

doi: 10.19562/j.chinasae.qcgc.2025.04.005

基于直方图特征蒸馏的密集交通目标检测*

张亿鸿¹, 钟铭恩¹, 谭佳威², 范康², 李正峰¹

(1. 厦门理工学院机械与汽车工程学院, 厦门 361024; 2. 厦门大学航空航天学院, 厦门 361005)

[摘要] 密集交通场景下的多类交通参与者目标检测仍是一项颇具挑战的视觉任务, 对于交通管理和安全至关重要。为此, 针对密集交通参与者的局部遮挡和小尺度特点, 提出一种深度神经网络检测算法 DSODet。首先采用轻量化的 CSPDarkNet 网络提取交通图像特征; 然后设计了多尺度特征融合上采样模块以增强对难检测目标的表达能力; 随后增加高分辨率检测分支来提升对小尺度目标的检测能力; 最后, 提出直方图特征蒸馏训练方法, 通过最小化教师模型与学生模型相同层特征直方图的交集比, 来有效引导学生模型训练, 实现参数优化与轻量化。实验结果表明, DSODet 对交通参与者的平均检测精度为 66.9%, 对局部遮挡的小尺度目标为 13.0%, 均超越现有主流算法, 模型参数量仅为 2.9 M, 体现了对边缘设备的友好性。相关代码将在 <https://github.com/XMUT-Vision-Lab> 分享。
关键词: 目标检测; 密集交通; 小尺度目标; 局部遮挡; 直方图特征蒸馏

Dense Traffic Object Detection Based on Histogram Feature Distillation

Zhang Yihong¹, Zhong Mingen¹, Tan Jiawei², Fan Kang² & Li Zhengfeng¹

1. School of Mechanical and Automotive Engineering, Xiamen University of Technology, Xiamen 361024;

2. School of Aerospace Engineering, Xiamen University, Xiamen 361005

[Abstract] Multi-class traffic participant detection in dense traffic scenarios remains a challenging visual task, which is crucial for traffic management and safety. To address this, a deep neural network-based detection algorithm, DSODet, is proposed to handle the challenges of partial occlusion and small-scale targets in dense traffic environment. Firstly, a lightweight CSPDarkNet network is used to extract features from traffic images. Then, a multi-scale feature fusion upsampling module is designed to enhance the representation capability for hard-to-detect targets. Next, a high-resolution detection branch is incorporated to improve detection accuracy for small-scale targets. Finally, a histogram feature distillation training method is proposed, which effectively guides the student model's training by minimizing the intersection ratio of feature histograms between the teacher and student models at corresponding layers, thus enabling parameter optimization and model compression. The experimental results show that DSODet achieves an average detection accuracy of 66.9% for traffic participants and 13.0% for small targets with partial occlusion, outperforming current state-of-the-art algorithms. The model contains only 2.9 M parameters, demonstrating its friendliness for edge device. The related code will be shared at <https://github.com/XMUT-Vision-Lab>.

Keywords: object detection; dense traffic; small-scale targets; partial occlusion; histogram feature distillation

* 福建省自然科学基金(2023J011439)资助。

原稿收到日期为 2024 年 10 月 16 日, 修改稿收到日期为 2024 年 12 月 09 日。

通信作者: 钟铭恩, 教授, 博士, E-mail: zhongmingen@xmut.edu.cn。

前言

随着城市化进程的加速和车辆保有量的快速增长,交通管理和安全问题日益突出,如何在复杂多变的交通场景中快速、准确地识别出各类交通参与者目标对于交通管理和交通安全都具有重要的现实意义^[1]。由于深度学习技术的快速发展,近年来利用神经网络从交通视频图像检测这些交通目标成为社会各界都积极探索的技术路径,诞生了诸如YOLO、Faster R-CNN、SSD和DETR等经典目标检测算法模型^[2]。这些模型具有优秀的综合性能,在各种普通交通场景中取得了良好的工作效果,但在面对密集交通场景时的目标检测能力还有待提升。密集交通时的目标遮挡问题仍然是当前研究共同面临的一大挑战。加大模型的复杂度和参数量被证明能够在一定程度上解决该问题。然而,由于大部分智能交通设施的存储和计算能力都相对较弱,因此如何在提升模型检测能力的同时尽可能减少模型对算力和存储资源的需求也是一个值得探究的现实课题。研究表明,引入蒸馏学习(knowledge distillation)可以在保持模型检测精度的同时降低参数规模,是解决该问题的一个可探索路径^[3]。

蒸馏学习利用复杂教师模型的知识来引导轻量化学生模型的训练,使得学生模型能够学习到更多有效的特征和模式,主要可分为逻辑蒸馏和特征蒸馏两大类。由于具有任务无关性和灵活性特点,特征蒸馏学习方法受到的关注日益增多,当前已经发展出各种先进的特征蒸馏方法,如经典的全局特征蒸馏法FitNet^[4]、前景和背景统一蒸馏框架FGD^[5]、通道概率图KL散度最小化法CWD^[6]等。最近的研究指出,在特征蒸馏方法中学生网络模型最好首先从教师网络模型中重构重要特征,而不是简单地追随教师网络模型,这样可以生成更有竞争性的特征表示,其中掩码特征蒸馏方法最受推崇。例如,MGD法通过随机生成的掩码来遮盖部分学生特征图,并迫使其通过一个简单网络来生成完整的教师特征图,从而提高学生特征图的表征能力^[7];AMD法利用注意力引导的掩码特征来识别学生特征图中的重要区域,而不再采用随机掩码^[8];DMKD法则是通过双重注意力机制来指导掩码分支,从而捕获空间重要性和通道信息,以实现更全面的掩码特征重建^[9]。存在的主要问题是:一方面,将原始学生模型的特征与掩码特征进行结合可能破坏原有学生模型

