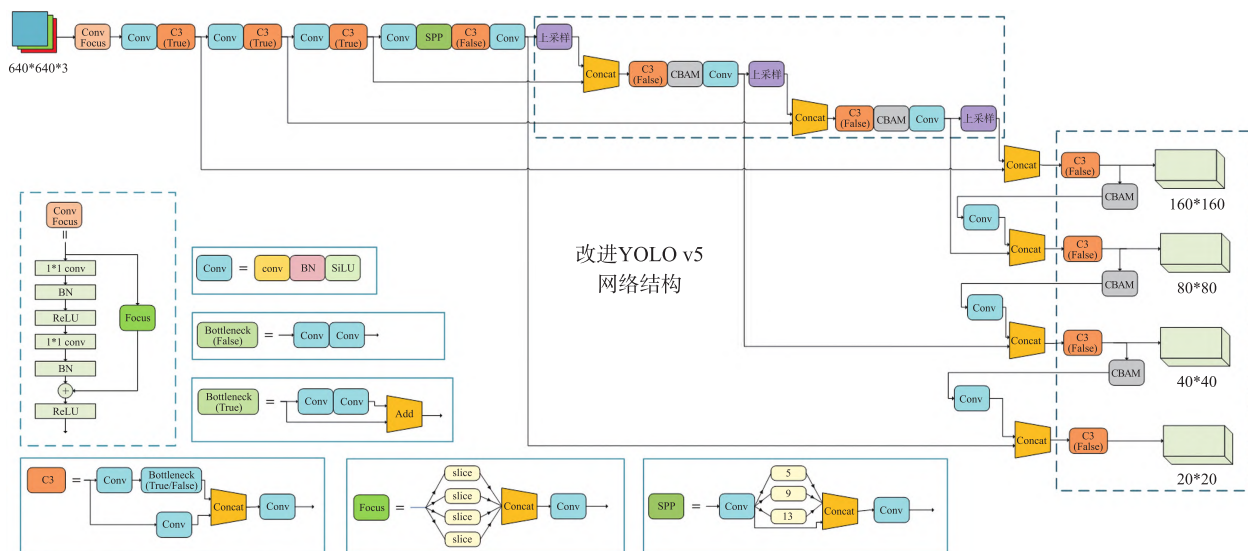


图 2 改进的 YOLO v5 网络结构

2.1 双线性插值法上采样操作

YOLO v5 模型中,在特征融合部分使用最近邻插值法实现特征图上采样。最近邻插值上采样法没有计算新插值特征图像素矩阵中点的数值,直接在原特征图像的位置信息找对应像素点的值,采样速度较快,但这种采样方法极易破坏原特征图的像素关系,导致图像失真,造成小目标特征及位置信息的损失,增大小目标漏检及误检率。针对该问题,本文引入双线性插值法上采样操作实现特征图的上采样,扩大特征图尺度。

双线性插值上采样在计算新特征图像素点值时,从原图像中选择 4 个点并在两个方向进行插值,是单线性插值的拓展。该采样方法通过两个方向上的三次插值,在保证原特征图像素间关系的同时对特征图进行上采样,可提高采样后特征图的清晰度,

减少小目标特征信息及位置信息的损失,降低小目标检测漏检及误检率。该采样方法原理如下。

假设在原图像选择的四个点分别为 $Q_{11}(x_1, y_1)$, $Q_{21}(x_2, y_1)$, $Q_{12}(x_1, y_2)$, $Q_{22}(x_2, y_2)$, 设 $R_1(x, y_1)$, $R_2(x, y_2)$, $P(x, y)$ 。首先在 x 方向上做两次插值得:

$$f(R_1) \approx \frac{x_2 - x}{x_2 - x_1} f(Q_{11}) + \frac{x - x_1}{x_2 - x_1} f(Q_{21}) \quad (1)$$

$$f(R_2) \approx \frac{x_2 - x}{x_2 - x_1} f(Q_{12}) + \frac{x - x_1}{x_2 - x_1} f(Q_{22}) \quad (2)$$

