

DOI:10.19651/j.cnki.emt.2314587

改进 YOLOv5s 的复杂交通场景下目标检测算法

卫策 吕进 曲晨阳

(长安大学信息工程学院 西安 710018)

摘要: 针对在实际的交通道路目标检测中,存在着小目标检测精度低,遮挡目标容易出现漏检误检等问题,提出了一种改进的 YOLOv5s 道路目标检测算法 YOLOv5s-OEAG。将 YOLOv5s 的标签分配策略更换为效率更高的 OTA 标签分配策略,提高模型的检测精度与泛化能力;提出了一种轻量化的解耦预测头对不同尺寸的特征层进行分类任务与回归任务的解耦,提高模型对道路中小目标的检测能力;将原始模型中的最近邻插值上采样模块替换为轻量级通用上采样 CARAFE 模块,有助于更好地保留图像中的细节信息,提高模型的精度;提出了一种新的 C3 模块 GMC3,在减小模型计算量的同时提高模型捕获特征的能力;为了提高模型的泛化能力,对 KITTI 数据集进行了扩充,增加了小目标的数量。实验结果表明,改进后的模型在经过扩充后的 KITTI 数据集的 mAP 达到了 90.4%,比原始模型的精度提高了 2.8%;FPS 为 75,满足实时性的要求,在一定程度上提高了对复杂交通场景的适应能力。

关键词: 目标检测;YOLOv5s;深度学习;轻量化解耦头;CARAFE 上采样

中图分类号: TP391.4 **文献标识码:** A **国家标准学科分类代码:** 510.4

Improved object detection algorithm for complex traffic scenes in YOLOv5s

Wei Ce Lyu Jin Qu Chenyang

(School of Information Engineering, Chang'an University, Xi'an 710018, China)

Abstract: In response to challenges in practical road target detection, such as low accuracy in detecting small targets and the occurrence of missed and false detections for occluded targets, an improved YOLOv5s road target detection algorithm, termed YOLOv5s-OEAG, is proposed in this study. The label assignment strategy of YOLOv5s is replaced with a more efficient OTA label assignment strategy to enhance the model's detection accuracy and generalization ability. Additionally, a lightweight decoupled prediction head is introduced to decouple classification and regression tasks for different-sized feature layers, thereby improving the model's capability to detect small targets on roads. The original nearest-neighbor interpolation upsampling module is replaced with the lightweight and versatile CARAFE module to better preserve fine details in the image, thereby enhancing the model's accuracy. Furthermore, a novel C3 module, GMC3, is proposed to reduce model computational complexity while improving the model's feature capturing capability. To enhance the model's generalization ability, the KITTI dataset is augmented, increasing the number of small targets. Experimental results demonstrate that the improved model achieves a mAP of 90.4% on the augmented KITTI dataset, representing a 2.8% improvement over the original model's accuracy. With a frame per second (FPS) rate of 75, meeting real-time requirements, the model exhibits enhanced adaptability to complex traffic scenarios.

Keywords: object detection;YOLOv5s;deep learning;lightweight decoupled head;CARAFE upsampling

0 引言

随着科学的进步与社会的发展,传统的交通领域不断的与人工智能技术相融合产生了一系列新的领域,例如:智能交通管理、自动驾驶技术、交通预测和优化、智能交通安全等。其中基于人工智能的自动驾驶技术已经逐步商用

化,为未来的交通运输带来了革命性的变化。目标检测是自动驾驶的重要组成部分,标检测可以实时识别和跟踪交通环境中的车辆、行人、交通标识等,提高自动驾驶车辆的安全性,避免交通事故的发生。但是在实际的行驶环境中道路场景复杂多样,需要更加可靠准确的目标检测算法。

最近几年来,卷积神经网络在图像分类和物体检测领

域有着显著的成就。当前较流行且基于深度学习的目标检测算法,可以分为一阶目标检测算法和二阶目标检测算法。一阶段算法直接检测目标的位置和类别,一般采用滑动窗口的方式进行扫描,然后使用分类器对每个窗口进行分类和识别。一些当前流行的一阶目标检测算法包括:YOLO 系列^[1-5]、RetinaNet^[6]、SSD^[7]等。二阶段算法首先检测出图像中可能存在的候选框,然后对每个候选框进行分类和回归处理,得到目标的位置和类别。一些当前流行的二阶目标检测算法包括:Mask R-CNN^[8]、R-CNN^[9]、Fast R-CNN^[10]、Faster R-CNN^[11]等。然而,在实际的交通道路下,往往存在一些非常小的目标,如行人、自行车等,这些目标很难被目标检测算法正确地识别和定位。并且车辆、行人等目标往往会出现形变,如变形、遮挡等,这可能会导致目标检测算法的准确率下降。

为了能够使目标检测算法更好地用在实际的交通场景里,道路目标检测算法主要从特征提取器的改进、采用注意力机制、增强模型的感受野、模型结构优化等方面进行探索。曾耀等^[12]通过对 YOLOv5s 的网络结构进行替换,更换为加权的双向特征金字塔,并使用了有坐标信息的注意力机制模块能够提高对小目标的检测能力。王鹏等^[13]使用自底向上的金字塔降采样网络结构,增强特征融合,并采用改进的 CIOU 目标回归损失函数,使模型的学些能力和目标检测精度得到提升。盛博莹等^[14]使用了基于反馈机制的特征提取网络(RFP-PAN),且提出级联注意力机制(SECA),充分融合了不同尺度的特征信息,提高了检测精度。张上等^[15],通过将 CBAM 和 CA 相融合来提高模型的检测精度,再通过 FPGM 剪枝对模型进行压缩。肖焱俊等^[16],通过对数据集进行水平翻转、随机添加像素等操作进行数据增强,并再 YOLOv5s 的主干网络添加 GAM 注意力模型,从而提高模型的检测精度。