特征的结构和表达;另一方面,这些方法通常须设定多个超参数,且针对不同任务需要不同超参数调整过程,这无疑增加了模型训练的难度和复杂性。

为此,本文针对密集交通场景下的2D目标检测,提出了一种新的检测算法模型DSODet(dense small object detector using distillation learning)。首先该模型基于CSPDarknet增加了多尺度特征融合上采样模块MSFU(multi-scale feature-fusion upsampling module)和高分辨率检测分支SHead(small head)来优化场景中较远距离小尺度目标的检测精度;同时针对目前蒸馏学习方法存在的两个问题,即掩码可能破坏学生模型特征结构和表达,以及超参数较多的问题,提出了基于直方图的蒸馏训练方法,基于两个模型特征的直方图计算损失来进行模型参数优化与轻量化。最后实验验证了所提出的模型及蒸馏学习方法的可行性和有效性。

1 算法设计

1.1 师生孪生模型结构

DSODet蒸馏学习网络模型整体结构如图1所示,包含两个层次对称的网络分支模型,分别称为教师模型和学生模型。二者仅在网络的层数和通道数目上存在差别,且教师模型的层数和特征通道数远大于学生模型的层数和特征通道数。一般情况下,模型的层次越深、特征通道越多,则参数规模越大,检测精度更高,但处理效率更低且需要更强大的硬件算力支持。在训练DSODet时,利用教师模型来引导学生模型,使得学生模型能够以更少的参数规模达到与教师模型相近的检测精度。在利用DSODet进行交通目标检测时,仅学生模型分支处于工作模式而教师模型分支不参与,如此可提高任务推理效率。采用轻量级的CSPDarknet主干网络来提取图像的基础特征,并进一步应用PANet中的路径聚合机制来实现对不同尺度的特征进行融合^[10],提高特征的语义表达能力,促进信息在不同尺度之间的传递和融合,增强模型对不同尺度目标的感知能力。为进一步提高模型对小尺度目标的识别能力,针对性地增加了一个尺度为160×160的目标检测头,且其输入特征经过了专门设计的MSFU模块的上采样融合处理,能够更加高效地聚合来自不同尺度的语义信息。目标检测模块DHead(decoupled head)采用解耦结构来提高模型的灵活性,以更好地适应不同尺度和不同场景下的目标检测要求^[11]。特征融合上

采样模块FUM中各层级图像特征通过拼接方式进行融合,用符号⊕表示。基础卷积(CBS)、C2f特征聚

合模块以及快速空间金字塔池化模块(SPPF)设计灵感均来自于文献[11],在此不再赘述。

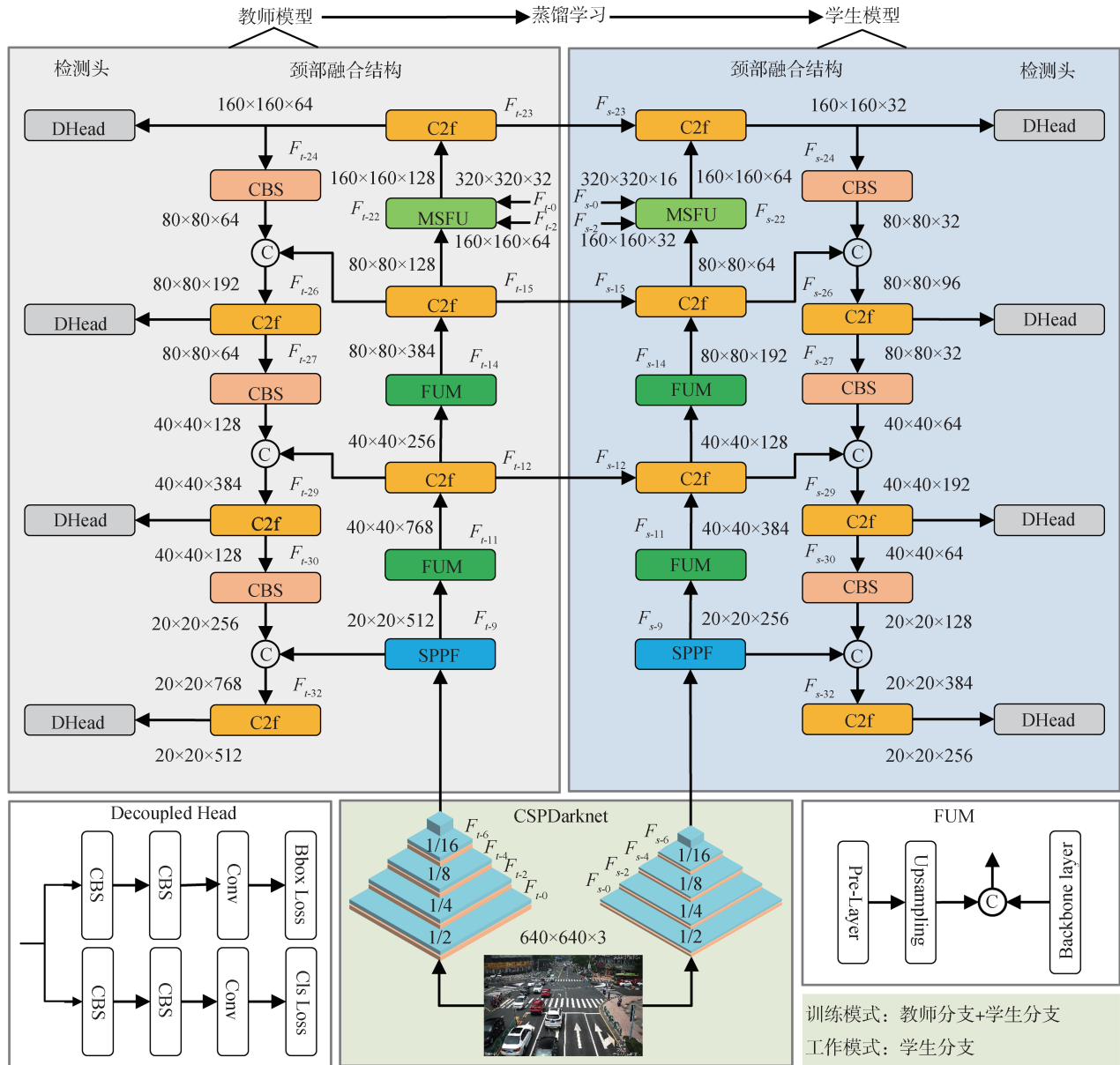


图1 DSODet整体结构

1.2 多尺度特征融合上采样

为获得更加精细的高分辨率图像特征,增强对小尺度目标的信息处理能力,本文结合多尺度特征信息构建了多尺度特征融合上采样模块MSFU,具体结构如图2所示。

MSFU模块的输入由两部分组成,一个是来自主干网络中不同分辨率层的空间细节特征 F_{high} 和 F_a ,另一个是待上采样的特征 F_{low} 。针对高分辨率特征 F_{high} ,首先通过C2f进行特征聚合,加强小尺度物