然后在 y 方向进行插值,得:

$$f(P) \approx \frac{y_2 - y}{y_2 - y_1} f(R_1) + \frac{y - y_1}{y_2 - y_1} f(R_2) \quad (3)$$

整理可得

$$f(x, y) \approx \frac{f(Q_{11})}{(x_2 - x_1)(y_2 - y_1)} (x_2 - x)(y_2 - y) +$$

$$\frac{f(Q_{21})}{(x_2-x_1)(y_2-y_1)}(x-x_1)(y-y_1)+$$

$$\frac{f(Q_{12})}{(x_2-x_1)(y_2-y_1)}(x_2-x)(y-y_1)+$$

$$\frac{f(Q_{22})}{(x_2-x_1)(y_2-y_1)}(x-x_1)(y-y_1) \quad (4)$$

2.2 网络结构改进

2.2.1 设计 ConvFocus 模块

在原 YOLO v5 模型中,输入网络的图片在 Backbone 部分首先会进入 Focus 模块。Focus 模块对每个通道的数据进行切片操作,通过隔像素取值,每一通道均得四个切片,每一切片相当于 2 倍下采样所得。所有切片将按照通道 concat 在一起,并通过卷积操作实现下采样。Focus 操作类似于计算量和参数量都较少的下采样操作,但在进行小目标检测时,直接对输入图像进行分片下采样操作,会损失原图像中小目标信息,造成小目标特征提取不充分,出现小目标漏检问题,影响小目标检测的精确度。针对该问题,设计了 ConvFocus 模块,如图 3 所示,该模块借鉴了 ResNet^[30] 的核心结构,将 Focus 模块和卷积操作并行连接,实现残差结构。在 ConvFocus 模块中,左侧分支包含两个 1×1 卷积,在不增大参数量和计算量的前提下实现图像的降维及升维;右侧分支是原 Focus 模块。将左右两侧分支用 add 方法连接融合并进行 ReLU 激活,通过融合连接,既增多了图像的特征数,也实现了在降低图像复杂度的前提下保持原图特征信息的完整性,减少小目标的信息损失。

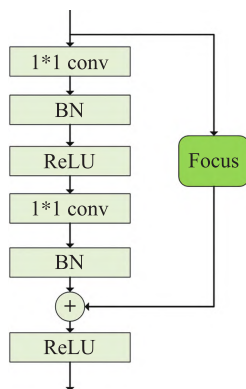


图 3 ConvFocus 模块

2.2.2 增加小目标检测层

YOLO v5 特征融合部分包含 FPN+PAN 结构,该部分进行两次上采样和两次下采样,融合深层特征与浅层特征。特征融合完成后,在检测层输出大小为 80×80 、 40×40 、 20×20 的 3 种不同尺度的特征图。特征图尺度越大,感受野越小,包含目标位置信

息和局部特征细节越详细,利于检测小目标;特征图尺度越小,感受野越大,语义信息越丰富,但局部细节不明显,适合检测大目标。为了提高小目标检测的平均精度,避免由于下采样导致目标的局部细节特征损失过大、小目标信息丢失,对原 YOLO v5 模型网络结构进行改进,在原网络中添加一层特征图尺度为 160×160 的小目标检测层,该层经过深层特征的传递以及与浅层特征的融合,包含更多的小目标轮廓信息以及位置信息,利于小目标定位与识别,降低小目标的漏检、误检。

2.2.3 插入注意力机制

注意力机制是对有限的输入进行权重偏好的选择,使用注意力机制,可更快地对输入选择权重偏好,提取感兴趣区域。在目标检测中,引入注意力机制,可以聚焦图像的局部信息,提取出检测中感兴趣区域。因此,本文在特征融合部分,共插入 5 个 CBAM 注意力机制模块,分别位于每个 C3 模块后,在通道和空间两个不同维度实现特征细化与提取,聚焦图像局部信息,降低小目标特征提取时的损失,增强网络表达能力,提升小目标检测的平均精度。

