

doi: 10.19562/j.chinasae.qcgc.2025.02.001

基于多粒度关系推理的自动驾驶域自适应视觉目标检测算法*

索锦辉¹, 王晓伟^{1,2}, 蒋沛文¹, 丁驰³, 高铭^{1,2}, 边有钢^{1,2}

(1. 湖南大学, 整车先进设计制造技术全国重点实验室, 长沙 410082; 2. 湖南大学无锡智能控制研究院, 无锡 214072;

3. 清华大学车辆与运载工程学院, 北京 100084)

[摘要] 现有域自适应视觉目标检测算法大多基于两阶段检测器设计, 且未能利用图像空间中不同元素之间的语义拓扑关系, 导致次优的跨域适应性能。为此, 本文提出一种基于多粒度关系推理的域自适应视觉目标检测算法。首先, 提出粗粒度图块关系推理模块, 使用粗粒度图块图结构来捕获前景和背景之间的拓扑关系, 对前景区域进行跨域适配。然后, 设计细粒度语义关系推理模块, 推理细粒度语义图结构来增强跨域多类别语义依赖关系。最后, 提出粒度诱导的特征对齐模块, 根据节点的亲和性调节特征对齐的权重, 提升检测模型面对场景整体变化时的适应性。多个自动驾驶跨域场景上的实验结果验证了所提算法的鲁棒性和实时性。

关键词: 自动驾驶; 视觉目标检测; 域自适应; 图推理

Domain Adaptive Visual Object Detection for Autonomous Driving Based on Multi-granularity Relation Reasoning

Suo Jinhui¹, Wang Xiaowei^{1,2}, Jiang Peiwen¹, Ding Chi³, Gao Ming^{1,2} & Bian Yougang^{1,2}

1. Hunan University, State Key Laboratory of Advanced Design and Manufacturing Technology for Vehicle, Changsha 410082;

2. Wuxi Intelligent Control Research Institute of Hunan University, Wuxi 214072;

3. School of Vehicle and Mobility, Tsinghua University, Beijing 100084

[Abstract] Most of the existing domain adaptive visual object detection algorithms are based on two-stage detector design and fail to exploit the semantic topological relationship between different elements in the image space, resulting in suboptimal cross-domain adaptation performance. Therefore, in this paper a domain adaptive visual object detection algorithm based on multi-granularity relationship reasoning is proposed. Firstly, a coarse-grained patch relationship reasoning module is proposed, which uses the coarse-grained patch graph structure to capture the topological relationship between the foreground and background and perform cross-domain adaptation on the foreground area. Then, a fine-grained semantic relationship reasoning module is designed to reason about the fine-grained semantic graph structure to enhance cross-domain multi-category semantic dependencies. Finally, a granularity-induced feature alignment module is proposed to adjust the weight of feature alignment according to the affinity of the nodes, thereby improving the adaptability of the detection model when facing overall scene changes. The experimental results on multiple cross-domain scenarios of autonomous driving verify the robustness and real-time performance of the proposed algorithm.

Keywords: autonomous driving; visual object detection; domain adaptation; graph reasoning

* 国家重点研发计划项目(2022YFB2503402)、长三角科技创新共同体联合攻关计划项目(2023CSJGG0801)、国家自然科学基金青年科学基金项目(52102461)、湖南省青年科技创新人才项目(2022RC1033)和湖南省自然科学基金项目(2023JJ40155)资助。

原稿收到日期为2024年07月14日, 修改稿收到日期为2024年08月28日。

通信作者: 王晓伟, 副研究员, 博士, E-mail: wxw9@163.com。

前言

视觉感知技术是自动驾驶汽车^[1-2]捕获外部环境的重要技术之一。其中,视觉目标检测^[3-4]作为自动驾驶视觉感知的基础任务,需要对车辆和行人等交通参与者进行定位和分类。然而,实际场景数据和训练数据之间存在“域偏移”,经过训练的检测模型可能无法适应新场景,面临性能退化问题。一般来说,可以采集新数据重新训练模型,但制作新场景的标注数据会耗费巨大成本。

作为迁移学习的重要分支,无监督域自适应(unsupervised domain adaptation, UDA)旨在将知识从有标签的源域传递到无标签的目标域,增强模型解决新问题的能力。近年来,已有一些域自适应视觉目标检测方法被提出来解决域偏移问题。其中,DA-Faster^[5]在图像级和实例级进行特征分布对齐,指导模型学习域不变特征表示。SW-Faster^[6]对低级特征进行强对齐,对高级特征进行弱对齐,更好地调整特征对齐效果。但这种类别无关的全局对齐只能适应色彩、风格和纹理上的低级域偏移,无法应用到更复杂的场景。为实现类别语义上的适应,MMCN^[7]使用特定类别的域分类器,同时负责预测特征的域类别和检测类别。C2F^[8]建模不同类别的实例原型,进行基于类别中心的约束对齐。但这些基于类别原型的方法忽视了域内类别方差,造成不稳定的类别适应。为此,SIGMA^[9]将类别像素建模为语义图,并进行跨域图匹配来完成精细化的类别条件适应。

以上域自适应目标检测方法提高了模型的跨域检测能力。但仍存在两方面的问题:一是现有大多数域自适应方法只针对经典两阶段目标检测模型Faster RCNN^[10]定制,不满足自动驾驶目标检测的实时性要求,虽然近几年也有单阶段检测模型SSD^[11]、FCOS^[12]被用于跨域检测研究,但它们很少被用于自动驾驶;二是现有方法没有考虑图像空间内的多元素的语义拓扑关系,例如前景目标和驾驶背景之间存在潜在的拓扑联系,不同目标类别之间具有不同的语义关系等等。这种不同域之间的语义拓扑关系的差异反映了更高级的域信息结构。以往方法忽视了语义拓扑关系的跨域匹配,语义空间的歧义必然导致次优的适应性能。

为此,本文基于实时目标检测模型YOLO^[13-16]提

出了一种基于多粒度关系推理(multi-granularity relation reasoning, MGR²)的域自适应目标检测算法,它构建粗细粒度图来推理多元素之间的语义拓扑关系。具体而言,粗粒度图块关系推理模块将前背景区域特征建模成粗粒度图节点,借助前景和背景区域的拓扑关系进行跨域适配。细粒度语义关系推理模块将多类别实例特征建模成细粒度图节点,增强源域和目标域类别语义的依赖关系。粒度诱导的特征对齐模块根据图节点的亲和性,调整源域和目标域全局特征对齐的权重,提升检测模型面对场景整体变化时的适应性。

本文的主要贡献包括以下4个方面:

(1)提出一种基于YOLO的实时域自适应目标检测算法。与之前算法相比,具有推理迅速、计算高效的优势,能满足智能汽车的实时性要求。

(2)提出一种多粒度关系推理的域自适应目标检测框架,能利用多元素的语义拓扑关系进行跨域适应,具有更强的跨域鲁棒性。

(3)提出粗粒度图块关系推理模块,完成感兴趣区域的跨域适配;提出细粒度语义关系推理模块,增强跨域多类别依赖关系;提出粒度诱导的特征对齐模块,提升检测模型对场景整体变化的适应性。