体的特征表示能力,再联合DWConv卷积、BN层和SiLu激活函数实现特征降维,将处理后的特征记为 F_a 。针对低分辨率特征 F_{low} ,首先将其上采样为 F_m ,其次将 F_a 和 F_m 融合为 $F_c \in \mathbb{R}^{H \times W \times 3C}$,然后再进行Conv卷积、BN批量归一化和SiLU激活处理,以获得更丰富的特征表示。最后将特征 F_c 与 F_a 进行拼接合并,获得新特征张量 $F_b \in \mathbb{R}^{H \times W \times 2C}$,并再次进行Conv卷积、BN和SiLU激活,最终得到包含多尺度信息的输出特征 F_o 。可以看出,MSFU在对特征进行上采样的同时融

合了各尺度空间细节信息,这有利于网络以更大的程度来重建图像特征,从而获取更丰富的特征信息。这种方案能够提高对局部遮挡和密集小尺度目标的特征收集和聚焦,从而提升针对密集交通参与者目标的检测精度。

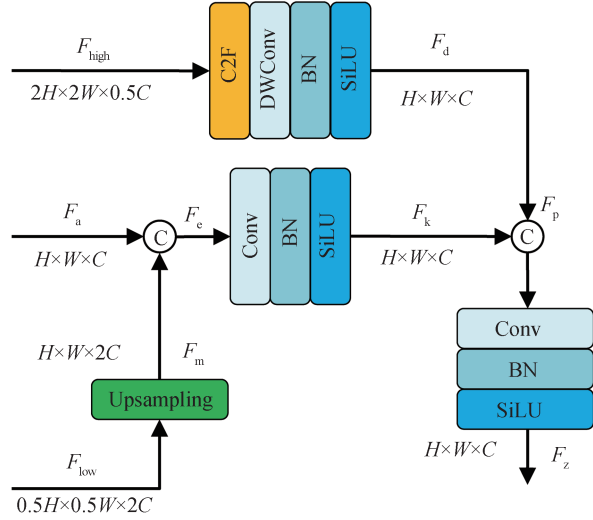


图2 多尺度特征融合上采样模块结构

1.3 直方图特征蒸馏训练原理

特征蒸馏是一种常用的模型压缩技术,一般利用教师网络模型提取的高级语义特征来引导学生网络模型训练,从而提升学生模型的综合性能。在传统特征蒸馏过程中,将损失函数定义为

$$Loss(F_t, F_{align}) = \sum_{c=1}^C \sum_{h=1}^H \sum_{w=1}^W (F_t^{h,w,c} - F_{align}^{h,w,c})^2 \quad (1)$$

$$F_{align} = \text{BatchNorm}(\text{Conv}(F_s(h, w, c_i))) \quad (2)$$

式中: C 、 H 和 W 分别表示特征图的通道数、高度和宽度; F_t 和 F_s 表示从教师模型和学生模型相同网络层的输出特征; F_{align} 表示将 F_s 与 F_t 进行通道对齐后的操作; $\text{Conv}(\cdot)$ 表示 1×1 卷积; BatchNorm 为批量归一化处理。

传统特征蒸馏通过直接模仿教师模型的特征来训练学生模型。在处理密集检测任务中,该方法可能忽视不同区域或通道的图像特征的重要性。针对该问题,当前主流的特征蒸馏转向通过生成教师模型掩码特征来重建学生模型特征,根据学生模型与教师模型的特征差异将损失函数改进为

$$M_s = \begin{cases} 0, & F_t^{h,w,c} \geq \tau \\ 1, & \text{其他} \end{cases} \quad (3)$$

$$F_{rec} = \alpha \times \text{Conv}(F_{align} \otimes M_s) \quad (4)$$

$$Loss(F_t, F_{rec}) = \sum_{c=1}^C \sum_{h=1}^H \sum_{w=1}^W (F_t^{h,w,c} - F_{rec}^{h,w,c})^2 \quad (5)$$

式中: M_s 表示掩码特征; τ 是一个阈值参数; α 是自适应权重因子; F_{rec} 为重建的学生模型特征; F_{align} 为特征适配层。超参数 τ 和 α 的引入一方面可能导致模型提取到较多的非重要区域的特征,影响学生模型的学习效果;另一方面,在模型训练时须调节多个超参数,增加了训练的复杂性。

为了更有效地引导学生模型学习教师模型的特征表示,本文提出一种直方图特征蒸馏HFD训练方法,将学生模型和教师模型特征间的直方图差异作为损失函数的一部分来引导学生模型的学习优化。具体算法流程如表1所示。首先分别计算学生模型和教师模型指定特征层 l 提取到的特征张量 $F_s^l \in \mathbb{R}^{H \times W \times C}$ 和 $F_t^l \in \mathbb{R}^{H \times W \times C}$ 的统计直方图并进行归一化处理,以确保它们的数值分布在相同尺度上,具体计算过程为

$$F_s^l = \sum_{i=1}^L \sum_{c=1}^C \sum_{h=1}^H \sum_{w=1}^W 1(b_i \leq F_s^l(h, w, c) < b_{i+1}) \quad (6)$$

$$F_t^l = \sum_{i=1}^L \sum_{c=1}^C \sum_{h=1}^H \sum_{w=1}^W 1(b_i \leq F_t^l(h, w, c) < b_{i+1}) \quad (7)$$

式中: F_{align}^l 表示特征适配层; $1(\cdot)$ 是指示函数表示如果条件为真,则取1,否则取0; B 表示直方图的bin数。

表1 直方图特征蒸馏算法流程

Algorithm: Histogram Feature Distillation (HFD)