CBAM 是一个轻量型即插即用的注意力机制模块,如图 4 所示,包括通道注意力模块和空间注意力模块,依次在通道和空间上两个维度实现映射。

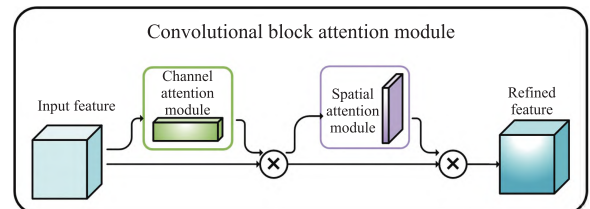


图 4 CBAM 注意力机制模块

CBAM 的通道注意力模块和 SE-Net^[31] 的通道注意力机制模块相像,如图 5 所示。通过通道注意力模块后,模型可学习不同通道特征的重要程度,从而更关注特征信息强的通道,抑制特征信息较少且不重要的通道。在 CBAM 的通道注意力模块,输入特征图 F ,首先会分别经过全局最大池化和全局平均池化获得通道注意力的一维向量,实现空间维度的压缩,聚合特征映射的空间信息,然后经过共享感知机 MLP 模块,将经过 MLP 后输出的两条分支加

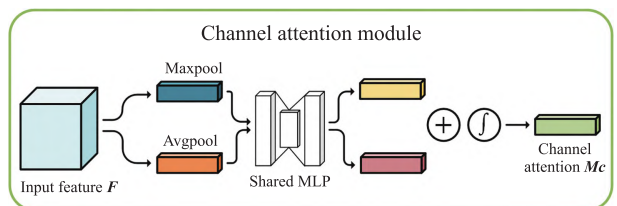


图 5 通道注意力模块

和,使用 sigmoid 激活函数,得到权重系数 M_c 。通道注意力机制表达式如下:

$$M_c(F) = \sigma \{ \text{MLP} [\text{AvgPool}(F)] + \text{MLP} [\text{MaxPool}(F)] \} = \sigma \{ W_1 (W_0(F_{\text{avg}}^C)) + W_2 (W_0(F_{\text{max}}^C)) \} \quad (5)$$

将通道注意力模块输出的权重系数 M_c 与原输入特征 F 进行相乘加权,得到新的特征 F' 并输入空间注意力模块(如图 6 所示)。空间注意力模块输出权重系数 M_s ,与输入的新特征 F' 相乘得到最终结果,该模块实现通道维度上的压缩。对于输入特征 F' ,首先进行全局最大池化和全局平均池化,并按通道拼接,最后通过卷积核为 7×7 的卷积层和激活函数得到权重系数 M_s 。空间注意力机制表达式如式(6):

$$M_s(F') = \sigma \{ f^{7 \times 7} [\text{AvgPool}(F'), \text{MaxPool}(F')] \} = \sigma \{ f^{7 \times 7} [(F'^S_{\text{avg}}; F'^S_{\text{max}})] \} \quad (6)$$

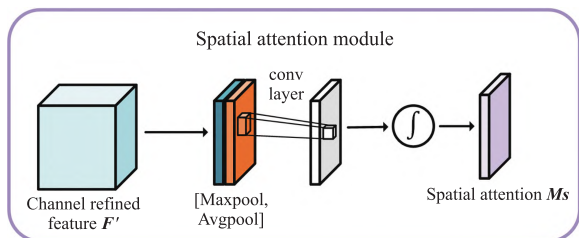


图 6 空间注意力模块

3 实验设置与结果分析

3.1 实验环境及参数配置

本次实验基于 Ubuntu 18.04.6 操作系统,使用 Python 语言实现,应用深度学习框架 Pytorch 1.8.0 搭建网络模型,Python 3.8 编译程序,硬件环境为 CPU: Intel(R) Core(TM) i7-9700 CPU@ 3.00GHz、31.2GiB 内存、NVIDIA GeForce RTX 2080 Ti。