上述算法在复杂的交通道路进行应用时还是会存在检测精度较低、道路目标漏检、模型结构复杂不利于实际应用部署等挑战。本文在 YOLOv5s 目标检测模型的基础上,对其进行改进,提出了 YOLOv5s-OEAG 算法。其主要改进方面如下:

1) 引入 OTA(optimal transport allocation)^[17]方法。该方法能够灵活地调整正负样本的分配策略,使得模型可以更好地适应各种目标特征和场景,并提高检测精度,而且考虑了目标之间的相似性,避免了多个目标之间对同一个正样本的竞争,减少了样本之间的冲突,提高了模型的稳定性和泛化能力。

2) 提出了一种轻量化的解耦预测头 Efficient Head,能够使模型的预测头分别独立的进行特征图的分类任务与回归任务,让模型能够更好的捕捉目标的位置信息和类别信息,提高了检测性能,特别是在处理小目标或密集目标时。

3) 使用轻量级上采样算子 CARAFE^[18]来替换原始模型中的最近邻插值上采样,使模型能够在上采样的过程中更好

的保留细节信息,提高模型对图像中微小特征的感知能力。

4) 本文受到分组卷积的启发,设计了一种新模块 GMC3。并使用该模块替换原始模型中的 C3 模块,在降低模型参数量和计算量的同时提高精度。

1 YOLOv5s 算法整体结构

在 YOLOv5s 的 v6.0 版本,其网络结构还是由输入端、Backbone、Neck 和 Head 这 4 个部分组成,如图 1 所示。在输入端 YOLOv5s 可以接受任意大小的图片作为输入,采用 Mosaic 数据增强、自适应锚框计算等方法对图像进行预处理。Backbone 主要由 Conv 模块、C3 模块和 SPPF 模块组成。Neck 部分主要使用的是路径聚合网络(path aggregation network,PAN)^[19]与特征金字塔网络(feature pyramid network,FPN)^[20]相结合的结构,以捕获不同尺度的特征图并在不同的层次上进行特征融合。Head 部分是由多个卷积层和池化层组成的,用于对 Neck 中提取的特征图进行处理以生成检测结果,通过非极大值抑制(non-maximum suppression,NMS)^[21]来对目标框进行筛选,以提高检测结果的准确性。

2 YOLOv5s 的改进

2.1 OTA 标签分配策略

YOLOv5s 的标签分配策略是通过定位 GT 的中心点来定位其所在特征图的网格单元(Grid Cell),并通过计算把距离该中心点最近的周围两个网格单元也当作正样本,其目的就是为了增加正样本的数量如图 2 所示。

虽然 YOLOv5s 在对 one-stage 目标检测算法中的区分正负样本做了贡献,但其对于每个 GT 分配锚框还是一种静态的策略,没有考虑到其存在着各种大小、形状或遮挡的情况,可能会导致同一个锚框分配给了多个 GT,如图 3 所示。基于此,本文引入 OTA,它能够当前训练批次里边的所有 GT 与所有的正样本锚框进行动态的重新分配,为每一个 GT 分配到其最合适的正样本锚框。其原理是将标签分配问题视作最优传输问题(optimal transport,OT)最优传输问题,在真实标签和所有正样本之间计算运输成本,通过寻找一个合适的映射关系,使得运输成本最低。

OT 问题就是在,拥有 s_i 单位的商品的供应商 i 与需要 d_j 单位的商品的需求者 j 之间找到一个最小运输成本的方案 $\pi = \{\pi_{i,j} \mid i = 1, 2, 3, \dots, m, j = 1, 2, 3, \dots, n\}$,其运输成本用 c_{ij} 表示,如式(1)。将 OT 应用到目标检测的标签分配策略中,也就是将 GT 当作正样本的供货商,每个 GT 拥有 k 个正样本, k 的确定会在下文中解释,每一个锚框作为一个需求商需要一个正样本。建立一个维度为 $GT_{all} \times anchor_{all}$ 的代价矩阵如图 4 所示,接下来的标签分配都是基于此代价矩阵的。计算每个 GT 与每一个锚框之间的运输成本并填入到代价矩阵中所对应的位置中,如式(2)所示:

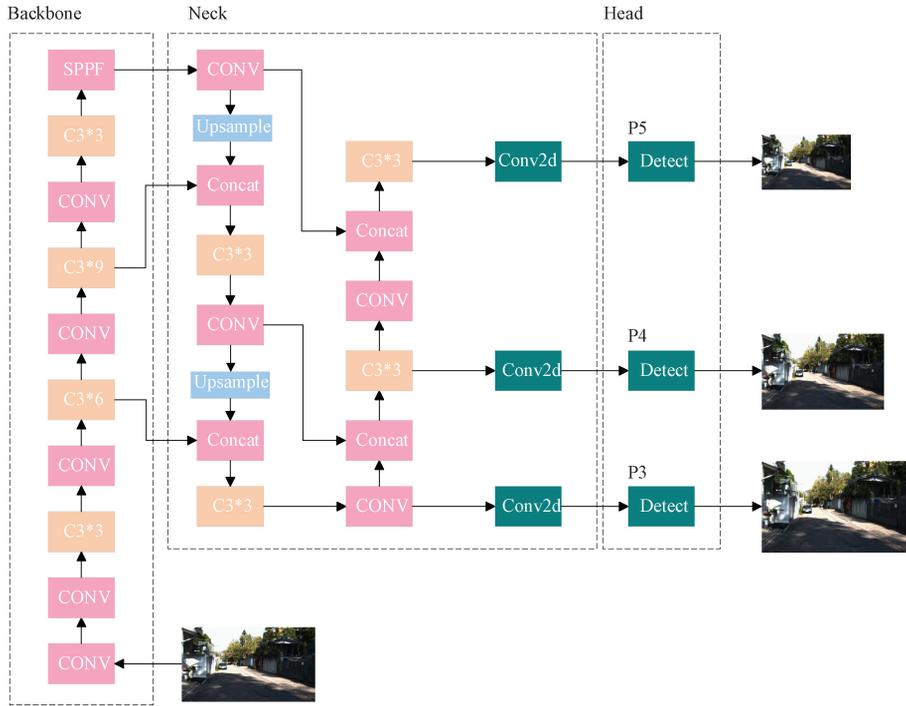


图 1 YOLOv5s 结构

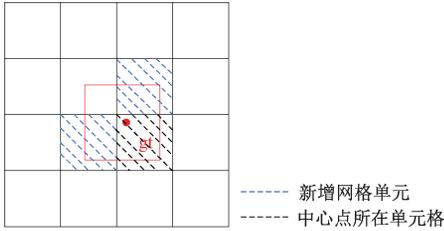


图 2 扩展网格单元

	锚框1	锚框2	锚框3	锚框4	...	锚框 $n-1$	锚框 n
目标框1	$C_{1,1}$	$C_{1,2}$	$C_{1,3}$	$C_{1,4}$	$C_{1,n-1}$	$C_{1,n}$
目标框2	$C_{2,1}$	$C_{2,2}$	$C_{2,3}$	$C_{2,4}$	$C_{2,n-1}$	$C_{2,n}$
目标框3	$C_{3,1}$	$C_{3,2}$	$C_{3,3}$	$C_{3,4}$	$C_{3,n-1}$	$C_{3,n}$
					:		
目标框 n	$C_{n,1}$	$C_{n,2}$	$C_{n,3}$	$C_{n,4}$	$C_{n,n-1}$	$C_{n,n}$

图 4 真实标签与锚框的代价矩阵

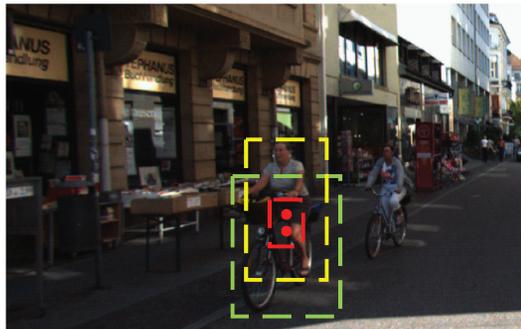


图 3 目标检测中模糊的锚点

$$\min \sum_{i=1}^m \sum_{j=1}^n c_{ij} \pi_{ij} \quad (1)$$

$$c_{ij} = L_{cls}(P_j^{cls}, G_i^{cls}) + \alpha L_{reg}(P_j^{box}, G_i^{box}) \quad (2)$$

其中, P_j^{cls} 和 P_j^{box} 分别表示 $anchor_j$ 的预测类别得分和边界框, G_i^{cls} 和 G_i^{box} 代表 GT_i 的真实类别和边界框, L_{cls} 和 L_{reg} 分别表示交叉熵损失和交并比损失, α 是两个损失的

平衡系数。

得到了代价矩阵,接下来需要确定参数 k 。由于每个 GT 的大小、尺度和遮挡条件是不同的,所以其所需的正样本 k 的数量也应该是不同的。先计算预测边界和真实框之间的交并比,然后选择前 q 个预测结果得到的就是正样本 k 的数量。根据所求出的代价矩阵和动态参数 k 得到最终用来选择每个 GT 所对应的锚框的匹配矩阵,将所选中的锚框置 1,其余置 0,如图 5 所示。

目标框1	0	1	0	0	0	1
目标框2	0	0	1	1	0	0
目标框3	1	0	0	1	1	0

图 5 锚框匹配矩阵

对于图 5 中所出现的同一个锚框被多个目标框同时选中的问题,其解决方法是对这两个位置的 c_{ij} 值进行比较,选择较小的值,并将其他值为 1 的位置都置为 0。最终得到每一个 GT 所对应的代价较小的一系列锚框,这些锚

框将会被用在训练模型过程中的损失计算当中。

2.2 轻量化的解耦预测头

YOLOv5 中的头部(head)部分是一种耦合的预测头,通过将分类和回归分支融合在一起,实现参数共享,从而从特征图中提取并预测目标的位置和类别信息。然而,在实际的目标检测中,分类任务和回归任务在模型训练过程中关注的重点不同。分类任务更关注特征图中的纹理内容,而回归任务更注重特征图的边缘信息。因此,本文将对 YOLOX 中所提出的解耦预测头进行轻量化的改进,并将其替换掉 YOLOv5s 中原始的耦合预测头。

YOLOX 中的解耦头将分类任务和回归任务分离开来,使模型能够分别专注于这两个任务。这种分离可以提高模型对目标位置和类别的准确性,尤其在存在小目标或密集目标的情况下效果显著。然而,YOLOX 的解耦头相对复杂,增加了模型的计算复杂度,这对于模型整体的轻量化是不利的。

因此,本文对 YOLOX 的解耦头进行了优化,在不损失太多检测精度和速度的前提下减少复杂度。首先,通过 1×1 卷积操作对 YOLOv5s 不同特征层输出的特征图进行了处理。 1×1 卷积可以有效地降低特征图的通道数,从而减少后续卷积层的计算负担,同时保持信息丰富性。接下来,将特征图输入到一个 3×3 的轻量化卷积层 GSConv 中,在降低计算量的同时,能够从特征图中提取局部特征;最后,回归任务和分类任务分开处理,在进行回归任务之前,再经过一个 3×3 的轻量化卷积层 GSConv。Efficient Head 的结构如图 6 所示。而且经过下文实验发现,这些改进使得模型在减少计算成本的同时,能够保持良好的检测精度和速度。