Input: Student model feature tensors $\{F_s^l\}_{l=1,2,\dots,L}$;

Teacher model feature tensors $\{F_t^l\}_{l=1,2,\dots,L}$;

Alignment modules $\{align_l\}_{l=1,2,\dots,L}$;

Normalization modules $\{norm_l\}_{l=1,2,\dots,L}$;

Number of bins B ;

Output: Total loss L_{HFD}

1. Initialize $L_{HFD} \leftarrow 0$
2. **for** $l = 1$ to L **do**
3. $s_l \leftarrow norm_l(align_l(F_s^l))$
4. $t_l \leftarrow norm_l(F_t^l)$
5. $H_s^l \leftarrow Histogram(s_l, B)$, $H_t^l \leftarrow Histogram(t_l, B)$
6. $L_l \leftarrow 1 - \sum_i \min(H_s^l[i], H_t^l[i])$
7. $L_{HFD} \leftarrow L_{HFD} + L_l$
8. **end for**
9. Normalize total loss: $L_{HFD} \leftarrow \frac{L_{HFD}}{L}$
10. **return** L_{HFD}

然后根据式(9)计算交集损失,以此度量学生模型和教师模型所提取图像特征的相似性。

$$H_s^l(i) = \frac{\sum_{i=1}^l F_s^l(i)}{\sum_{i=1}^B F_s^l(i)}; \quad H_t^l(i) = \frac{\sum_{i=1}^l F_t^l(i)}{\sum_{i=1}^B F_t^l(i)} \quad (8)$$

$$Loss(l) = 1 - \sum_{i=1}^l \sum_{i=1}^B \min(H_s^l(i), H_t^l(i)) \quad (9)$$

式中: $H(i)_s^l$ 和 $H(i)_t^l$ 分别表示学生和教师模型在特征层 l 下,第 i 个bin的归一化直方图值; $\min(\cdot)$ 操作取每个bin上的最小值,这种做法考虑了特征在每个区域的最小匹配程度,能更准确地反映特征分布之间的相似性。

为得到整体的蒸馏损失,最后将所有特征层上的直方图交集损失累加起来,得到总的直方图特征蒸馏损失 L_{HFD} ,将其添加到网络训练总损失函数中,用于指导学生模型的训练优化过程。

$$L_{\text{HFD}} = \frac{1}{L} \sum_{l=1}^L (Loss(l)) \quad (10)$$

不难看出,相比于现有特征蒸馏方法,直方图特征蒸馏在保持图像重要区域关注度的同时无须额外引入超参数,简化了模型训练过程。

1.4 损失函数

在密集交通目标检测任务中,本文采用二元交叉熵损失 L_{Bce} 作为网络训练时的目标分类损失^[12],计算式为

$$L_1 = -\frac{1}{N} \sum_{i=1}^N y_i \log(p_i) \quad (11)$$

$$L_2 = -\frac{1}{N} \sum_{i=1}^N (1 - y_i) \log(1 - p_i) \quad (12)$$

$$L_{\text{Bce}} = L_1 + L_2 \quad (13)$$

式中: N 表示样本数量; y_i 表示样本的真实类别标签(0或1); p_i 表示网络预测的类别概率。

在对交通目标进行边界框回归定位时,采用联合交并比损失 L_{Ciou} 来衡量预测边界框和真实边界框之间的匹配程度^[13]。

$$IOU = \frac{|A \cap B|}{|A \cup B|} \quad (14)$$

$$\alpha = \frac{v}{(1 - IOU) + v} \quad (15)$$

$$L_{\text{Ciou}} = 1 - IOU + \frac{\rho^2(A, B)}{c^2} + \alpha \times v \quad (16)$$

$$v = \frac{4}{\pi^2} (\arctan(\frac{w_{\text{gt}}}{h_{\text{gt}}}) - \arctan(\frac{w}{h}))^2 \quad (17)$$

式中: IOU 表示预测边界框 A 和真实边界框 B 的交并比; α 是一个平衡参数; v 是关于 A 和 B 的宽高比和长宽比的函数; ρ 表示 A 和 B 的中心点之间的欧式距离; c 表示包围 A 和 B 的最小外接框的对角线长度。

在处理类别不平衡问题时,使用焦点损失 L_{Focal} 以减少易分类样本的损失,增强对难分类样本的关注^[14]。具体计算过程为

$$L_1 = \log(p_n(c)) \quad (18)$$

$$L_2 = g_n(c)(1 - p_n(c))^2 \quad (19)$$

$$L_{\text{Focal}} = -\frac{1}{N} \sum_{c=0}^{C-1} \sum_{n=1}^N L_1 \times L_2 \quad (20)$$

式中: N 表示样本数量; C 表示类别数量; $g_n(c)$ 是样本 n 是否属于类别 c 的指示函数; $p_n(c)$ 是模型对样本 n 属于类别 c 的概率。

最后,将训练总损失定义为分类损失、边界框回归损失、焦点损失和直方图特征蒸馏损失的加权和:

$$L_{\text{total}} = \alpha \times L_{\text{Bce}} + \beta \times L_{\text{Ciou}} + \gamma \times L_{\text{Focal}} + L_{\text{HFD}} \quad (21)$$

式中 α 、 β 和 γ 表示平衡网络不同部分对整体影响的权重参数。为保证算法性能对比的公平性,本实验中采用与文献[11]中相同的权重系数设置,即分别取值为0.5、7.5、1.5。

2 实验与分析

2.1 数据集

本文在密集交通场景公开数据集SEU_PML^[15]上进行实验。该数据集主要为城镇混合道路高清监控摄像头视野下的多类密集交通场景图像,覆盖了不同的天气和照明条件,包含270 684个带2D边界框注释的交通参与对象,可分为行人、机动车、非机动车和其它4大类。本文利用其训练集来训练DSODet和其它对比算法模型,但因该数据集未公开测试集的真实标签,本文在验证集上进行各算法的性能对比。