网络训练阶段,将批量大小设置为 8,初始学习率设置为 0.01,动量为 0.937,权重衰减系数为 0.000 5。使用 Warmup 预热学习率方式使模型逐渐趋于稳定,当 Warmup 将学习率调整到一定数值时,开始进行余弦退火方式调整学习率。

3.2 数据集

小目标指在图像中分辨率低,像素占比小的物体,其定义^[8]方式有两种,即相对尺寸定义和绝对尺寸定义。相对尺寸根据国际组织 SPIE 将 256×256 的图像中像素少于 80 的目标称为小目标;绝对尺寸根据 COCO^[32] 数据集定义,尺寸小于 32×32 像素的目标即可称为小目标。此外,小目标也可根据场景进行定义,如 Braun 等^[33] 针对交通场景数据,定义像素在 30 到 60 的物体为小目标;行人识别数据集 CityPersons^[34] 中定义像素高度小于 75 的为小目标。

本次实验使用 KITTI 数据集,部分图片如图 7 所示,该数据集包括从市区、乡村以及高速公路等多个场景中获取的现实图像数据,图像像素大小为 $1\,224 \times 370$,每幅图像最多包含 15 部车辆及 30 名行人,多数目标像素低于 60×25 或 45×30 ,可定义为道路小目标。原数据集共包含 Car、Van、Truck、Pedestrian、Pedestrian(sitting)、Cyclist、Tram 以及 Misc 八个类别,本实验对原数据集进行了处理,将 Cyclist 和 Misc 这两类目标标注不清晰且数据量较少的数据删除,并将 Pedestrian 和 Pedestrian(sitting)进行合并。从处理后的数据集中选取 7481 张图片做为实验数据,其中 5 984 张图片用于训练,1 497 张图片用于测试。



图 7 KITTI 数据集示意图

3.3 评价指标

本实验选用平均精度 AP、平均精度均值 mAP 作为评价指标。平均精度 AP 根据每个类别计算,每个类别的平均精度值是该类别精确率-召回率曲线的面积值,平均精度均值 mAP 表示所有类别平均精度的一个

平均值。其中,在精确率-召回率曲线中,精确率表示实际正样本占预测正样本的比例,召回率代表实际正样本中被正确检测的样本的比例。其计算公式如下:

$$\text{mAP} = \frac{\sum(\text{AP})}{N} \quad (7)$$

$$P = \frac{TP}{TP+FP} \quad (8)$$

$$R = \frac{TP}{TP+FN} \quad (9)$$

式中: N 表示数据类别总数, TP 表示被正确识别的正样本数, FP 表示被错误识别的负样本数, FN 表示被错误识别的正样本数。

3.4 实验结果

3.4.1 本文算法检测结果

在经本文方法处理的数据集上使用训练集分别训练 YOLO v5 算法和改进后的算法,然后分别在训练好的模型上用测试集验证。YOLO v5 原算法和改进后的算法在测试集上的检测结果如表 1 所示。

表 1 YOLO v5 算法改进前后检测结果对比

算法模型	AP/%					mAP/%
	Van	Tram	Car	Pedestrian	Truck	
YOLO v5	96.5	96.9	97.7	87.1	97.3	95.1
改进的 YOLO v5	97.5	98.5	98.0	89.3	99.2	96.5

由表 1 得,改进后 YOLO v5 算法的 mAP 指标提高了 1.4%,数据集中 Van、Tram、Car、Pedestrian、Truck 这五大类别的 AP 值分别提高了 2.0%、1.6%、1.7%、2.2%、1.9%。改进的 YOLO v5 算法对 Pedestrian 类别的检测 AP 值提升最大, Pedestrian 在 KITTI 数据集中