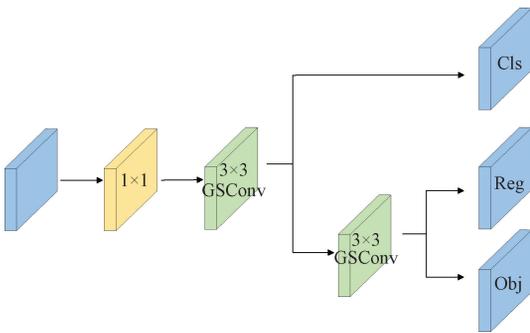


图 6 解耦预测头

GSConv^[22]是一种新型的卷积操作,通过引入 Shuffle 操作,将通道密集卷积(SC)生成的信息与深度可分离卷积(depth-wise separable convolution, DSC)生成的信息混合,以减少计算成本并提高模型性能。这种操作使得 GSConv 能够弥补 DSC 在进行分离操作时所造成的信息丢失。

2.3 引入上采样算子 CARAFE

YOLOv5s 在网络的不同层之间对提取到的特征进行

放大的使用的是最近邻插值上采样,其作用是将特征图的尺寸增大。最近邻插值是一种最简单的插值方式,其原理是将原始图片中的像素进行映射放大,从而放大特征图,而不引入新的信息。所以,最近邻插值会使得特征图在上采样时会丢失图像的一些信息和细节,而且在面对复杂像素时,效果较差。针对该问题,本文提出使用轻量级上采样算子 CARAFE 来替换最近邻插值。CARAFE 主要由两部分组成,CARAFE 主要由上采样核预测模块和特征重组模块两部分组成。

上采样核预测模块的主要作用是生成用于上采样的核,这些核能够将低分辨率特征图上采样到目标分辨率。其流程为,先使用 1×1 的卷积操作将输入的形状为 $H \times W \times C$ 的特征图通道数进行压缩,使其形状变为 $H \times W \times C_m$,这样做能够减少后续的计算成本;然后,使用一个大小为 $K_{encode} \times K_{encode} \times \sigma^2 K_{up}^2$ 的卷积层来对压缩后的特征图进行上采样卷积核的预测,将特征图的形状变为 $H \times W \times \sigma^2 K_{up}^2$;接下来,将特征图通道维度重新组织,使其在空间维度上连续排列,得到的上采样核形状为 $\sigma H \times \sigma W \times K_{up}^2$;最后使用 softmax 函数对得到的上采样内核进行归一化,并且从特征重组模块中取出以该特征点为中心的大小为 $K_{up} \times K_{up}$ 的区域,来和该点的上采样核作点积运算,得到该点的上采样特征。整个流程如图 7 所示。

总之,CARAFE 上采样能够使用实时生成的自适应的上采样核在更大的感受野内聚合上下文信息,能够有助于更好地保留图像中的细节信息和处理不同尺度的特征图,使得模型对微小特征的感知更为准确。

2.4 GMC3 模块

分组卷积是深度学习中的一种卷积操作,其核心思想是将输入通道分成多个组,每个组内的通道进行独立卷积操作,然后将它们的输出进行通道融合,以降低模型参数数量和计算复杂度。这一操作方式的优势包括减少模型参数、提高计算效率、处理多尺度特征以及增强模型表现,因此在卷积神经网络中得到广泛应用。例如,在分组卷积的思想进一步发展得来的深度可分离卷积(depthwise separable Conv, DSC)。

$$FLOPs_n = H \times W \times C_1 \times K \times K \times C_2 \quad (3)$$

$$FLOPs_g = H \times W \times \frac{C_1}{G} \times K \times K \times C_2 \quad (4)$$

$$Params_n = (K \times K \times C_1 \times C_2) \quad (5)$$

$$Params_g = (K \times K \times \frac{C_1}{G} \times C_2) \quad (6)$$

上述式(3)、(5)计算的是普通卷积的计算量和参数量,式(4)、(6)计算的是分组卷积的计算量和参数量。 $H \times W \times C_1$ 表示的是输入特征图的大小和通道数, $K \times K$ 表示的是使用的卷积核大小,输出通道数为 C_2 ,分组卷积将输入通道分成 G 组,每组有 C_1/G 个通道,输出通道数也为 C_2 。他们的比值计算公式如所示。

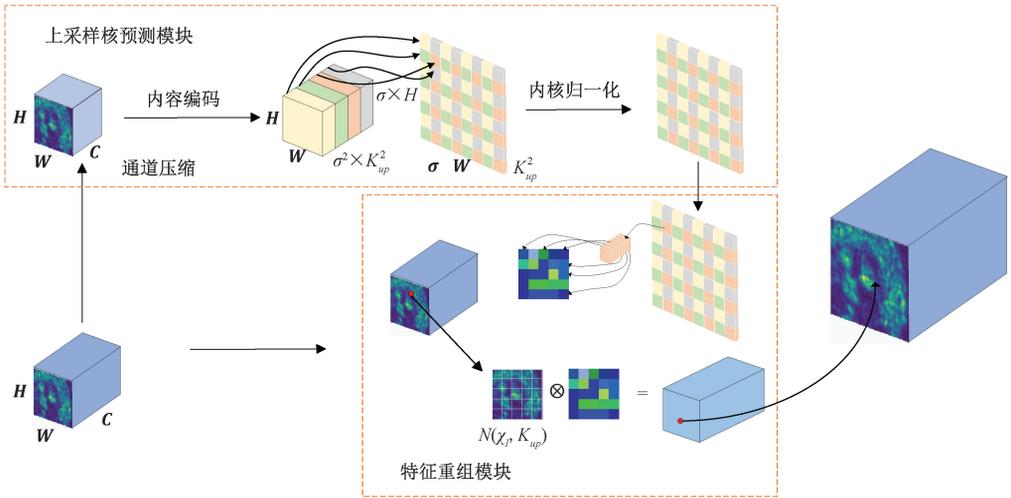


图 7 CARAFE 结构图

$$\frac{FLOPs_n}{FLOPs_g} = \frac{H \times W \times C_1 \times K \times K \times C_2}{H \times W \times \frac{C_1}{G} \times K \times K \times C_2} = G \quad (7)$$

$$\frac{Params_n}{Params_g} = \frac{K \times K \times C_1 \times C_2}{K \times K \times \frac{C_1}{G} \times C_2} = G \quad (8)$$