(4)在多个跨域场景下开展实验,验证了算法设计的合理性和有效性。

1 基本原理

1.1 域自适应目标检测问题定义

本文遵循视觉目标检测的无监督域自适应通用设置^[5],具有 N^s 数量的带标签图像源域记为 $\mathcal{D}^s = \{(x_i^s, y_i^s)\}_{i=1}^{N^s}$,具有 N^t 数量的无标签图像目标域记为 $\mathcal{D}^t = \{(x_i^t)\}_{i=1}^{N^t}$ 。其中, x_i^s 表示第 i 张源域图像, $y_i^s = (b_i^s, c_i^s)$ 表示第 i 张源域图像对应的标签,即边界框的位置标签 b_i^s 及其类别标签 c_i^s ; x_i^t 表示第 i 张目标域图像。虽然源域和目标域来自不同的联合概率分布(即 $\mathcal{P}^s \neq \mathcal{P}^t$),但共享相同的类别空间(即 $c_i^{st} \in \{0, 1, \dots, K-1\}$)。其中, c_i^{st} 表示第 i 张源域或目标域图像中真实存在的目标类别, K 表示跨域视觉目标检测任务中待检测类别的总数。域自适应视觉目标检测的目的是借助带标签的源域,学习一个在无标签的目标域上最优的视觉目标检测模型。

1.2 网络概述

图1展示了本文提出的多粒度关系推理的域自

目标检测算法框架。该框架基于单阶段实时检测模型YOLO搭建,主要包含粗粒度图块关系推理(coarse-grained patch relational reasoning, CGPR²)模块、细粒

度语义关系推理(fine-grained semantic relational reasoning, FGSR²)模块和粒度诱导的特征对齐(granularity-induced feature alignment, GIFA)模块。

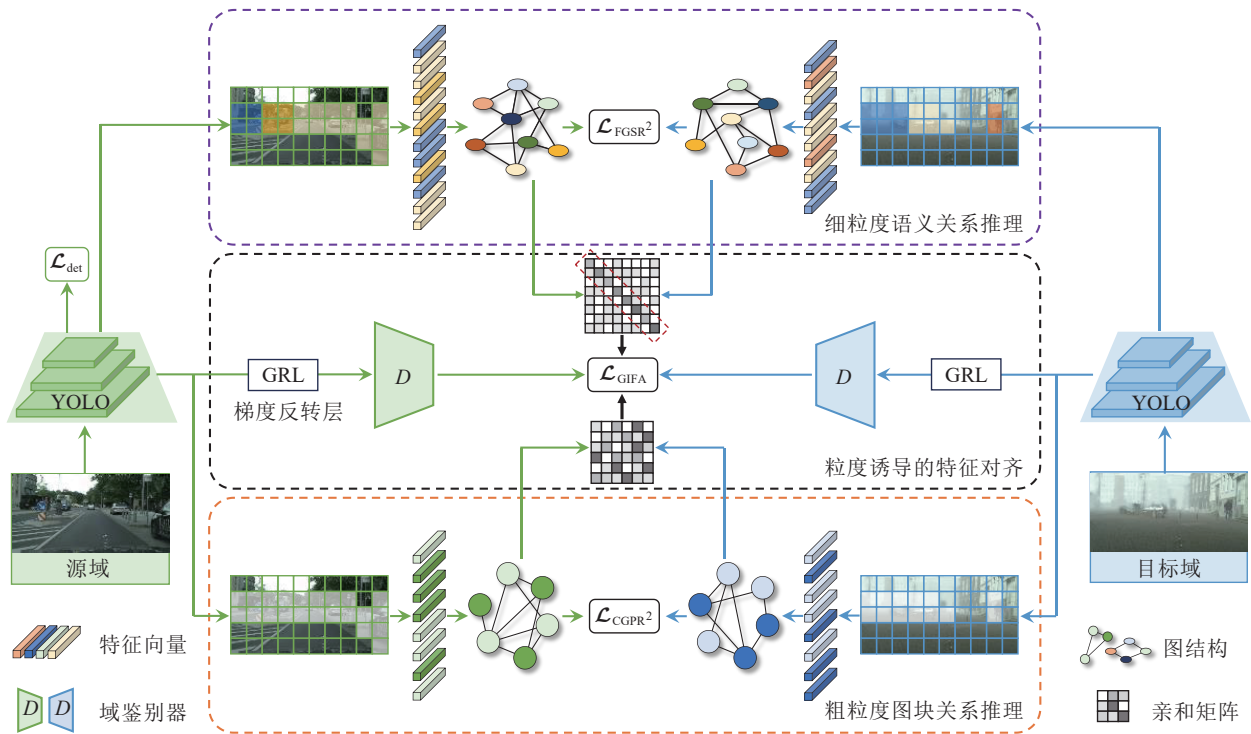


图1 基于多粒度关系推理的跨域目标检测算法框架

1.3 模块细节

1.3.1 粗粒度图块关系推理模块

图1下半部分为粗粒度图块关系推理模块的基本流程。第1步,获取源域和目标域的图块特征向量。如图2所示,首先使用YOLO目标检测模型特有的边界框置信度 $\hat{O}_{s/t}$ 与对应尺度的主干特征图 $f_{s/t}$ 作哈达玛积^[17],拆解主干特征图中倾向于前景或背景的特征图块,即

$$u_{s/t}^{fg} = \hat{O}_{s/t} \times f_{s/t} \quad (1)$$

$$u_{s/t}^{bg} = (1 - \hat{O}_{s/t}) \times f_{s/t} \quad (2)$$

式中: $u_{s/t}^{fg}$ 为前景区域特征图; $u_{s/t}^{bg}$ 为背景区域特征图。然后,对 $u_{s/t}^{fg}$ 和 $u_{s/t}^{bg}$ 从空间位置上取平均,获得能够代表前景和背景区域的中心特征 $\bar{u}_{s/t}^{fg/bg}$,即

$$\bar{u}_{s/t}^{fg/bg} = \frac{\sum_{j=1}^{H \times W} u_{s/t}^{fg/bg,j}}{H \times W} \quad (3)$$

式中: H 和 W 分别表示特征图的高和宽; j 是特征图上像素的位置索引。

考虑物体边界的混淆像素的干扰,并非特征图

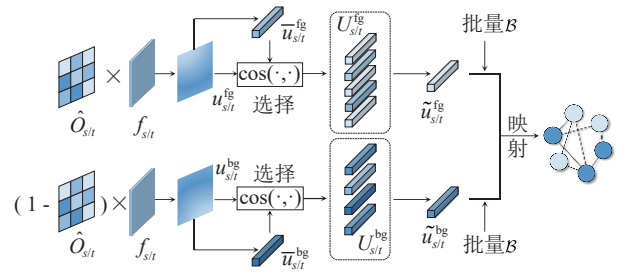


图2 粗粒度图结构的构建过程

上每个位置的像素特征都能表征前背景语义。因此,须对特征图块向量进行必要的过滤和选择。对于前背景特征图 $u_{s/t}^{fg/bg}$ 每个像素位置的像素特征向量 $u_{s/t}^{fg/bg,j}$,计算其与对应域的中心特征向量 $\bar{u}_{s/t}^{fg/bg}$ 的余弦相似度,即 $\cos(u_{s/t}^{fg/bg,j}, \bar{u}_{s/t}^{fg/bg})$ 。接着,根据相似度结果,将位于前 μ 比例的 $u_{s/t}^{fg/bg,j}$ 置于前背景区域特征的集合 $U_{s/t}^{fg/bg}$ 中。随后,分别对高相似度集合中的像素特征计算平均值,得到增强前背景图块特征向量 $\tilde{u}_{s/t}^{fg/bg}$,有

$$\tilde{u}_{s/t}^{fg/bg} = \frac{1}{|U_{s/t}^{fg/bg}|} \sum_{u_{s/t}^j \in U_{s/t}^{fg/bg}} u_{s/t}^j \quad (4)$$