目标尺寸较小,单张图片上目标易集中重叠,识别困难程度大,由此可验证改进后的 YOLO v5 算法对于小目标的检测有一定的提升。

YOLO v5 原算法和改进后的算法网络在测试集上的 P-R 曲线如图 8 所示。

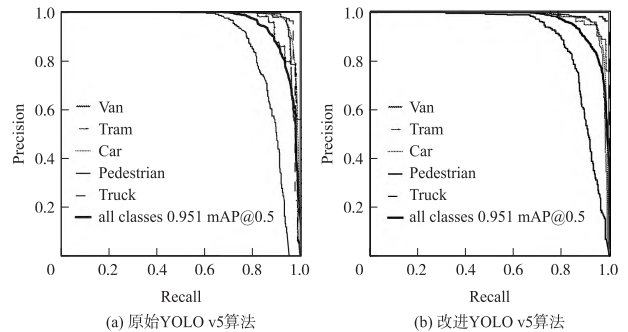


图 8 P-R 曲线示意图

由图 8 得,在改进后的 YOLO v5 算法测试所得的 P-R 曲线中,随着召回率的增加,各个类别以及总类别的精确率下降趋势更快,改进的 YOLO v5 算法针对每个类别以及总类别的 P-R 曲线面积更大,曲线收敛效果更好,显然其对小目标的检测具有更好效果。

本文采用训练完成后的 YOLO v5 以及改进 YOLO v5 算法对从测试集中选取的四组图片进行检测,检测结果如图 9 所示。



图 9 检测结果示意图

由图 9 可看出改进后算法模型的小目标检测得分相较原始 YOLO v5 算法检测得分更高,并且在原始算法中未定位识别的小目标也得以识别,改进后的算法模型误检和漏检均得到减少。

3.4.2 不同模块对检测的提升

将本文中的改进模块依次加入到原 YOLO v5 模型中,得到的检测结果如表 2 所示。

由表 2 可以看出,ConvFocus 模块对 YOLO v5 算法的小目标检测能力提升较多,对于行人这类极小目标的检测能力更强,行人类别检测的平均精度提升了 1.2%。将 YOLO v5 网络模型增加一个小目标检测层可使检测的 mAP 值提升 0.5%,在增加小目标检测层的基础上,使用双线性插值上采样操作并插入 CBAM 注意力机制模块, mAP 值再次提升

0.2%,最后将 Focus 模块替换为设计的 ConvFocus 模块,模型的 mAP 上升到 96.5%。由表 2 可得,随着每一模块的加入,模型小目标检测的能力逐渐增强,检测的平均精度均值 mAP 指标也越来越高。

表 2 不同模块加入的效果对比

算法 模型	AP/%					mAP/ %
	Van	Tram	Car	Pedestrian	Truck	
YOLO v5	96.5	96.9	97.7	87.1	97.3	95.1
YOLO v5+ 四层检测层	97.6	96.4	97.7	87.3	98.7	95.6
YOLO v5+ ConvFocus	97.3	97.8	97.9	88.3	98.8	96.0
YOLO v5+ 四层检测层+ 双线性插值上 采样+CBAM	97.0	97.9	97.8	87.5	98.7	95.8
改进的 YOLO v5	97.5	98.5	98.0	89.3	99.2	96.5

3.5 与其他目标及小目标检测算法的对比

为进一步对比实验,将本文算法与目标及小目标检测算法进行比较,其中包括一阶段目标检测算法 SSD、YOLO v2、YOLO v3,二阶段目标检测算法 Fast R-CNN、Faster R-CNN,融合一阶段和二阶段优点的 RefineDet 算法以及同样针对道路小目标检测的 Vehicle-YOLO、Refined YOLO v4 算法。参考所选算法原作者提出的最优参数及数据集特点,将所选算法参数设为最优。将上述算法在经过文中方法处理的 KITTI 数据集上进行训练与测试,并选择平均精度均值 mAP 和检测速度两项指标评价算法,表 3 展示了不同算法在 KITTI 数据集目标检测上的实验结果。