2.2 实验条件与参数设置

实验主机的操作系统为64位Windows 10,硬件采用Intel Core i5-13600KF CPU和NVIDIA GeForce RTX 3060显卡。算法开发环境采用Python 3.9.18和PyTorch深度学习框架。模型训练时,采用SGD优化器,初始学习率设置为0.01,权重衰减率设置为 5×10^{-4} ,学习率衰减策略选择线性衰减。不同对比模型都进行300轮次训练,批处理大小都设置为8。数据加载和预处理时,将图像大小统一调整为640×640,并采用随机缩放、翻转、扭曲等操作对数据集进行几何增强,以及随机调整图片颜色、饱和度和亮度进行光照增强。

2.3 评价指标

选择平均精度 mAP (mean average precision, mAP)作为目标检测性能的评价指标:

$$P = \frac{TP}{TP + FP} \quad (22)$$

$$R = \frac{TP}{TP + FN} \quad (23)$$

$$AP = \int_0^1 P(R) dR \quad (24)$$

$$mAP = \frac{1}{N} \sum AP_i^{IOU=0.7} \quad (25)$$

$$MR = 1 - R \quad (26)$$

式中: TP 表示真阳性目标数量; FP 表示假阳性目标数量; FN 表示假阴性目标数量; P 为准确率; R 为召回率; MR 表示漏检率; N 为类别总数量; AP_i 表示第*i*个类别的平均精度; $IOU=0.7$ 表示预测框与真实框重叠程度达到或超过70%即为有效检测。与此同时,引入针对不同尺度目标的平均精度指标,包括小尺度目标 AP_s 、中等尺度目标 AP_m 和大尺度目标 AP_l 。选取模型参数量 N_p 和每秒浮点运算数GFLOPs来衡量模型的内存占用程度和计算复杂度,采用每秒传输帧数FPS衡量模型的推理速度。

2.4 消融实验

本文进行了一系列消融实验,旨在探明所提出的多尺度特征融合上采样模块MSFU、小尺度目标检测头SHead和直方图特征蒸馏HFD训练方法对DSODet算法模型的性能影响情况,具体实验结果如表2所示。

表2 消融实验结果 %

模型	mAP	AP_s	AP_m	AP_l	MR
Baseline	55.6	6.3	29.4	47.0	49.1
+SHead	63.1	11.2	34.9	47.5	43.1
+MSFU	65.6	12.6	37.3	49.7	41.4
+HFD	66.9	13.0	38.4	50.8	40.7

可以看出:(1)相比于基线模型Baseline,添加具有解耦结构的小目标检测头SHead后, mAP 和 AP_s 分别增加7.5%和4.9%, MR 下降6.0%。这表明添加一个高分辨率浅层图像特征能加强小尺度特征的代表能力,对于密集小尺度目标的检测很有帮助。(2)进一步添加多尺度特征融合上采样模块MSFU后,平均检测精度 mAP 提升了2.5%,针对小尺度目标、中等尺度目标和大尺度目标的检测精度分别提升1.4%、2.4%和2.2%,漏检率 MR 下降了1.7%。这表明MSFU能够较好地融合图像中的各尺度特

征,帮助模型获得更全面丰富的语义信息。(3)采用直方图特征蒸馏训练方法HFD后, mAP 再次提高了1.3%,达到66.9%, AP_s 提升至13.0%,且 MR 下降至40.7%,模型性能获得了进一步优化,证明了直方图特征蒸馏训练方法的有效性。

综上所述,相较于基线模型Baseline,DSODet通过设计的高分辨率小尺度目标检测头、多尺度特征融合上采样模块和直方图特征蒸馏训练方法,不仅显著提升了 mAP ,从55.6%提高到66.9%,还大幅降低了漏检率,从49.1%降至40.7%,显著地提高了在密集交通流下对各类交通参与者目标的识别精度。

图3对比了消融实验中基线模型Baseline(图左列,标注为1)和最终模型DSODet(图右列,标注为2)在不同交通十字路口场景的推理结果。图中红色、粉色、橙色和黄色边界框分别代表行人、车辆、非机动车和车牌。对于图3(a)所示交通场景,从局部放大图中可以看出,基线模型对于远处的小尺度车辆目标存在较多的漏检案例,而DSODet都能够成功检测到这些目标对象。对于图3(b)所示存在较多前后纵向局部遮挡车辆的交通场景,DSODet的漏检数量相对较少,且对于斑马线附近的二轮车和行人对象具有更好的检测效果。对于图3(c)所示的存在密集行人的交通场景,DSODet的检测效果相比更好。

2.5 不同特征蒸馏训练方法的对比

为验证直方图特征蒸馏HFD的有效性,本文将其应用于DSODet模型的训练。与此同时,选取了现有主流蒸馏方法来进行实验对比,包括针对密集目标检测任务的最小化概率散度法CWD、基于掩码特征重建的FGD法和MGD法。最后获得的实验结果如表3所示。可以看出,与现有主流特征蒸馏方法相比,本文提出的直方图特征蒸馏训练方法HFD对于网络模型的平均检测精度 mAP 提升效果最明显,相比于不采用蒸馏训练的基本模型增加了1.3%,达到最高值66.9%。对于小尺度目标和大尺度目标的检测精度 AP_s 和 AP_l 都取得最优值,而对于中等尺度目标的检测精度 AP_m 取得了次优值38.4%,仅比最小化概率散度法取得的最优值低了0.1%。这些结果证明,HFD确实能够提高DSODet模型对于交通场景中的密集小尺度目标的检测能力。

2.6 与不同检测算法的性能对比

为评估DSODet的性能优劣,将其与现有主流目标检测算法进行实验对比,对比对象包括采用CNN



图3 消融实验中Baseline(左)与DSODet(右)目标检测效果对比

表3 采用不同特征蒸馏方法的实验结果

蒸馏方法	$mAP\%$	$AP_s\%$	$AP_m\%$	$AP_l\%$	$N_p/10^6$
DSODet	65.6	12.6	37.3	49.7	2.93
MGD	66.3	12.7	38.0	50.3	2.93
FGD	66.5	12.9	38.2	50.5	2.93
CWD	66.6	12.8	38.5	50.4	2.93
HFD	66.9	13.0	38.4	50.8	2.93