从上述的化简公式能够看出普通卷积的计算量是分组卷积的 G 倍,普通卷积的参数量是分组卷积的 G 倍。

基于该思想本文设计了一个新的 Conv 模块 GMConv,并使用其替换模型中 C3 模块的普通卷积,得到 GMC3 模块。将输入的特征图的通道数进行分组,一半进行分组卷积操作,另一半什么也不做。进行分组卷积的一半每一组可以通过不同大小的卷积核。另一半什么也不做的原因是因为 Han 等^[23]已经在 GhostNet 这篇文章中说明了,经过神经网络的特征提取后,在得到的许多特征图中有一些特征图的相似性很高,这就是所谓的冗余特征图。基于此,该模块只对一般的通道进行分组卷积操作,这样能够减小模型的计算量和复杂度。当另一半的分组卷积完成后,将他们重新在通道维度上连接在一起并通过一个 Channel Shuffle 操作,打乱原始的通道顺序,引入更多的通道交叉信息,从而增强了模型的代表能力。这有助于网络更好地捕获图像或特征之间的关联性,提高模型对复杂特征的理解能力,其结构如图 8 所示。

3 实验结果与分析

3.1 参数设置与实验环境

实验中为了能够公平对比,使用 YOLOv5s 本身所使用的参数设置,使用随机梯度下降(stochastic gradient descent,SGD)优化器,循环学习率为 0.1,初始学习率为 0.01,SGD 优化器的动量系数取值为 0.937,权重衰减系数取值为 0.0005,训练的批量大小为 32,训练轮次为 300 个 epoch,输入图像大小为 640×640 。实验环境如表 1 所示。

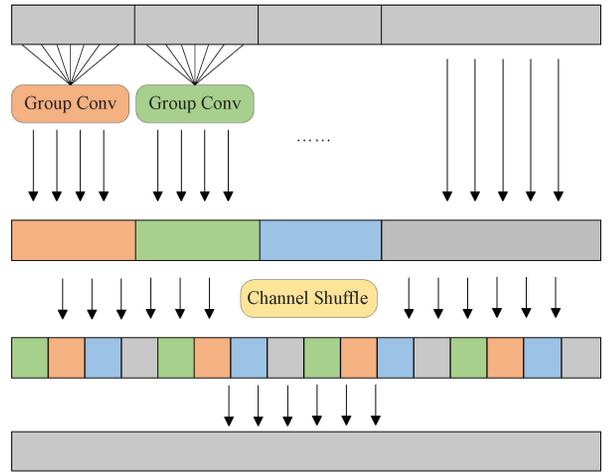


图 8 GMConv 结构图

表 1 实验环境

参数	实验环境
CPU	AMD Ryzen 7 6800H 3.20 GHz
GPU	GeForce RTX 3060
操作系统	Windows11
深度学习框架	Pytorch1.12
Python	3.8

3.2 实验数据集

为了对改进的模型——验证其有效性,本文选取 KITTI^[24]数据集进行实验探究。

KITTI 数据集是一个广泛用于自动驾驶场景下的计算机视觉数据集。该数据集中有 9 类标签,为了能够更好地适应实际情况,本文对数据集中的类别进行重新划分,将 Pedestrian、Person_sitting 类合并为 Person 类, Van、Truck、Car 合并为 Car 类,Cyclist 类保留不变,删除 Misc 类、DontCare 类和 Tram 类。KITTI 数据集中总共有

14 999 张图像,其中有标注的图像有 7 481 张,为了能体现本文所设计的模型的有效性,从未标注的数据图像中挑选了 1 000 张左右进行手工标记,在标记过程中主要是标记数据标签中的 Cyclist 类与 Person 类,因为他们在数据集中所占的占比比较少。目前对于小目标的定义有两种:1)在 MS COCO 数据集中规定,像素面积小于 32×32 的物体被规定为小目标,如表 2 所示;2)国际光学工程学会定义小目标的尺寸为小于原图像的 $0.12^{[25]}$;本文使用国际光学工程学会对于小目标的定义,基于此在标注中也增加了小目标的数量。最终 Car 类、Cyclist 类、Person 类的标签数量分别为 34 451、1 819、5 440。为了能够更加直观的了解数据集中大中小目标的分布,本文通过计算每张图像中的标签的宽高与图像的宽高比,并画出散点图,如图 9 所示。从图 9 中的分布情况可以看出,KITTI 数据集中的目标大小都集中在中、小尺寸。