表 3 改进后的算法与其他目标及小目标检测算法的对比

算法	平均精度均值/mAP%	检测速度/FPS
Fast R-CNN	63.69	10.29
Faster R-CNN	74.6	12.8
SSD	83.91	50.18
YOLO v2	67.0	90.36
YOLO v3	90.07	42.17
RefineDet	89.3	41.5
Vehicle-YOLO	94.2	41.24
Refined YOLO v4	94.53	52.1
YOLO v5	95.1	77.3
改进的 YOLO v5	96.5	74.63

由表 3 可以看出,改进的 YOLO v5 算法平均精度均值达到了 96.5%,较表中一阶段检测、二阶段检测以及融合一二阶段优点的 RefineDet 算法有很大提升,较两个道路小目标检测算法 Vehicle-YOLO、

Refined YOLO v4 也分别有 2.3%和 1.97%的提升。对比表中算法,本文算法在有较好检测速度的同时实现了平均精度均值更高,目标检测更准确。

4 结论

本文针对小目标像素少、可提取特征少,易出现小目标漏检、误检,拉低检测精度的问题,以 YOLO v5 网络模型为基础进行改进。首先使用双线性插值法进行上采样操作,在采样过程中降低图像的失真率,保证检测过程中目标的清晰度,其次设计了一个 ConvFocus 模块,替换原始算法的 Focus 模块,减少直接对输入图像进行下采样导致的小目标特征损失严重以及漏检问题,提升小目标的检测能力;然后增加了一个小目标检测层,该层包含更多的小目标位置信息,利于小目标的定位与识别;最后插入 CBAM 注意力机制模块,通过注意力机制模块从通道和空间两个维度聚焦图像的重要特征,降低目标误检率。将改进的 YOLO v5 算法在本文处理的 KITTI 数据集上与原始 YOLO v5 算法做对比实验。在检测精度方面,改进的 YOLO v5 算法在小目标检测上平均精度更高;在检测速度方面,本文算法由于设计并添加提升小目标检测的模块,检测速度较原始算法稍有下降,由于 YOLO v5 算法本身的速度优势,该下降可忽略不计。同时,将改进的 YOLO v5 算法与 Faster R-CNN、SSD、RefineDet、Vehicle-YOLO 等算法在本文数据集上进行对比实验发现,改进的 YOLO v5 算法在具有更高检测速度的同时也具有更高的检测精度。

参考文献:

- [1] 吴伟浩,李青. 基于改进 YOLO v3 的电连接器缺陷检测[J]. 传感技术学报,2020,33(2):299-307.
- [2] 李运堂,谢梦鸣,王鹏峰,等. 基于改进 YOLO v3 算法的斜拉桥拉索表面缺陷检测方法[J]. 传感技术学报,2021,34(11):1509-1517.
- [3] 葛青青,张智杰,袁珑,等. 融合环境特征与改进 YOLO v4 的安全帽佩戴检测[J]. 中国图象图形学报,2021,26(12):2904-2917.
- [4] 赵朵朵,章坚武,傅剑峰. 基于深度学习的实时人流统计方法研究[J]. 传感技术学报,2020,33(8):1161-1168.
- [5] Shan Y, Lu W, Chew C. Pixel and Feature Level Based Domain Adaptation for Object Detection in Autonomous Driving[J]. Neurocomputing,2019,367:31-38.
- [6] 赵文清,孔子旭,周震东,等. 增强小目标特征的航空遥感目标检测[J]. 中国图象图形学报,2021,26(3):