式中 $| \cdot |$ 为集合的基数。

第2步,使用增强前背景图块特征向量 $\tilde{u}_{sl}^{fg/bg}$ 构建粗粒度前背景语义图。首先,将选择后的增强前背景图块特征向量 $\tilde{u}_{sl}^{fg/bg} \in \mathbb{R}^{\mathcal{B} \times d_1}$ 组合在一起,构成粗粒度节点特征向量 $\tilde{u}_{sl} = \{\tilde{u}_{sl}^{fg}, \tilde{u}_{sl}^{bg}\} \in \mathbb{R}^{2\mathcal{B} \times d_1}$ 。其中, \mathcal{B} 表示批量大小, d_1 表示节点特征向量维度。然后,将节点特征映射到图节点空间,使用图边连接节点,构建跨域粗粒度语义图结构 $\mathcal{G}_c = \langle \mathcal{U}_s, \mathcal{U}_t, \mathcal{A} \rangle$ 。其中 \mathcal{U}_s 和 \mathcal{U}_t 表示源域和目标域粗粒度图结构的节点,分别由 \tilde{u}_s 和 \tilde{u}_t 映射而来; \mathcal{A} 为跨域粗粒度图结构的边,编码了 \mathcal{U}_s 和 \mathcal{U}_t 之间的依赖关系,能适时传递对立域的信息来聚合特定的语义关系,有

$$\mathcal{A} = \begin{bmatrix} \tilde{u}_s^{fg} \mathcal{W}_e (\tilde{u}_t^{fg} \mathcal{W}_e)^T & \tilde{u}_s^{fg} \mathcal{W}_e (\tilde{u}_t^{bg} \mathcal{W}_e)^T \\ \tilde{u}_s^{bg} \mathcal{W}_e (\tilde{u}_t^{fg} \mathcal{W}_e)^T & \tilde{u}_s^{bg} \mathcal{W}_e (\tilde{u}_t^{bg} \mathcal{W}_e)^T \end{bmatrix} \quad (5)$$

式中 \mathcal{W}_e 为表征节点之间关系的可学习参数,可在训练中根据梯度更新。考虑到现有粗粒度图结构不便于直接使用图卷积来学习节点之间的关系,对原始粗粒度图结构的节点和边进行增广,有

$$\hat{\mathcal{U}} = [\mathcal{U}_s, \mathcal{U}_t]; \hat{\mathcal{A}} = \begin{bmatrix} \mathbf{0} & \mathcal{A} \\ \mathcal{A}^T & \mathbf{0} \end{bmatrix} \quad (6)$$

式中: $\hat{\mathcal{U}}$ 表示增广节点; $\hat{\mathcal{A}}$ 表示增广边。增广后的粗粒度图结构就表示为 $\hat{\mathcal{G}}_c = \langle \hat{\mathcal{U}}, \hat{\mathcal{A}} \rangle$ 。

第3步,通过图卷积网络^[18]学习前背景跨域语义拓扑关系。推理网络由 $\hat{\mathcal{U}}^{(u+1)} = \text{ReLU}(\hat{\mathcal{A}}\hat{\mathcal{U}}^{(u)}\mathcal{W}^{(u)})$ 的多个图卷积层堆叠而成,从而实现跨域粗粒度节点之间的消息传递,聚合前景和背景之间的语义知识,获得增强节点特征。其中, $\mathcal{W}^{(u)}$ 表示第 l 个图卷积层的可学习参数, $\hat{\mathcal{U}}^{(l)}$ 和 $\hat{\mathcal{U}}^{(u+1)}$ 分别表示第 l 层和第 $l+1$ 层的节点特征。在模块训练中,式(5)的可学习参数 \mathcal{W}_e 能从初始状态,逐渐学习表征跨域语义拓扑关系。为在训练中维持图节点的语义可辨别性,使用全连接层预测输出节点的前背景属性 \hat{p} ,该过程表示为

$$\hat{p} = \sigma(\text{FC}(\text{GCN}(\tilde{u}_s, \tilde{u}_t, \hat{\mathcal{G}}_c))) \quad (7)$$

式中: $\text{GCN}(\cdot, \cdot, \cdot)$ 表示图卷积网络输出的节点特征; $\text{FC}(\cdot)$ 表示全连接层的预测分数; $\sigma(\cdot)$ 表示经过Sigmoid激活函数处理后输出的概率值。然后,使用二元交叉熵损失函数来监督节点特征的语义属性预测结果,有

$$\mathcal{L}_{\text{CGPR}^2} = -[p \log(\hat{p}) + (1-p) \log(1-\hat{p})] \quad (8)$$

式中: $\mathcal{L}_{\text{CGPR}^2}$ 表示粗粒度图块关系推理损失函数; p 表示语义属性真值,0为背景,1为前景。通过CGPR²模块,源域和目标域之间的图块区域的语义拓扑关

系被跨域传递交互,从而促进前背景拓扑关系间的跨域适应。

1.3.2 细粒度语义关系推理模块

为实现细粒度的类别条件适应,本文设计细粒度语义关系推理模块。具体原理如图1上半部分所示。与CGPR²模块类似,第1步,获取类别级实例特征。为避免使用像素特征建模带来的高计算成本和内存开销,本文提出一种语义特征向量获取和表征方式。如图3所示,首先对于YOLO头部网络输出的类别分类概率图 \hat{C}_{sl} ,将它作为与头部网络的特征图 g_{sl} 合并的权重,得到对应类别的感兴趣实例的基础平均嵌入 $v_{sl}^{(k)}$,即

$$v_{sl}^{(k)} = \frac{\sum_{j=1}^{H \times W} \hat{C}_{sl}^{j, (k)} \times g_{sl}^j}{\sum_{j=1}^{H \times W} \hat{C}_{sl}^{j, (k)}} \quad (9)$$

式中: $v_{sl}^{(k)}$ 为第 k 个类别感兴趣实例的基础特征向量; $\hat{C}_{sl}^{j, (k)}$ 表示第 k 个类别在第 j 个像素位置上的分类概率; g_{sl}^j 表示输入YOLO头部网络的特征图 g_{sl} 在第 j 个像素位置上的特征值。

