面向多种天气场景下目标检测的多域动态平均教师模型

刘袁缘[1)], 王超凡[1)], 王文斌[1)], 张浩宇[1,2)], 罗忠文[1)*]

[1)] (中国地质大学(武汉)计算机学院 武汉 430074)
[2)] (中国地质大学(武汉)国家地理信息系统工程技术研究中心 武汉 430074)
(luozw@cug.edu.cn)

摘 要: 现有的基于深度学习的目标检测模型中, 由于复杂天气使得现有方法的精度大幅降低. 因此为了有效地消除不同天气场景带来的域差异问题, 提出一种多域动态平均教师模型. 首先引入多域平均教师模块, 为多个不同天气场景下目标域数据生成伪标签; 然后引入基于学生网络的风格迁移模块, 解决多域任务中学生网络对不同目标域泛化能力差的问题, 可有效地减小源域与不同目标域之间的差异, 提升学生网络对不同目标域的泛化能力; 最后提出基于教师网络的动态过滤伪标签模块, 根据教师网络对不同目标域的学习效果动态地调整过滤伪标签的阈值, 提升每个目标域伪标签质量. 在 FoggyCityscapes & RainCityscapes 和 Dusk-rain & Night-rain 数据集上的实验结果表明, 所提模型分别获得了 40.3%和 31.4%的精度, 在雨天、雾天和夜晚等多种复杂天气场景下都优于对比方法.

关键词: 领域自适应; 目标检测; 多目标域; 风格迁移; 伪标签
中图法分类号: TP391.41 **DOI:** 10.3724/SP.J.1089.2024.19820

Multi-Domain Dynamic Mean Teacher for Object Detection in Complex Weather

Liu Yuanyuan[1)], Wang Chaofan[1)], Wang Wenbin[1)], Zhang Haoyu[1,2)], and Luo Zhongwen[1)*]

[1)] (*School of Computer Science, China University of Geosciences(Wuhan), Wuhan* 430074)
[2)] (*National Engineering Research Center for Geographic Information System, China University of Geosciences(Wuhan), Wuhan* 430074)

Abstract: The performance of existing deep learning-based object detection models significantly degrades due to the influence of complex weather. To effectively eliminate the problem of domain differences caused by different weather scenes, propose a multi-domain dynamic mean teacher model. First, introduce a multi-domain mean teacher module to generate pseudo-labels for target domain data in multiple different weather scenes. Then, introduce a student network-based style transfer module to solve the problem of weak generalization ability of student network to different target domains in multi-domain tasks. It can effectively reduce the difference between source domain and different target domains and improve the generalization ability of the student network to different target domains. Finally, a teacher network-based dynamic filtering pseudo-label module is used to dynamically adjust the threshold values of filtering pseudo labels according to the learning effect of the teacher network on different target domains and improve the quality of pseudo labels for each target domain. Experiments on FoggyCityscapes & RainCityscapes and Dusk-rain & Night-rain datasets show that the proposed model achieves 40.3% and 31.4% accuracy respectively, and outperforms other comparison methods in complex weather scenes such as rain, fog and night.

收稿日期: 2022-07-05; 修回日期: 2023-03-01. **基金项目**: 国家自然科学基金面上项目(62076227); 武汉市科技局基础前沿项目(2020010601012166). **刘袁缘**(1984—), 女, 博士, 副教授, 博士生导师, CCF 会员, 主要研究方向为计算机视觉、模式识别; **王超凡**(1998—), 男, 硕士研究生, 主要研究方向为目标检测; **王文斌**(1997—), 男, 硕士研究生, 主要研究方向为人脸表情识别; **张浩宇**(1997—), 男, 硕士研究生, 主要研究方向为多模态情感分析; **罗忠文**(1963—), 男, 硕士, 教授, 博士生导师, CCF 会员, 论文通信作者, 主要研究方向为机器人与机器智能.

Key words: domain adaptation; object detection; multi-target domains; style transfer; pseudo-labeling

近年来，交通目标检测与识别是人工智能和计算机视觉研究和应用的热点与难点问题. 现有的基于深度学习的方法已经取得了不错的进展[1-3]. 然而，真实交通场景中不同天气下带来的跨域差异，使得现有目标检测方法的鲁棒性大大降低. 如雾天场景下图像比较模糊，与正常天气下采集并标注的图像存在严重的数据分布差异，导致出现模型无法在雾天场景下工作的域偏移问题，使得恶劣天气条件下的图像缺少有效的标注信息，难以有效地使用传统的有监督目标检测算法.

为了解决不同场景下跨域目标检测的问题，研究者开始关注领域自适应技术[4]；其基本设置是源域为大量带标注数据，目标域为无标注数据，且源域与目标域数据分布不同. 领域自适应技术分为基于距离度量的方法和基于对抗学习的方法 2 大类. 基于距离度量的方法倾向将源域与目标域投影到一个潜在的特征空间，通过缩小源域与目标域之间的距离实现跨域目标检测[5]. Long 等[6]将原始数据映射到一个再生核希尔伯特空间，通过最大平均差(maximum mean discrepancy, MMD)[7]拉近源域与目标域之间的距离. 自从 Goodfellow 等[8]引入生成对抗网络以来，基于对抗学习的方法成为主流. Chen 等[9]分别从图像级和实例级 2 个层级上，通过对抗训练的方法学习域分类器，并通过一致性正则化进一步加强不同层级上的域分类器. Saito 等[10]提出一种基于强局部对齐和弱全局对齐的域自适应方法. 最近，基于伪标签的平均教师模型由于在半监督领域出色的表现，受到了研究者的广泛关注，其通过蒸馏增强未标记数据的一致性，在跨域目标检测领域取得了不错的结果[11-13].