- 644-653.
- [7] 罗会兰,陈鸿坤. 基于深度学习的目标检测研究综述[J]. 电子学报,2020,48(6):1230-1239.
- [8] 赵永强,饶元,董世鹏. 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.
- [9] 刘颖,刘红燕,范九伦,等. 基于深度学习的小目标检测研究与应用综述[J]. 电子学报,2020,48(3):590-601.
- [10] Girshick R, Donahue J, Darrell T, et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation [C]//2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, IEEE, 2014:580-587.
- [11] Girshick R. Fast R-CNN [C]//2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, IEEE, 2015:1440-1448.
- [12] 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.
- [13] Liu W, Anguelov D, Erhan D, et al. SSD: Single Shot Multibox Detector [C]//14th European Conference on Computer Vision, Amsterdam, The Netherlands, Springer, 2016:21-37.
- [14] Redmon J, Divvala S, Girshick R, et al. You Only Look Once: Unified, Real-Time Object Detection [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, IEEE, 2016:779-788.
- [15] Redmon J, Farhadi A. YOLO9000: better, faster, stronger [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, IEEE, 2017:6517-6525.
- [16] Redmon J, Farhad I A. YOLOv3: an Incremental Improvement [EB/OL]. (2018-04-08) [2022-01-05]. <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]. (2020-04-22) [2022-01-05]. <https://arxiv.org/pdf/2004.10934.pdf>.
- [18] Petsiuk V, Jain R, Manjunatha V, et al. Black-Box Explanation of Object Detectors via Saliency Maps [C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, IEEE, 2021:11438-11447.
- [19] Griffin B A, Corso J J. Depth from Camera Motion and Object Detection [C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, IEEE, 2021:1397-1406.
- [20] He Y H, Zhu C C, Wang J R, et al. Bounding Box Regression with Uncertainty for Accurate Object Detection [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, IEEE, 2019:2883-2892.
- [21] Zhang S F, Wen L Y, Bian X, et al. Single-Shot Refinement Neural Network for Object Detection [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, IEEE, 2018:4203-4212.
- [22] Rezaatofghi H, Tsoi N, Gwak J Y, et al. Generalized Intersection over Union: A Metric and A Loss for Bounding Box Regression [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, IEEE, 2019:658-666.
- [23] Joseph K, Khan S, Khan F, et al. Towards Open World Object Detection [C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, IEEE, 2021:5826-5836.
- [24] 顾恭,徐旭东. 改进YOLOv3的车辆实时检测与信息识别技术[J]. 计算机工程与应用,2020,56(22):173-184.
- [25] Mahto P, Garg P, Seth P, et al. Refining YOLOv4 for Vehicle Detection [J]. International Journal of Advanced Research in Engineering and Technology, 2020, 11(5):409-419.
- [26] 刘洪江,王懋,刘丽华. 基于深度学习的小目标检测综述[J]. 计算机工程与科学,2021,43(8):1429-1442.
- [27] Liu L, Ouyang W L, Wang X G, et al. Deep Learning for Generic Object Detection: A Survey [J]. International Journal of Computer Vision, 2020, 128(2):261-318.
- [28] Sanghyun W, Jongchan P, Joon-Young L, et al. CBAM: Convolutional Block Attention Module [C]//15th European Conference on Computer Vision, Munich, Germany, Springer, 2018:3-19.
- [29] Geiger A, Lenz P, Stiller C, et al. Vision Meets Robotics: the KITTI Dataset [J]. The International Journal of Robotics Research, 2013, 32(11):1231-1237.
- [30] He K M, Zhang X Y, Ren S Q, et al. Deep Residual Learning for Image Recognition [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA: IEEE, 2016:770-778.
- [31] Hu J, Shen L, Sun G. Squeeze-and-Excitation Networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 42(8):2011-2023.
- [32] Lin T Y, Maire M, Belongie S, et al. Microsoft COCO: Common Objects in Context [C]//13th European Conference on Computer Vision, Zurich, Switzerland, Springer, 2014:740-755.

- [33] Braun M, Krebs S, Flohr F, et al. The EuroCity Persons Dataset: a Novel Benchmark for Object Detection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 41(8): 1844–1861.
- [34] Zhang S S, Benenson R, Schiele B. CityPersons: A Diverse Dataset for Pedestrian Detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, IEEE, 2017: 3213–3221.



牛为华(1978—),女,天津人,博士,研究生导师,研究方向为数字图像处理及其在电力系统中的应用, niuw@ncepu.edu.cn;



殷苗苗(1998—),女,河南安阳人,硕士研究生,研究方向为图像处理及应用, mmyin2021@163.com。