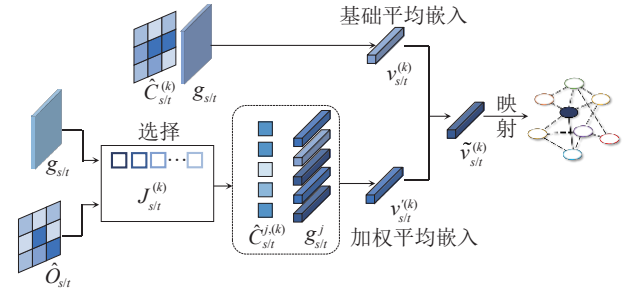


图3 细粒度图结构的构建过程

然后,根据检测头对特征图 g_{sl} 的类别预测标签,获取第 k 个类别实例向量的索引集合 $J_{sl}^{(k)}$ 。对于每种类别的索引集合,依据该位置对应的边界框置信度概率(即 $\hat{O}_{sl}^j, j \in J_{sl}^{(k)}$)从高到低进行排序,选取位于前 μ 比例的元素,组合得到对应类别的高置信度索引集合 $\hat{J}_{sl}^{(k)}$ 。将 $\hat{J}_{sl}^{(k)}$ 中每个元素对应的类别分类概率作为权重,与特征图 g_{sl} 对应位置的特征向量相乘,计算各个类别的加权平均嵌入,即

$$v_{sl}^{(k)} = \frac{\sum_{j \in \hat{J}_{sl}^{(k)}} \hat{C}_{sl}^{j, (k)} \times g_{sl}^j}{\sum_{j \in \hat{J}_{sl}^{(k)}} \hat{C}_{sl}^{j, (k)}} \quad (10)$$

式中 $v_{sl}^{(k)}$ 为源域或目标域第 k 个类别感兴趣实例的加权平均特征向量。最后,对特定类别的基础特征向量和加权特征向量相加,得到能表征不同类别感

兴趣实例的语义特征,有

$$\tilde{v}_{slt}^{(k)} = \frac{1}{2} (v_{slt}^{(k)} + v'_{slt}^{(k)}) \quad (11)$$

式中 $\tilde{v}_{slt}^{(k)} \in \mathbb{R}^{d_2}$ 为第 k 个类别感兴趣实例的语义特征向量。其中, d_2 表示语义特征向量的维度。

第2步,使用实例语义特征向量 $\tilde{v}_{slt}^{(k)}$,构建细粒度类别语义图。首先,组合各个类别的实例特征向量,得到 $\tilde{v}_{slt} = \{\tilde{v}_{slt}^{(0)}, \tilde{v}_{slt}^{(1)}, \dots, \tilde{v}_{slt}^{(K-1)}\} \in \mathbb{R}^{K \times d_2}$ 细粒度语义特征向量集。随后,将 \tilde{v}_{slt} 映射到图节点空间,构建跨域细粒度图结构 $\mathcal{G}_F = \langle \mathcal{V}_s, \mathcal{V}_t, \mathcal{E} \rangle$, 表征跨域多类别语义拓扑关系。其中, \mathcal{V}_{slt} 为源域或目标域的细粒度图节点,来源于 \tilde{v}_{slt} ; \mathcal{E} 为跨域细粒度图结构的边,负责传递同域或跨域类别之间的语义信息,使用式(12)初始化图边结构,即

$$\mathcal{E} = \begin{bmatrix} \tilde{v}_s^{(0)} \mathcal{W}_e (\tilde{v}_t^{(0)} \mathcal{W}_e)^T & \dots & \tilde{v}_s^{(0)} \mathcal{W}_e (\tilde{v}_t^{(K-1)} \mathcal{W}_e)^T \\ \vdots & & \vdots \\ \tilde{v}_s^{(K-1)} \mathcal{W}_e (\tilde{v}_t^{(0)} \mathcal{W}_e)^T & \dots & \tilde{v}_s^{(K-1)} \mathcal{W}_e (\tilde{v}_t^{(K-1)} \mathcal{W}_e)^T \end{bmatrix} \quad (12)$$

式中 \mathcal{W}_e 表示可学习参数。使用式(6)对 \mathcal{G}_F 进行增广,得到经过增广的细粒度类别语义图结构,表示为 $\hat{\mathcal{G}}_F = \langle \hat{\mathcal{V}}, \hat{\mathcal{E}} \rangle$ 。

第3步,使用图卷积网络推理并增强跨域多类别语义依赖关系。与粗粒度图推理过程相似,推理网络由 l 个图卷积层 $\hat{\mathcal{V}}^{(l+1)} = \text{ReLU}(\hat{\mathcal{E}} \hat{\mathcal{V}}^{(l)} \mathcal{W}^{(l)})$ 的堆叠而成,其中, $\mathcal{W}^{(l)}$ 表示第 l 个图卷积层的可学习参数, $\hat{\mathcal{V}}^{(l)}$ 和 $\hat{\mathcal{V}}^{(l+1)}$ 分别表示第 l 层和第 $l+1$ 层的节点特征。经过图卷积网络后,增强的类别节点特征为

$$\hat{g}_s, \hat{g}_t = \text{GCN}(g_s, g_t, \hat{\mathcal{G}}_F) \quad (13)$$

式中 $\hat{g}_s, \hat{g}_t \in \mathbb{R}^{K \times d_2}$ 为图卷积网络输出的增强类别节点特征。

为保证在细粒度图推理中类别语义信息完整,须约束各层图节点的语义可判别性。如式(14)所示,采用归一化反欧氏距离^[19](normalized inverse euclidean distance)来计算图节点间的相似度矩阵 ζ_{**} ,符号 $** \in \{ss, tt, st\}$,分别表示源域内、目标域内、源和目标域之间,即

$$\zeta_{**}^{m,n} = \frac{1}{1 + \|a^{(m)} - b^{(n)}\|_2} \quad (14)$$

式中: $\zeta_{**}^{m,n}$ 表示相似度矩阵第 m 行、第 n 列的元素; $a^{(m)}$ 和 $b^{(n)}$ 分别表示图卷积网络输出第 m 个类别和第 n 个类别的增强节点特征。理论上,相同类别的图节点相似度尽量高,不同类别的图节点相似度尽量低。因此,相似度真值应对角线元素全为1、其它位置元素全为0的对角矩阵。类别相似度约束损失 \mathcal{L}_{**} 为

$$\mathcal{L}_{**} = (1 - \zeta_{**}) \times \mathbf{M} + \max(0, \zeta_{**} + \delta) \times (\sim \mathbf{M}) \quad (15)$$

式中: \mathbf{M} 表示掩码,为单位矩阵,维度与 ζ_{**} 相同; \sim 表示矩阵按位取反; δ 为控制不同语义区分程度的参数,默认取0。因此,FGSR²模块的总损失函数 $\mathcal{L}_{\text{FGSR}^2}$ 为

$$\mathcal{L}_{\text{FGSR}^2} = \mathcal{L}_{ss} + \mathcal{L}_{tt} + \mathcal{L}_{st} \quad (16)$$