架构的 Faster R-CNN^[2]、Cascade R-CNN^[2]、Sparse R-CNN^[16]、SSD^[2]、Dyhead^[17]、YOLOX^[11]、YOLOv5^[11]、YOLOv8^[11]、YOLOv9^[18]、YOLOv10^[19]、YOLOv11^[20]和采用 Transformer 架构的 DETR^[2]、DAB-DETR^[21]、DINO^[22]、RT-DETR^[23]。详细结果如表4所示。

从表4可以看出:(1)DSODet在密集交通目标检测任务上的平均检测精度 mAP 、小尺度目标检测精度 AP_s 、中等尺度目标检测精度 AP_m 和漏检率 MR 分别为66.9%、13.0%、38.4%和40.7%,都取得了全局最优值。(2)针对大尺度目标的检测精度 AP_l 取

得了全局次优值50.8%,仅次于YOLOv9取得的全局最优值52.7%,但DSODet的参数规模 N_p 和浮点运算总量GFLOPs仅分别为YOLOv9模型的20.7%(2.9/14.0)和25.9%(13.7/52.9),相对轻量化,对算力较弱的边缘设备更为友好。(3)平均每秒时间可以检测144帧图像,即单帧图像推理所需的耗时仅为6.94ms,具有较高的处理效率。这些结果说明所设计的DSODet算法模型相比现有主流目标检测网络具有更好的综合性能。

综上,在多尺度融合上采样模块MSFU和高分辨率小尺度目标检测头Shead的作用下,DSODet在较低参数量下仍然表现优异。

图4展示了本文算法DSODet与对比算法YOLOv9在不同光照条件下的交通场景图像特征提取的热力图可视化结果。其中第1列为输入原图,第2列为标签热力图,第3、4列分别为YOLOv9和DSODet的特征热力图。所谓标签特征热力图是仅

表4 不同算法模型的性能对比结果

模型名称	$mAP_s/\%$	$AP_s/\%$	$AP_m/\%$	$AP_l/\%$	$MR/\%$	$N_p/10^6$	GFLOPs	FPS
Faster R-CNN	53.4	0.7	17.9	46.4	67.9	41.4	64.0	30
Cascade R-CNN	52.9	0.8	19.0	46.5	66.9	69.4	92.0	19
Sparse RCNN	54.6	5.0	19.4	33.6	61.3	106.0	45.8	30
Dyhead	46.8	3.1	21.8	44.5	58.3	38.9	28.4	23
SSD	41.1	2.7	15.2	37.6	58.0	24.2	30.5	57
YOLOX	54.0	6.1	24.9	42.6	42.6	5.1	3.2	70
YOLOv5	51.7	5.6	23.9	38.1	55.1	1.8	4.1	77
YOLOv8	56.9	6.3	29.4	47.0	47.8	3.2	8.9	167
YOLOv9	63.3	9.4	34.7	52.7	42.2	14.0	52.9	59
YOLOv10	54.0	6.1	29.1	46.5	46.0	2.7	8.2	184
YOLOv11	55.1	6.4	29.5	47.3	50.0	2.6	6.3	193
DETR	29.9	1.0	7.6	29.9	69.2	41.6	25.1	60
DAB-DETR	37.8	4.6	15.3	21.7	64.4	43.7	29.4	47
DINO	57.5	7.4	28.6	46.9	45.2	47.6	80.4	19
RT-DETR	53.5	7.3	27.5	41.7	49.7	41.9	125.6	34
DSODet(ours)	66.9	13.0	38.4	50.8	40.7	2.9	13.7	144