表 2 小目标定义

类型	规定
small	$area < 32^2$
medium	$32^2 < area < 96^2$
large	$96^2 < area$

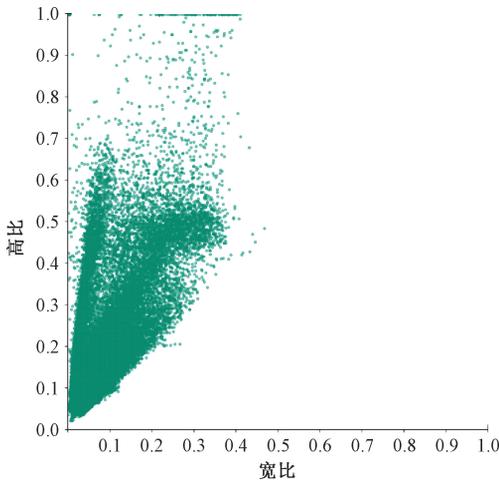


图 9 KITTI 数据集标签宽高比

3.3 评价指标

对于本文所提出的模型,为了能够更好的评价该模型的性能表现,本文采用的评价指标为召回率(recall,R)、精确度(precision,P)、平均精度均值(mean average precision,mAP)和每秒帧数(FPS)。mAP 的计算基于 Precision-Recall 曲线,即在不同的置信度阈值下计算模型的 P 和 R。例如常用的目标检测模型指标 $mAP@0.5$ 和 $mAP@0.5:0.95$, $mAP@0.5$ 代表当 IoU 阈值为 0.5 时的平均精度,而 $mAP@0.5:0.95$ 则是在 IoU 阈值从 0.5 到 0.95 之间以 0.05 为步长计算出来的平均精度的平均值,更全面地评价了模型在不同 IoU 阈值下的表现。相关公式如下:

$$P = \frac{TP}{TP + FP} \tag{9}$$

$$R = \frac{TP}{TP + FN} \tag{10}$$

$$AP = \int_0^1 P dR \tag{11}$$

$$mAP = \frac{\sum_{i=1}^n AP_i}{n} \tag{12}$$

在上述公式中,TP 表示模型预测为正样本且与真实标签匹配的样本数目,FP 表示模型预测为正样本但与真实标签不匹配的样本数目,FN 表示模型预测为负样本但与真实标签匹配的样本数目。式(11)表示每个类别的 Average Precision (AP)是该类别 Precision-Recall 曲线下的面积。式(12)表示将每个类别的 AP 求和取平均,得到的就是所有类别的 mAP。

3.4 实验分析

1) 改进的标签分配策略的有效性验证

通过使用 OTA 的标签分配策略对每个 GT 的锚框进行筛选后,将所得到的锚框与真实框进行损失计算。这些损失包括置信度损失(obj_loss)、分类损失(cls_loss)和边界框回归损失(box_loss)。这 3 种损失的计算都与正样本锚框有联系,因此使用经过优化的锚框与真实目标进行计算可以得到更小的损失值。为了验证该结论,只对 YOLOv5s 的目标分配策略进行改进,训练完成后与原始的 YOLOv5s 进行对比,结果如图 10 所示。

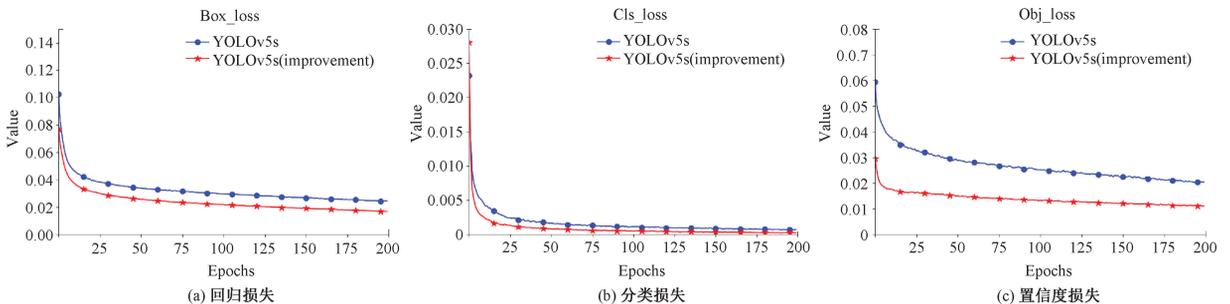


图 10 3 种损失值对比

从图中可以看到优化过的模型的损失明显低于原模型,说明模型更好地学习了训练数据集中的特征,使得预测到的结果与真实结果之间的差距变小,模型也具有更好的拟合能力。

2) 解耦头轻量化实验

表3 解耦头对比

实验模型	模型大小/M	参数量/ 10^6	GFLOPs	mAP@0.5/%	mAP@0.5:0.95/%
YOLOv5s	13.6	7.0	15.8	87.6	57.3
YOLOv5s-E1	19.6	10.1	21.5	89.5	61.3
YOLOv5s-E2	38.0	19.7	39.1	89.8	61.8

从表中可以看出,经过本文改进后的轻量化解耦头YOLOv5s-E1在mAP@0.5和mAP@0.5:0.95上相对于原始模型分别提升了1.9%和4.0%,说明了其对检测精度的有效性;虽然其精度低于原始解耦头0.3%、0.5%,但其在模型大小、参数量、GFLOPs上分别减少了48%、48%、45%,有效的降低了模型的大小和计算量,提升了训练的效率。

3) CARAFE算子有效性验证

为了验证CARAFE上采样是否能够更好地融合不同尺度的特征图,从而提升模型对于上下文的特征提取能力,提升检测精度。将使用CARAFE上采样算子的模型与原始模型进行对比实验,结果如表4所示。

表4 CARAFE算子验证实验

实验模型	模型	mAP@0.5/%	mAP@.5:.95
	大小/M		
YOLOv5s	13.6	87.6	57.3
YOLOv5s-R	13.9	89.7	61.5

从表中可以看出,YOLOv5s-R相对于原模型在mAP@0.5和mAP@0.5:0.95上分别有2.1%、4.2%的提升,且模型大小只增加了0.2M,说明了该模块对模型的

为了验证对YOLOX中所提出的解耦头轻量化改进是否能够在保持良好精度的同时降低模型的大小、参数量和计算量,本文将在原始YOLOv5s的基础上只更换解耦预测头,并且对改进后的解耦头与原始解耦头进行对比实验,其结果如表3所示。