通过FGSR²模块,能实现细粒度的类别条件适应,避免了类别原型方法的次优适应。

1.3.3 粒度诱导的特征对齐模块

图推理模块构建元素间的语义拓扑关系来进行跨域适应。但在整体场景剧烈变化时,可能出现失效的风险。为此,本文提出粒度诱导的特征对齐模块,使用图边结构提取亲和性分数来调整全局特征对齐的强度。对于粗粒度图,处理前景图边 $\mathcal{A}^{\text{fg}} = \sigma(\tilde{u}_s^{\text{fg}} \mathcal{W}_e (\tilde{u}_t^{\text{fg}} \mathcal{W}_e)^T)$, 得到跨域亲和性矩阵。该亲和分数矩阵编码了不同域节点特征之间的亲和性,将每一行或列高于所在行或列的均值 $\bar{\mathcal{A}}_{m/n}^{\text{fg}}$ 的亲和值取平均,得到粗粒度层面源域和目标域之间的双向亲和权重。上述过程表示为

$$w_{s \leftrightarrow t}^C = 1 + \frac{\sum_{m=1}^B \sum_{n=1}^B \Pi(\mathcal{A}_{mn}^{\text{fg}} > \bar{\mathcal{A}}_{m/n}^{\text{fg}}) \cdot \mathcal{A}_{mn}^{\text{fg}}}{\sum_{m=1}^B \sum_{n=1}^B \Pi(\mathcal{A}_{mn}^{\text{fg}} > \bar{\mathcal{A}}_{m/n}^{\text{fg}})} \quad (17)$$

式中: $w_{s \leftrightarrow t}^C$ 指源域到目标域、目标域到源域的双向亲和权重; m, n 分别表示行、列的索引; $\Pi(\cdot)$ 为指示函数,满足括号条件时取1,反之取0。

对于细粒度图,从语义图 $\hat{\mathcal{G}}_F$ 的图边中得到跨域类别间的亲和性矩阵 $\mathcal{E} \in \mathbb{R}^{K \times K}$ 。该矩阵代表不同类别之间的亲和度关系。逐行或逐列归一化,并取对角元素求平均,得到细粒度层面源域和目标域之间的双向亲和权重。以上过程表述为

$$w_{s \leftrightarrow t}^F = \exp\left(\frac{1}{K-1} \sum_{m/n=0}^{K-1} \text{Diag}(\text{Norm}(\mathcal{E}_{m/n}))\right) \quad (18)$$

式中: $w_{s \leftrightarrow t}^F$ 为源域到目标域、目标域到源域的双向亲和权重; $\text{Diag}(\cdot)$ 表示经过矩阵对角线元素提取操作而得到的结果; $\text{Norm}(\cdot)$ 表示归一化结果,默认使用 L_1 归一化。组合粗粒度和细粒度双向亲和权重,得到粒度诱导的加权权重 $w_{s \leftrightarrow t}$, 即

$$w_{s \leftrightarrow t} = \frac{1}{2} (w_{s \leftrightarrow t}^C + w_{s \leftrightarrow t}^F) \quad (19)$$

使用 $w_{s \leftrightarrow t}$ 调整全局特征对齐的强度。因此粒度诱导的全局特征对损失 $\mathcal{L}_{\text{GIFA}}$ 为

$$\mathcal{L}_{\text{GIFA}} = \mathbb{E}_{f_s \sim \mathcal{D}'} [w_{s \rightarrow t} \cdot \log(D(f_s))] + \mathbb{E}_{f_t \sim \mathcal{D}'} [w_{t \rightarrow s} \cdot (1 - \log(D(f_t)))] \quad (20)$$

通过粒度诱导的全局对齐损失,网络能自发地根据当前批次的亲和性调整域自适应训练,从而更好地适应行驶场景的全局变化。

1.4 整体训练过程

模型训练时,每个训练批次包含了源域和目标域图像,输入到MGR²模型中,检测器首先提取输入图像的特征,并从检测头中输出预测结果,只对标注的源域预测进行监督训练。同时,粗细粒度推理模块使用骨干网络、颈部和头部输出的源域和目标域特征构建跨域图结构,并完成跨域语义拓扑关系适应。然后,特征对齐模块借助域鉴别器和梯度反转层(gradient reversal layer, GRL)形成对抗训练架构,激励检测主干输出域不变的特征。此外, GIFA模块提取跨域亲和性自适应调整对齐强度。

综上,每个批次迭代中,MGR²模型的总损失包括目标检测模型的损失、CGPR²模块的损失、FGSR²模块的损失和GIFA模块的损失,有

$$\mathcal{L}_{\text{MGR}^2} = \mathcal{L}_{\text{det}} + \lambda_1 \mathcal{L}_{\text{CGPR}^2} + \lambda_2 \mathcal{L}_{\text{FGSR}^2} + \lambda_3 \mathcal{L}_{\text{GIFA}} \quad (21)$$

式中: $\mathcal{L}_{\text{MGR}^2}$ 表示MGR²的总损失; λ_1 、 λ_2 和 λ_3 为域自适应任务的平衡因子。域自适应作为辅助任务,平衡因子一般小于1,多组实验确定最优平衡因子组合,取 $\lambda_1 = 0.1$ 、 $\lambda_2 = 0.01$ 、 $\lambda_3 = 0.5$ 。平衡因子小范围变化时,性能波动很小。

2 实验分析

2.1 实验数据集

本文使用4种公开自动驾驶数据集对MGR²算法的性能进行验证,这些数据集包括Cityscapes^[27]、Foggy Cityscapes^[28]、KITTI^[29]、BDD100k^[30]。

2.2 实验设置

在YOLO^[13-15]系列中,YOLOv5具有广泛的认可度和较高的使用频率,本文将YOLOv5作为基础视觉目标检测模型。训练时,使用在COCO^[31]预训练的权重对基础检测模型进行初始化。每个训练批次的样本由64张图像组成,包括32张来自源域的图像,32张来自目标域的图像。所有训练图像都保持高宽比不变,缩放到600×600范围之内。使用随机梯度下降优化器。该优化器的动量参数为0.937,权重衰减参数为 5×10^{-4} 。训练epoch总数为100。初始学习率为 1×10^{-2} ,使用余弦退火策略来衰减学习率,衰减倍率为0.1。若无特别说明,特征向量的选取比例 μ 均设置为0.8。

2.3 定量比较结果

本节在跨天气条件、跨相机设置、跨复杂街景和跨光照条件4种自动驾驶域自适应场景中开展实验,并与其它方法进行对比。每个实验场景中包含一个有标注源域和一个无标注目标域,源域训练集和目标域训练集用于MGR²模型的训练。训练后使用目标域的验证集来评测模型检测性能。同时,将只使用源域训练的原始YOLOv5模型在目标域验证集上的验证结果记为“基准”(Baseline),将使用带标签目标域数据训练的原始YOLOv5模型在目标域验证集上的验证结果记为“上限”(Oracle)。除非特别说明,本文所有性能结果,即各类别平均精度的平均值(mean average precision, mAP),均在IoU阈值为0.5的情况下计算得到。

2.3.1 跨天气条件自适应

将Cityscapes作为源域,Foggy Cityscapes作为目标域,来模拟不同天气条件下的域偏移。两种数据集共享8个类别。表1展示了Cityscapes到Foggy