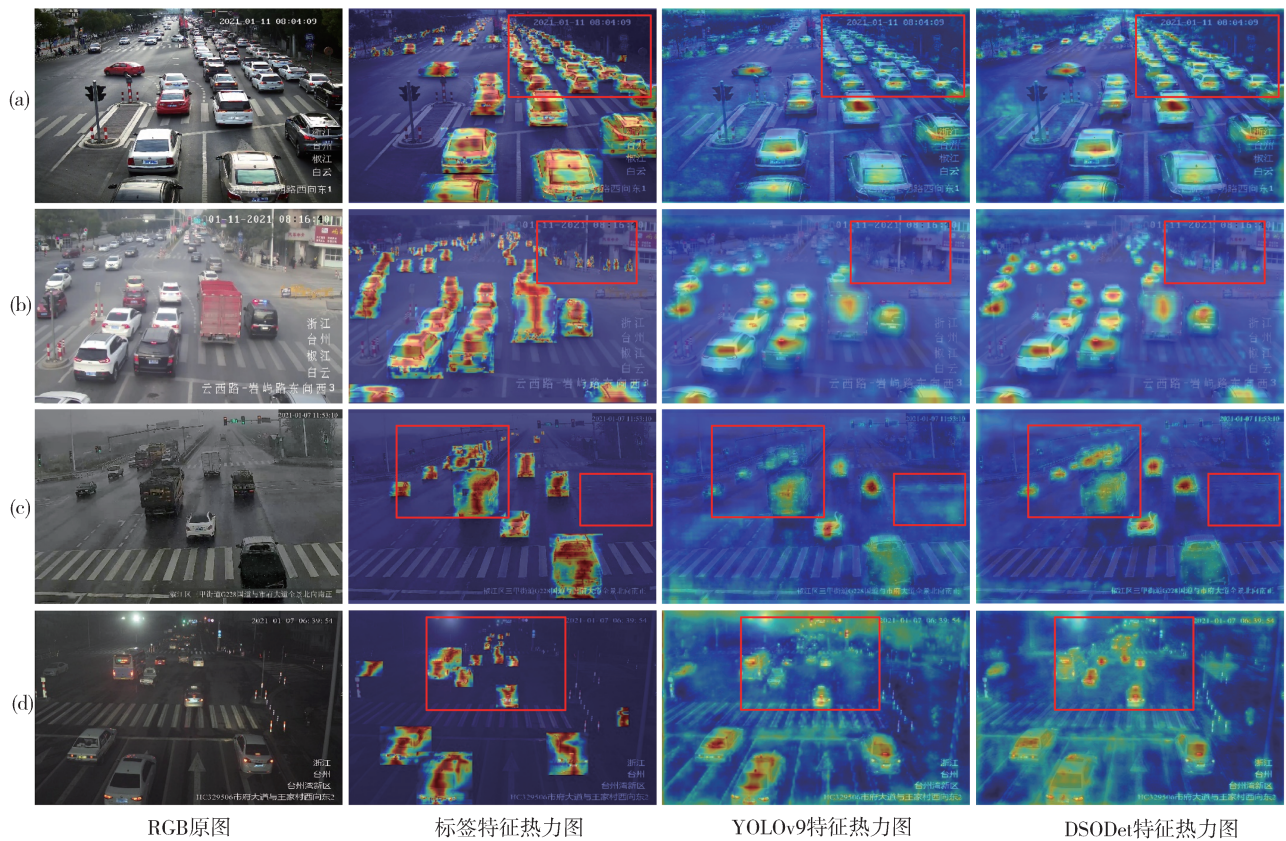


图4 不同算法的特征热力图效果对比

在目标真实标签定位框内进行渲染的热力图。具体而言,首先通过 GradCAM 等方法生成图像的完整类激活映射^[24],用以表征模型在不同位置的关注强度。然后通过目标的真实定位标签进行区域裁剪并进行

标准化处理,使得生成的特征热力图仅聚焦于目标区域而忽略其他无关区域。最后将生成的热力图与原始图像叠加融合,从而更好地展示模型对检测目标的关注情况。标签特征热力图能够为各算法的特

征热力图对比提供一种对照参考。图4(a)为天气晴朗的白天时段内的典型交通场景,从其中红色方框所圈选的内容可以看出DSODet对远处密集车辆区域和局部遮挡车辆区域的目标特征关注度更高,这对于提升检测的准确率极其有益。图4(b)为雾天白天时的典型交通场景,显然可见DSODet捕获小尺度目标行人特征的能力更强。图4(c)为雨天黄昏时的典型交通场景,可以看出DSODet在交通目标特征提取时具有更强的抗扰能力,能有更好的过滤雨水等全局背景影响。图4(d)为夜间光照较弱时的典型交通场景,此时车辆的红色尾灯易与红灯相混淆,从而影响到目标检测的准确度。可以看出DSODet模型能够更好处理这些局部干扰而更关注于感兴趣的交通参与者目标,鲁棒性更好。这些案例表明,DSODet能够更好地挖掘道路交通环境感知所需的各类关键要素信息,降低各类全局和局部干扰的影响,在密集交通目标感知视觉任务上具有更优的检测性能和更好的适应能力。

3 结论

为更好地应对密集交通场景下多类参与者目标的视觉感知挑战,本文设计了一种深度卷积神经网络模型DSODet来提升针对局部遮挡和密集小尺度交通目标的检测能力,并提出直方图特征蒸馏训练方法实现模型轻量化,最终在检测精度和参数规模上获得了较好均衡。在密集交通流公开数据集SEU_PML上的实验结果表明,该模型针对各类交通参与者的平均检测精度达到66.9%,针对局部遮挡、密集小尺度目标的检测精度达到13.0%,均超越了现有主流算法,且模型参数量仅为2.9M,对边缘设备较为友好,可为解决城市密集交通目标检测难题提供技术支持和经验参考。后续研究中可增加和细化交通参与者类别,并考虑增加算法对各类交通违法行为的自动判断和对交通事故的实时侦测,为城市交通场景的综合感知技术研究贡献力量。

参考文献

- [1] GHAREMANNEZHAD H, SHI H, LIU C, et al. Object detection in traffic videos: a survey [J]. IEEE Transactions on Intelligent Transportation Systems, 2023, 24(7): 6780-6799.
- [2] AMJOURD A B, AMROUCH M. Object detection using deep learning, CNNs and vision transformers: a review [J]. IEEE Access, 2023, 11: 35479-35516.
- [3] HUANG G, SHEN A, HU Y, et al. Optimizing YOLOv5s object detection through knowledge distillation algorithm [J/OL]. Computer Science, arXiv preprint arXiv: 2410.12259, 2024.
- [4] GOU J, YU B, MAYBANK S J, et al. Knowledge distillation: a survey [J]. International Journal of Computer Vision, 2021, 129(6): 1789-1819.
- [5] YANG Z, LI Z, JIANG X, et al. Focal and global knowledge distillation for detectors [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, LA, USA, IEEE Press, 2022: 4633-4642.
- [6] SHU C, LIU Y, GAO J, et al. Channel-wise knowledge distillation for dense prediction [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal, QC, Canada, IEEE, 2021: 5291-5300.
- [7] YANG Z, LI Z, SHAO M, et al. Masked generative distillation [J/OL]. Computer Science, arXiv preprint arXiv: 2205.01529, 2022.
- [8] YANG G, TANG Y, LI J, et al. AMD: adaptive masked distillation for object detection [C]. 2023 International Joint Conference on Neural Networks. Gold Coast, Australia, IEEE Press, 2023: 1-8.
- [9] YANG G, TANG Y, WU Z, et al. DMKD: improving feature-based knowledge distillation for object detection via dual masking augmentation [C]. IEEE International Conference on Acoustics, Speech and Signal Processing. Seoul, Korea, Republic of, IEEE Press, 2024: 3330-3334.
- [10] DENG C, WANG M, LIU L, et al. Extended feature pyramid network for small object detection [J]. IEEE Transactions on Multimedia, 2021, 24: 1968-1979.
- [11] TERVEN J R, CORDOVA-ESPARZA D M, ROMERO-GONZÁLEZ J A. A comprehensive review of YOLO architectures in computer vision: from YOLOv1 to YOLOv8 and YOLO-NAS [J]. Machine Learning and Knowledge Extraction, 2023, 5(4): 1680-1716.
- [12] WADDAR R, RATHOD V, NETRAVATI H, et al. A CNN-based stutter detection using MFCC features with binary cross-entropy loss function [C]. IEEE International Conference on Contemporary Computing and Communications. Bangalore, India, IEEE Press, 2024, 1: 1-6.
- [13] HUANG P, TIAN S, SU Y, et al. IA-CIOU: an improved IOU bounding box loss function for SAR ship target detection methods [J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2024, 17: 10569-10582.
- [14] YANG B, ZHANG X, ZHANG J, et al. EFLNet: enhancing feature learning network for infrared small target detection [J]. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62: 1-11.
- [15] ZHOU W, WANG C, XIA J, et al. Monitoring-based traffic participant detection in urban mixed traffic: a novel dataset and a tailored detector [J]. IEEE Transactions on Intelligent Transportation Systems, 2023, 25(1): 189-202.