有效性。

4) GMC3模块有效性验证

为了能够验证本文所设计卷积模块GMConv是否相对于YOLOv5中的Conv有更少的计算量和参数量,将对这个两个卷积模型进行单独的对比实验。将输入的batch_size设置为1,输入图像的宽高设置为 32×32 ,输入通道设置为64,输出通道设置为64,训练次数设置为1000次,结果如表5所示。从表中可以看出输入同样的参数,本文所设计的模块的FPS相对于Conv有所下降,但是计算量和参数量分别相对于Conv降低了64%、66%,证明了模型的有效性。

表5 GMConv模块对比实验

实验模型	FPS	FLOPs/M	参数量/K
Conv	3 007	76.022	36.992
CMConv	1 040	27.001	12.992

接下来将GMConv替换C3模块中的普通卷积,得到新模块GMC3,并与原始模型进行对比实验。结果如表6所示,其参数量和GFLOPs相较于原模型都有所下降,但是其精度却分别提升了1.4%、3.8%,证明了本文所设计模块的有效性。

表6 GMC3模块有效性实验

实验模型	参数量/ 10^6	GFLOPs	mAP@0.5/%	mAP@.5:.95
YOLOv5s	7.0	15.8	87.6	57.3
YOLOv5s-GMC3	5.6	12.7	89.0	61.1

3.5 实验分析与对比

1) 消融实验

为了能够验证对于模型的各个部分的改进是否有效果,在KITTI数据集上对模型展开消融实验,其结果如表7所示。可以看到,模型2所引入的OTA分配策略,使得模型的mAP@0.5和mAP@0.5:0.95分别提升了0.6%、1.1%,优化了模型在训练过程中对于标签的分配,使得模型预测的结果更加准确。模型3使用了轻量化的

解耦预测头,使得模型能够捕捉到图像中的更多特征,mAP@0.5和mAP@0.5:0.95分别提升了1.9%、4.0%。模型4使用通用轻量化上采样算子CARAFE,使模型能够更好地融合特征图,提升对小目标的敏感度,mAP@0.5和mAP@0.5:0.95分别提升了2.1%、4.2%。模型5将C3模块替换为GMC3模块,不但减少了模型的计算量和参数量,还使得mAP@0.5和mAP@0.5:0.95分别提升了1.4%、3.8%。

表 7 消融实验结果

模型	OTA	Efficient head	CARAFE	GMC3	mAP@0.5/%	mAP@0.5:0.95/%	参数量/ 10^6
1					87.6	57.3	7.0
2	✓				88.2	58.4	7.0
3		✓			89.5	61.3	10.1
4			✓		89.7	61.5	7.1
5				✓	89.0	61.1	5.7
6	✓	✓	✓	✓	90.4	62.1	8.9

从实验的结果来看,每个改进都提升了模型的性能,最终得到的模型 6 即 YOLOv5s-OEAG 模型与原模型相比虽然参数量有所增加,但在 KITTI 数据集上 mAP@0.5 和 mAP@0.5:0.95 提升了 2.8%、4.8%。

2) 对比试验

为了能够说明改进模型的优越性,与当前比较主流目标检测算法在扩展后的 KITTI 数据集上进行实验对比,其中 YOLOv5-GCB、YOLOs-CPI、YOLOv5-KGA 是对引言中所提到的模型算法的复现,并将他们的模型在本文所使用的数据集上进行训练,实验结果如表 8 所示。可以看出曾耀等^[12]的改进模型虽然其参数量和计算量虽然是当前

流行算法中最低的,但是其模型精度还不如基线模型;王鹏等^[13]改进后的模型其精度要比基线模型高,但是其增加的计算量要比本文所提出的算法高,且精度没有本文所提出的算法高;肖黎俊等^[16]改进的模型其精度虽然与基线模型相比有所提升,但是其增加了过多的参数量和计算量,而且精度也没有本文所提出模型的高。经过对比实验,本文所提出的算法在精度上位列第一,证明了本文所提出模型的创新性和有效性。虽然其计算量和参数量相较于原始基线模型有一些增加,但相对于其他网络模型来说还是小很多。其检测速度 FPS 为 75,虽然不是最优,但仍然能够满足实时检测的要求。

表 8 不同算法对比结果

模型	模型大小/M	参数量/ 10^6	GFLOPs	mAP@0.5/%	mAP@0.5:0.95/%	FPS
Faster R-CNN	110.2	30.2	209.8	84.2	54.1	20
SSD	103.3	28.5	73.2	71.6	46.8	58
YOLOv4	235.7	58.1	59.6	86.7	52.6	29
YOLOX	28.6	48.7	148.3	88.2	57.4	23
YOLOv5s	13.6	7.0	15.8	87.6	57.3	116
YOLOv5s-OEAG	17.4	8.9	18.7	90.4	62.1	75
YOLOv5-GCB ^[12]	9.89	5.0	10.8	85.5	54.3	111
YOLOs-CPI ^[13]	16.6	8.4	22.9	88.8	59.7	80
YOLOv5-KGA ^[16]	27.4	14.2	25.6	87.8	58.5	94

3) 结果分析

为了能够更加清晰的展示改进后的模型性能,从 KITTI 数据集的测试集中选取部分图片进行测试,如图 11 所示。图中的左侧为 YOLOv5s-OEAG 模型所检测出来的结果,右侧为基线方法所检测出来的。图 11(a)是对图像深处小目标进行检测的结果,可以看出改进后的模型能够准确的预测到较小的行人目标,这对于在实际的交通道路中有着非常重要的作用。图 11(b)是对图像中的目标被一些其他物体遮挡的检测结果,经过改进后的模型能够很好的检测到被路灯杆遮挡的行人和没有被拍摄完全的行人。图 11(c)中图像左侧的汽车目标集中密集而且所在的地方光照不足,导致基线模型出现了目标漏检的情况,而经过改进后的模型可以准确的识别出目标的真实数量。