表1 跨天气条件自适应的定量比较结果

%

方法	检测器	行人	骑手	汽车	货车	公交车	火车	摩托车	自行车	mAP
Baseline	YOLOv5	36.9	38.4	49.0	20.6	30.1	5.2	14.5	28.7	27.9
C2F ^[8]	Faster RCNN	34.0	46.9	52.1	30.8	43.2	29.9	34.7	37.4	38.6
MeGA ^[20]	Faster RCNN	37.7	49.0	52.4	25.4	49.2	46.9	34.5	39.0	41.8
MMCN ^[7]	Faster RCNN	33.4	46.8	51.9	29.1	48.4	43.2	36.0	37.4	40.8
FLDMN ^[21]	Faster RCNN	33.4	45.4	50.9	29.9	55.4	38.3	33.4	36.5	40.4
EPM ^[22]	FCOS	41.5	43.6	57.1	29.4	44.9	39.7	29.0	36.1	40.2
KTNet ^[23]	FCOS	43.0	42.7	60.0	32.3	46.6	38.4	31.2	38.2	41.5
SIGMA ^[9]	FCOS	46.9	48.4	63.7	27.1	50.7	35.9	34.7	41.4	43.5
DA-YOLO ^[24]	YOLOv3	29.5	27.7	46.1	9.1	28.2	4.5	12.7	24.8	36.1
S-DAYOLO ^[25]	YOLOv5	42.6	42.1	61.9	23.5	40.5	39.5	24.4	37.3	39.0
ConfMix ^[26]	YOLOv5	45.0	43.4	62.6	27.3	45.8	40.0	28.6	33.5	40.8
MGR²(本文)	YOLOv5	44.1	47.8	62.4	28.1	51.8	54.0	29.7	41.2	44.9
Oracle	YOLOv5	46.4	49.4	67.5	29.8	55.1	52.2	35.5	40.9	47.1

Cityscapes 的实验结果。可以看出, MGR² 相较于“基准”, mAP 提升了 +17.0%, 相较于“上限”仅有 2.2% 的差距。比目前最优的 SIGMA^[9] 方法提升 +1.4%。与单纯依靠图像翻译或对抗性特征对齐的 DA-YOLO^[24]、S-DAYOLO^[25] 等方法相比, MGR² 通过建模跨域语义关系, 实现了最优的域自适应性能。

2.3.2 跨相机配置自适应

为研究 MGR² 对不同相机配置(如分辨率、畸变、方位)引起的域偏移的域自适应能力, 本实验将 Cityscapes 作为源域, 将 KITTI 作为目标域。数据集处理遵循之前的文献[8], 共享 5 种目标类别。实验结果如表 2 所示。可以看出, MGR² 明显超过“基准”, 相较于次优方法 PCRT 获得 +1.7% 的增益, 证明所提算法在这种更复杂的跨域场景中, 仍具备良好的域自适应能力。

2.3.3 跨复杂街景自适应

理论上, 自动驾驶的目标检测模型应兼容开放世界的设置, 但数据集只包含有限的场景布局。因此不同街景布局间的适应能力对自动驾驶至关重要。本实验将 Cityscapes 作为源域, 将 BDD100k 作为目标域。遵循文献[38]中的设置, 目标域为 BDD100k 的白天子集, 共享 7 个目标类别。跨街景

表 2 跨相机配置自适应的定量比较结果 %

方法	检测器	行人	骑手	汽车	货车	火车	mAP
Baseline	YOLOv5	55.5	15.3	80.3	26.1	21.4	39.7
MLDA ^[32]	Faster RCNN	53.0	24.5	72.2	28.7	25.3	40.7
C2F ^[8]	Faster RCNN	50.4	29.7	73.6	29.7	21.6	41.0
DI-FR ^[33]	Faster RCNN	58.5	37.2	75.4	30.6	18.5	44.0
PCRT ^[34]	Faster RCNN	58.8	19.4	80.1	29.9	39.6	45.6
MGR²(本文)	YOLOv5	56.2	16.5	82.6	48.3	32.7	47.3
Oracle	YOLOv5	84.4	88.0	96.0	87.6	80.4	87.3

的实验结果如表 3 所示。可以看出, 所提 MGR² 相较于“基准”, mAP 提升 +7.8%, 与其它方法对比, 取得最优 36.0% 的 mAP, 证明 MGR² 通过多跨域语义关系的匹配, 能弥合不同复杂街景的域差距, 这在之前方法中往往被忽视。

2.3.4 跨光照条件自适应

自动驾驶汽车需要在弱光的夜晚稳定工作, 要求模型完成从白天到夜晚的域自适应任务。本实验使用 BDD100k 数据集, 将白天子集作为源域, 将夜晚子集作为目标域。与跨复杂街景的目标类别一致。从表 4 可知, 所提 MGR² 算法相对于“基准”提升 +12.3%, 超越当前最先进算法 S-DAYOLO, 与上限仅相差 3.8%, 这证明所提算法能有效缓解光照条件引起的域偏移。

表 3 跨复杂街景自适应的定量比较结果 %

方法	检测器	行人	骑手	汽车	货车	公交车	摩托车	自行车	mAP
Baseline	YOLOv5	37.4	24.6	58.9	19.1	20.0	16.3	21.2	28.2
PCRT ^[34]	Faster RCNN	39.1	30.4	55.9	15.3	17.5	21.8	30.1	30.0
UAMA ^[35]	Faster RCNN	37.3	32.9	55.8	19.0	15.4	17.6	27.0	29.3
ILLUME ^[36]	Faster RCNN	33.2	20.5	47.8	20.8	33.8	24.4	26.7	29.6
TDD ^[37]	Faster RCNN	39.6	38.9	53.9	24.1	25.5	24.5	28.8	33.6
SIGMA++ ^[38]	FCOS	47.5	30.4	65.6	21.1	26.3	17.8	27.1	33.7
S-DAYOLO ^[25]	YOLOv5	48.4	29.1	64.5	29.5	28.6	14.4	20.5	33.6
MGR²(本文)	YOLOv5	45.2	34.7	65.0	25.2	29.7	21.1	31.0	36.0
Oracle	YOLOv5	52.8	38.0	73.2	50.4	48.3	32.9	37.0	47.5

表 4 跨光照条件自适应的定量比较结果 %

方法	检测器	行人	骑手	汽车	货车	公交车	摩托车	自行车	mAP
Baseline	YOLOv5	40.4	20.2	60.7	31.4	36.6	10.2	27.5	32.4
SDA ^[39]	CenterNet	42.8	26.4	53.9	33.5	36.5	20.4	28.2	34.5
S-DAYOLO ^[25]	YOLOv5	44.8	25.1	63.9	39.4	42.6	27.5	32.5	39.4
MGR²(本文)	YOLOv5	45.8	31.0	67.7	49.9	48.7	29.7	40.0	44.7
Oracle	YOLOv5	49.6	32.3	73.6	52.8	52.3	38.5	40.1	48.5

2.4 消融实验

表 5 为 MGR² 算法的消融结果, “w/o” 表示移除模块。可以看出, 移除任何一个模块, 性能都出现不

同程度的下降。如果去掉粗粒度图块关系推理模块 CGPR², mAP 降低 2.1%。移除细粒度语义关系推理模块 FGSR², 性能下降 1.9%。没有粒度诱导的特征

对齐模块 GIFA 时, mAP 也只能达到 41.1%。以上结果说明, 粗细粒度语义关系的跨域适应能有效提高模型的跨域检测能力, 同时证明了各个模块设计的合理性。

表 5 各模块消融实验结果

方法	mAP/%	净提升/%
Baseline	27.9	
w/o CGPR ²	42.8	14.9
w/o FGSR ²	43.0	15.1
w/o GIFA	41.1	13.2
全模型	44.9	17.0
Oracle	47.1	