(下转第679页)

- 踪控制[J]. 汽车工程, 2023, 45(11): 2092-2103.
- HU J, CHEN R P, ZHANG Z H, et al. Path tracking control of autonomous trucks based on RMPC[J]. *Automotive Engineering*, 2023, 45(11): 2092-2103.
- [16] CHENGLIANG Z, GUANLEI G, CHUNZHAO Z, et al. Research on 4WS agricultural machine path tracking algorithm based on fuzzy control pure tracking model[J]. *Machines*, 2022, 10(7): 597-597.
- [17] WANG H, CHEN X, CHEN Y, et al. Trajectory tracking and speed control of cleaning vehicle based on improved pure pursuit algorithm[C]. 中国自动化学会控制理论专业委员会(Technical Committee on Control Theory, Chinese Association of Automation), 中国自动化学会(Chinese Association of Automation), 中国系统工程学会(Systems Engineering Society of China). 第三十八届中国控制会议论文集(3). School of Mechatronic Engineering and Automation, Shanghai University; Chinese Academy of Agriculture Mechanization Sciences; Shanghai Key Laboratory of Power Station Automation Technology, 2019:6.
- [18] 潘世举,李永乐,李子先,等. 基于改进纯跟踪的智能车路径跟随方法[J]. 汽车工程, 2023, 45(1): 1-8, 19.
- PAN S J, LI Y L, LI Z X, et al. Path following method f-or intelligent vehicles based on improved pure pursuit [J]. *Automotive Engineering*, 2023, 45(1): 1-8, 19.
- [19] QINPENG S, ZHONGHUA W, MENG L, et al. Path tracking control of wheeled mobile robot based on improved pure pursuit algorithm [C]. 2019 Chinese Automation Congress (CAC). IEEE, 2019: 4239-4244.
- [20] 肖礼明,张发旺,陈良发,等. 依托多风格强化学习的车辆轨迹跟踪避撞控制[J]. 汽车工程, 2024, 46(6): 945-955.
- XIAO L M, ZHANG F W, CHEN L F, et al. Vehicle trajectory tracking and collision avoidance control based on multi-style reinforcement learning [J]. *Automotive Engineering*, 2024, 46(6): 945-955.
- [21] 韩陌,何洪文,石曼,等. 基于学习的无人驾驶车辆模型预测路径跟踪控制研究[J]. 汽车工程, 2024, 46(7): 1197-1207.
- HAN M, HE H W, SHI M, et al. Research on model predictive path tracking control of autonomous vehicles based on learning [J]. *Automotive Engineering*, 2024, 46(7): 1197-1207.
- [22] 刘聪,刘辉,韩立金,等. 基于学习型滑模预测控制的无人驾驶车辆非结构化环境轨迹跟踪及稳定性控制[J]. 机械工程学报, 2024, 60(10): 399-412.
- LIU C, LIU H, HAN L J, et al. Trajectory tracking and stability control of autonomous vehicles in unstructured environments based on learning-type sliding mode predictive control [J]. *Journal of Mechanical Engineering*, 2024, 60(10): 399-412.
- [23] 杨阳阳,何志刚,汪若尘,等. 智能车辆路径跟踪横向混合控制器设计[J]. 重庆理工大学学报(自然科学), 2018, 32(11): 7-14.
- YANG Y Y, HE Z G, WANG R C, et al. Design of a lateral hybrid controller for path tracking of intelligent vehicles [J]. *Journal of Chongqing University of Technology (Natural Science)*, 2018, 32(11): 7-14.
- [24] 吴西涛,魏超,翟建坤,等. 考虑横摆稳定性的无人车轨迹跟踪控制优化研究[J]. 机械工程学报, 2022, 58(6): 130-142.
- WU X T, WEI C, ZHAI J K, et al. Optimization research on trajectory tracking control of unmanned vehicles considering yaw stability [J]. *Journal of Mechanical Engineering*, 2022, 58(6): 130-142.

(上接第644页)

- [16] SUN P, ZHANG R, JIANG Y, et al. Sparse R-CNN: an end-to-end framework for object detection[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, 45(12): 15650-15664.
- [17] 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*. Nashville, TN, USA, IEEE Press, 2021: 7373-7382.
- [18] WANG C Y, YE H I, LIAO H Y M. YOLOv9: learning what you want to learn using programmable gradient information [J/OL]. *Computer Science*, arXiv preprint arXiv: 2402.13616, 2024.
- [19] WANG A, CHEN H, LIU L, et al. YOLOv10: real-time end-to-end object detection [J/OL]. *Computer Science*, arXiv preprint arXiv: 2405.14458, 2024.
- [20] KHANAM R, HUSSAIN M. YOLOv11: an overview of the key architectural enhancements [J/OL]. *Computer Science*, arXiv preprint arXiv: 2410.17725, 2024.
- [21] LIU S, LI F, ZHANG H, et al. DAB-DETR: dynamic anchor boxes are better queries for detr[J/OL]. *Computer Science*, arXiv preprint arXiv: 2201.12329, 2022.
- [22] ZHANG H, LI F, LIU S, et al. DINO: DETR with improved denoising anchor boxes for end-to-end object detection [J/OL]. *Computer Science*, arXiv preprint arXiv: 2203.03605, 2022.
- [23] LV W, ZHAO Y, XU S, et al. DETRs beat YOLOs on real-time object detection[C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, WA, USA, IEEE Press, 2024: 16965-16974.
- [24] ZHANG Y, ZHU Y, LIU J, et al. An interpretability optimization method for deep learning networks based on Grad-CAM [J]. *IEEE Internet of Things Journal*, 2024, Accession number 20244517311551: 1-8.