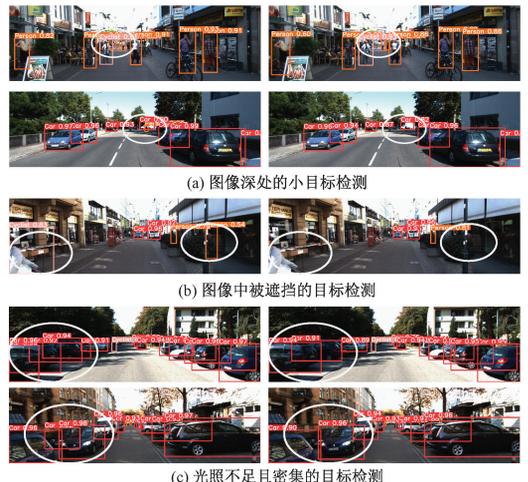


图 11 检测结果对比

4 结 论

本文在基于YOLOv5s的基础上提出了改进的YOLOv5s-OEAG目标检测算法。使用高效的OTA标签分配策略提高模型的检测精度;将网络的预测头换成进行轻量化改进的解耦预测头,提高模型预测不同尺度目标的能力;将模型的上采样算子更换为CARAFE上采样算子,增强模型对微小特征的感知能力;使用本文自研的GMC3模块不仅减少了模型的计算量与参数量,并且提升了检测精度。为了提高模型的泛化能力和鲁棒性,对KITTI数据集进行了扩充,且增加了小目标的数量。本文区别于引言中在模型中直接引入注意力模块机制的方法,而是通过设计一种新的卷积模块和引入新的上采样算子,在卷积和上采样的过程中增强模型捕获特征的能力。并且本文所引入OTA标签分配策略和轻量化解耦头与王鹏等人改进损失的方法不同,但都是为了在训练的时候让模型能够更加精确的对目标进行分配,从而降低损失值,提升模型的精度。实验表明,本文所提出的YOLOv5s-OEAG模型相较于目前最新的基于YOLOv5s改进的模型有着更高的精度,体现了本文所提出模型的先进性。在后续的研究中,希望可以在不降低模型的精确度的情况下,对模型的大小与参数量进行进一步的减小,使其能够应用到实际中。

参考文献

- [1] 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, 2016: 779-788.
- [2] REDMON J, FARHADI A. Yolov3: An incremental improvement [J]. ArXiv Preprint, 2018, ArXiv: 1804.02767.
- [3] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. Yolov4: Optimal speed and accuracy of object detection [J]. ArXiv Preprint, 2020, ArXiv: 2004.10934.
- [4] LI C, LI L, JIANG H, et al. YOLOv6: A single-stage object detection framework for industrial applications [J]. ArXiv Preprint, 2022, ArXiv: 2209.0297.
- [5] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for realtime object detectors [J]. ArXiv Preprint, 2022, ArXiv: 2207.02696.
- [6] 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.
- [7] LIU W, ANGUÉLOV D, ERHAN D, et al. SSD: Single shot multibox detector [C]. European Conference on Computer Vision. Amsterdam: Springer, Cham, 2016: 21-37.
- [8] HE K, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN [C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 2961-2969.
- [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, 2014: 580-587.
- [10] GIRSHICK R. Fast R-CNN [C]. Proceedings of the IEEE international conference on computer vision, 2015: 1440-1448.
- [11] 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.
- [12] 曾耀,高法钦.基于改进YOLOv5的电子元件表面缺陷检测算法[J].浙江大学学报(工学版),2023,57(3): 455-465.
- [13] 王鹏,王玉林,焦博文,等.基于YOLOv5的道路目标检测算法研究[J].计算机工程与应用,2023,59(1): 117-125.
- [14] 盛博莹,侯进,李嘉新,等.面向复杂交通场景的道路目标检测方法[J].计算机工程与应用,2023,59(15): 87-96.
- [15] 张上,王恒涛,冉秀康.基于YOLOv5的轻量化交通标志检测方法[J].电子测量技术,2022,45(8):129-135.
- [16] 肖黎俊,潘睿志,李超,等.基于改进YOLOv5s绝缘子缺陷检测技术研究[J].电子测量技术,2022,45(24): 137-144.
- [17] ZHENG G, SONGTAO L, ZEMING L, et al. OTA: Optimal transport assignment for object detection [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021: 303-312.
- [18] WANG J Q, CHEN K, XU R, et al. CARAFE: Content aware reassembly of features [C]. Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision, 2019: 3007-3016.
- [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] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature

- pyramid networks for object detection[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu: IEEE, 2017: 2117-2125.
- [21] QIAN S, CHEN L C, YUILLE A. Detectors: Detecting objects with recursive feature pyramid and switchable atrous convolution[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville: IEEE, 2021: 10213-10224.
- [22] LI H, LI J, WEI H, et al. Slim-neck by GSConv: A better design paradigm of detector architectures for autonomous vehicles [J]. ArXiv Preprint, 2022, ArXiv:2206.02424.
- [23] HAN K, WANG Y, TIAN Q, et al. Ghostnet: More features from cheap operations[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 1580-1589.
- [24] 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.
- [25] ZHANG W, CONG M, WANG L. Algorithms for optical weak small targets detection and tracking[C]. International Conference on Neural Networks and Signal Processing, 2003. Proceedings of the 2003, IEEE, 2003, 1: 643-647.

作者简介

卫策, 硕士研究生, 主要研究方向为计算机视觉目标检测。

E-mail: weice_5003_0@163.com

吕进(通信作者), 博士, 副教授, 主要研究方向为人工智能及智能优化、控制等。

E-mail: lvjin@chd.edu.cn

曲晨阳, 硕士研究生, 主要研究方向为计算机视觉目标检测。

E-mail: q18411012150423@163.com