2.5 参数敏感度实验

粗细粒度图构建时, 以 μ 比例收集节点特征。为此, 本节以跨天气自适应为例, 分析超参数 μ 对算法性能的影响。如表 6 所示, 当 μ 取 0.8 时, mAP 最高, 达到 44.9%。 μ 过大或过小时, 性能都有所下降, 波动在 1.8% 之间。这说明过大的 μ 会引入大量带有错误伪标签的感兴趣区特征, 过小的 μ 反而会过滤掉过多足够可靠的节点特征, 二者都会使粗细粒度关系推理偏离正确的方向, 造成域自适应性能退化。

2.6 推理性能实验

本节在跨天气场景下进一步分析 MGR² 的推理性能。实验结果如表 7 所示, 实验对比了使用 Faster

表 6 超参数 μ 的敏感度实验结果

μ	0.6	0.7	0.8	0.9	1.0
mAP/%	43.3	44.0	44.9	44.1	43.1

RCNN^[10]、FCOS^[12] 检测模型的域自适应方法, 表中 FPS 表示每秒能处理的帧数。可以看出, 由于本文基于实时检测器 YOLO 开发, FPS 达到 106.61, 远超基于其它检测模型的方法。同时, 所设计的域自适应模块都具备即插即用的特性, 只在训练时执行域自适应, 推理时可以去除。因此, 与“基准”相比, 检测模型在保持高 mAP 的同时, 不会产生额外的推理耗时, 实现鲁棒且实时的视觉目标检测。

表 7 不同模型的推理性能对比

方法	检测器	FPS	mAP/%
GPA ^[40]	Faster RCNN	22.98	39.5
SIGMA ^[9]	FCOS	78.25	44.2
Baseline	YOLOv5	106.27	27.9
MGR²(本文)	YOLOv5	106.61	44.9

2.7 可视化实验

(1) 检测结果可视化

为展示 MGR² 的跨域视觉目标检测能力, 图 4 对跨天气状况、跨相机配置、跨复杂街景和跨光照条件(从左到右)自适应的实际检测结果进行可视化。可看出在域偏移的影响下, “基准”仅能检测出少数显著的目标。但 MGR² 算法能有效地抑制假阴性, 甚至能规避语义层面的歧义, 在多种极具挑战性的跨域场景中表现出足够优秀的性能。例如, 在第 1 列中, MGR² 能检测“公交车”这种狭长的困难目标。在第 3 列中, MGR² 能将“基准”错检为“货车”的“公交车”进行纠正。以上结果证明了 MGR² 的跨域检测鲁棒性。

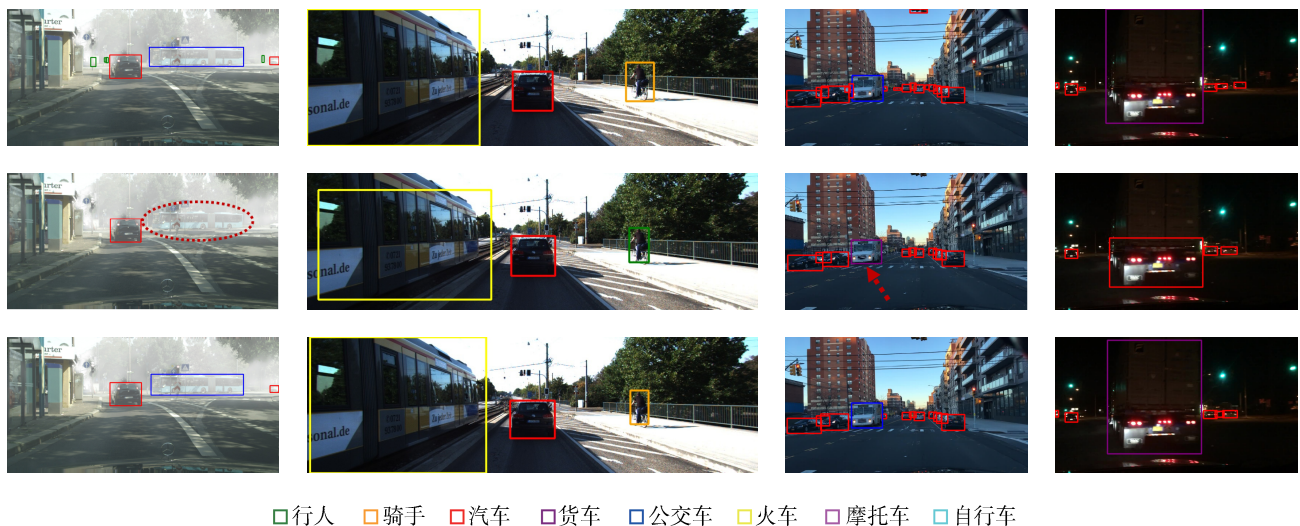


图 4 检测结果可视化实验结果(从上到下依次为真实标注、基准、本文方法的检测结果)

(2)特征分布可视化

为进一步呈现MGR²对源域和目标域特征的对齐效果,本节使用t-SNE^[19]技术可视化跨天气状况自适应下的图像级和实例级特征分布。图像级特征是由主干特征图经过全局平均池化后得到;实例级特征是由感兴趣区域池化后得到。图像级特征分布情况如图5所示。“基准”的源域和目标域图像级特征几乎完全分离,特征分布明显不同。MGR²的源域和目标域图像级特征分布基本重合,说明所提算法能够有效缩小域差异。图6为实例级特征分布可视化。对于“基准”模型,源域(浅色点)和目标域(深色点)的特征不仅没有对齐,且属于不同目标类别的特征也混杂在一起。得益于跨域语义关系推理,MGR²的域内和域间相同类别的实例特征可以实现聚拢,不同类别的特征能够得到分离,实现了很好的类别条件适应效果。

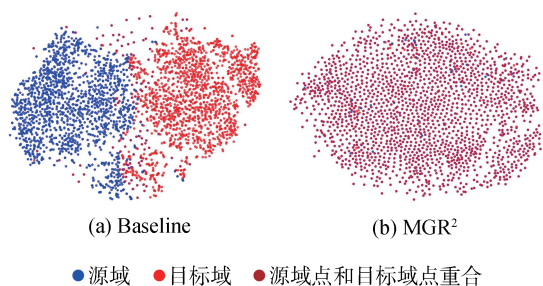


图5 图像级特征分布情况

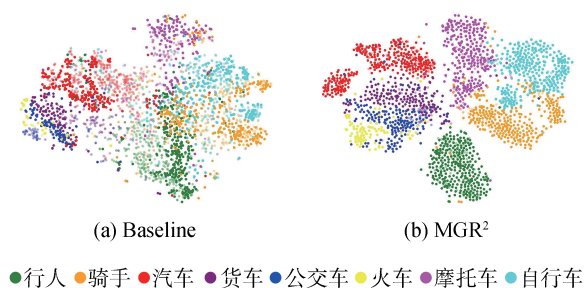


图6 实例级特征分布情况

3 结论

本文提出一种基于YOLO的多粒度关系推理域自适应目标检测方法MGR²,由粗细粒度关系推理模块和粒度诱导的全局对齐模块组成,能利用跨域多元素的拓扑关系来实现域自适应。在4种域自适应场景上进行了对比和定性实验。实验结果证明了算

法的有效性和先进性。MGR²不但优于目前的先进方法,且能满足自动驾驶汽车的实时性要求。但本文提出的模块在训练过程中会增加计算成本,并不便扩展到其它目标检测器,如何优化模块的计算量和扩展性有待后续进一步研究。

参考文献

- [1] FAHRENKROG F, REITHINGER S, GÜLSEN B, et al. European research project's contributions to a safer automated road traffic[J]. *Automotive Innovation*, 2023, 6(4): 521-530.
- [2] HE X, LV C. Towards safe autonomous driving: decision making with observation-robust reinforcement learning [J]. *Automotive Innovation*, 2023, 6(4): 509-520.
- [3] 赵东宇, 赵树恩. 基于级联YOLOv7的自动驾驶三维目标检测[J]. *汽车工程*, 2023, 45(7): 1112-1122.
ZHAO D Y, ZHAO S E. Autonomous driving 3D object detection based on cascade YOLOv7 [J]. *Automotive Engineering*, 2023, 45(7): 1112-1122.
- [4] 张炳力, 秦浩然. 基于RetinaNet及优化损失函数的夜间车辆检测方法[J]. *汽车工程*, 2021, 43(8): 1195-1202.
ZHANG B L, QIN H R. A method of vehicle detection at night based on RetinaNet and optimized loss functions [J]. *Automotive Engineering*, 2021, 43(8): 1195-1202.
- [5] CHEN Y, LI W, SAKARIDIS C, et al. Domain adaptive faster R-CNN for object detection in the wild [C]. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018: 3339-3348.
- [6] SAITO K, USHIKU Y, HARADA T, et al. Strong-weak distribution alignment for adaptive object detection [C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019: 6956-6965.
- [7] 胡杰, 徐博远, 熊宗权, 等. 基于多尺度掩码分类域自适应网络的跨域目标检测算法[J]. *汽车工程*, 2022, 44(9): 1327-1338.
HU J, XU B Y, XIONG Z Q, et al. Cross-domain object detection algorithm based on multi-scale mask classification domain adaptive network [J]. *Automotive Engineering*, 2022, 44(9): 1327-1338.
- [8] ZHENG Y, HUANG D, LIU S, et al. Cross-domain object detection through coarse-to-fine feature adaptation [C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 13766-13775.
- [9] LI W, LIU X, YUAN Y. SIGMA: semantic-complete graph matching for domain adaptive object detection [C]. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022: 5281-5290.
- [10] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [J]. *Advances in Neural Information Processing Systems*, 2015, 28.
- [11] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot mul-

- tibox detector[C]. ECCV 2016: 21–37.
- [12] TIAN Z, SHEN C, CHEN H, et al. FCOS: fully convolutional one-stage object detection[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 9627–9636.
- [13] 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.
- [14] REDMON J, FARHADI A. Yolov3: an incremental improvement [J]. arXiv preprint arXiv:1804.02767, 2018.
- [15] WANG C Y, YE H I, LIAO H Y M. Yolov9: learning what you want to learn using programmable gradient information[J]. arXiv preprint arXiv:2402.13616, 2024.
- [16] REDMON J, FARHADI A. Yolo9000: better, faster, stronger [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 7263–7271.
- [17] CHEN C, ZHENG Z, DING X, et al. Harmonizing transferability and discriminability for adapting object detectors[C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020: 8866–8875.
- [18] KIPF T N, WELLMING M. Semi-supervised classification with graph convolutional networks[J]. arXiv, 2017.
- [19] MAATEN L, HINTON G E. Visualizing data using t-SNE [J]. Journal of Machine Learning Research, 2008, 9(11).
- [20] CORDTS M, OMRAN M, RAMOS S, et al. The cityscapes dataset for semantic urban scene understanding[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 3213–3223.
- [21] SAKARIDIS C, DAI D, VAN GOOL L. Semantic foggy scene understanding with synthetic data[J]. International Journal of Computer Vision, 2018, 126(9): 973–992.
- [22] 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.
- [23] YU F, CHEN H, WANG X, et al. BDD100K: a diverse driving dataset for heterogeneous multitask learning [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 2636–2645.
- [24] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: common objects in context[C]. ECCV 2014: 740–755.
- [25] VS V, GUPTA V, OZA P, et al. MeGA-CDA: memory guided attention for category-aware unsupervised domain adaptive object detection [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 4514–4524.
- [26] 刘正发, 吴亚, 刘佩根, 等. 基于特征和标签联合分布匹配的智能驾驶跨域自适应目标检测[J]. 汽车工程, 2023, 45(11): 2082–2103.
- LIU Z F, WU Y, LIU P G, et al. Cross-domain object detection for intelligent driving based on joint distribution matching of features and labels [J]. Automotive Engineering, 2023, 45(11): 2082–2103.
- [27] HSU C C, TSAI Y H, LIN Y Y, et al. Every pixel matters: center-aware feature alignment for domain adaptive object detector [C]. ECCV 2020: 733–748.
- [28] TIAN K, ZHANG C, WANG Y, et al. Knowledge mining and transferring for domain adaptive object detection [C]. IEEE/CVF International Conference on Computer Vision (ICCV), 2021: 9113–9122.
- [29] ZHANG S, TUO H, HU J, et al. Domain adaptive YOLO for one-stage cross-domain detection [C]. Proceedings of The 13th Asian Conference on Machine Learning, 2021: 785–797.
- [30] LI G, JI Z, QU X, et al. Cross-domain object detection for autonomous driving: a stepwise domain adaptive YOLO approach [J]. IEEE Transactions on Intelligent Vehicles, 2022, 7(3): 603–615.
- [31] MATTOLIN G, ZANELLA L, RICCI E, et al. Confmix: unsupervised domain adaptation for object detection via confidence-based mixing [C]. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2023: 423–433.
- [32] LI W, LIU X, YUAN Y. SIGMA++: improved semantic-complete graph matching for domain adaptive object detection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(7): 9022–9040.
- [33] XIE R, YU F, WANG J, et al. Multi-level domain adaptive learning for cross-domain detection [C]. IEEE/CVF International Conference on Computer Vision (ICCV), 2019: 3213–3219.
- [34] YANG X, WAN S, JIN P. Domain-invariant region proposal network for cross-domain detection [C]. IEEE International Conference on Multimedia and Expo (ICME), 2020: 1–6.
- [35] WANG X, JIANG P, LI Y, et al. Progressive critical region transfer for cross-domain visual object detection [J]. IEEE Transactions on Intelligent Transportation Systems, 2024: 1–15.
- [36] CAI M, LUO M, ZHONG X, et al. Uncertainty-aware model adaptation for unsupervised cross-domain object detection [J]. arXiv preprint arXiv:2108.12612, 2021.
- [37] KHINDKAR V, ARORA C, BALASUBRAMANIAN V N, et al. To miss-attend is to misalign! residual self-attentive feature alignment for adapting object detectors [C]. IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 2022: 376–386.
- [38] HE M, WANG Y, WU J, et al. Cross domain object detection by target-perceived dual branch distillation [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022: 9560–9570.
- [39] LI G, JI Z, QU X. Stepwise domain adaptation (SDA) for object detection in autonomous vehicles using an adaptive centernet [J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(10): 17729–17743.
- [40] XU M, WANG H, NI B, et al. Cross-domain detection via graph-induced prototype alignment [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020: 12352–12361.