尽管现有的领域自适应方法取得了一定的进展，然而目前工作大多仅关注单目标域领域自适应，即将源域的知识迁移到单一目标域的情景. 这种方法存在一定的局限性，难以应对真实交通场景下多种天气的变化，如雾天、雨天和夜晚等. 为了同时解决真实应用场景下的多目标域检测问题，现有的多目标域领域自适应方法分为 2 种：(1) 为每个目标域分别训练一个模型；(2) 将多个目标域混合在一起视为一个目标域. 第 1 种方法通常需要训练多个模型，耗时耗力，需要大量的软硬件资源；第 2 种方法直接混合不同目标域，忽略了各个目标域之间存在的分布差异，模型的泛化性能不高.

为了解决上述问题，本文提出多域动态平均教师模型(multi-domain dynamic mean teacher, MDMT)，包含 3 个模块：多域平均教师模块(multi-domain mean teacher, MMT)，基于学生网络的风格迁移模块(student network-based style transfer, SST)和基于教师网络的动态过滤伪标签模块(teacher network-based dynamic filtering pseudo-label, TDFP). (1) 以 MMT 模块为基础框架，包括学生网络和教师网络. 学生网络使用有标注的源域数据进行训练，教师网络为多个无标注的目标域数据生成伪标签，并且指导学生网络利用伪标签从目标域数据上学习新的知识. (2) 为了解决多域任务中由于源域与不同目标域互不相同，导致学生网络对不同目标域泛化能力差的问题，引入 SST 模块. 通过风格迁移得到类目标域风格图像，并输入学生网络训练，有效地提升了学生网络对不同目标域的泛化能力. (3) TDFP 模块在训练过程中自适应地评估教师网络在不同目标域上的学习效果，对于学习效果较好的类别继续保持较高阈值，保证伪标签的精准性，对于学习效果较差的类别设置较低的阈值. 通过动态调整阈值，使教师网络能够自适应地在每个目标域上都生成高质量的伪标签.

在 Cityscapes[14]和 BDD100K[15]这 2 个真实街景数据集上进行大量实验，结果表明，MDMT 模型在雨天、雾天和夜晚等多种复杂天气场景下都表现出良好的性能.

1　相关工作

1.1　目标检测

深度卷积神经网络(convolutional neural network, CNN)的发展提高了目标检测的性能. 当前，目标检测网络可以分为 2 种类型：单阶段目标检测网络，两阶段目标检测网络. Ren 等[16]提出的 Faster R-CNN 是一种具有代表性的两阶段目标检测网络，它在第 1 阶段使用区域提议网络产生粗略的候选区域，第 2 个阶段将候选区域送入感兴趣区域(region of interest, ROI)池化层进行后续准确的分类和回归. 单阶段目标检测网络直接对候选区域进行分类和回归，得到最终的目标框和类别，如 YOLO[17]，SSD[18]，其速度通常比两阶段目标检测网络快但精度稍低. MDMT 模型选用 Faster R-CNN 作为基础网络.

1.2 领域自适应

领域自适应是将模型从有标记的源域自适应到未标记的目标域[5]. 按照目标域的数量, 可以将领域自适应问题分为单目标域领域自适应和多目标域领域自适应. 针对单目标域领域自适应的方法分为基于距离度量的方法和基于对抗学习的方法 2 类. 基于距离度量的方法[6]将源域和目标域特征映射到一个共同的空间中, 通过最小化域间分布差异实现源域和目标域对齐. Carlucci 等[19]在不引入额外参数的前提下, 引导模型在不同层次上通过 MMD 对齐源域与目标域的分布; 基于对抗学习的方法利用梯度翻转实现了对抗式学习来获得域不变特征[9-10]; Liu 等[4]在像素级、图像级和实例级 3 个层次上对齐源域与目标域, 通过一致性损失优化训练过程. 目前, 多目标域领域自适应的研究较少, 并且大多专注于解决分类问题[20-21]. Wu 等[22]提出一种基于向量分解的方法(vector-decomposed disentanglement, VDD), 可以应用于多目标域领域自适应检测问题, 但在多种复杂的天气情况下分离域不变信息是困难的. 本文分析多目标域目标检测中存在的问题, 提出相应的解决方案.

1.3 教师学生模型

教师学生模型是知识蒸馏技术[23]的常用方法, 该模型包括学生网络和教师网络, 通过预测伪标签将知识从教师网络迁移到学生网络. 平均教师模型[24]是一种更精确的教师学生模型, 在各种基准上取得了良好的结果. 其中, 学生网络使用梯度下降的方式进行训练, 教师网络的权重是学生网络权重的指数移动平均值(exponential moving average, EMA). Cai 等[12]首次将平均教师模型应用到跨域检测任务中, 通过 3 种一致性正则化优化整体结构, 在单一数据上取得了良好的效果. Deng 等[13]提出一个无偏差平均教师模型(unbiased mean teacher, UMT), 通过消除教师网络和学生网络的偏差, 进一步提升跨域蒸馏的效果. 与上述方法相比, MDMT 模型能够学习多个目标域的知识, 在复杂的天气场景下正常工作.

2 MDMT 模型

为了解决复杂天气带来的多跨域问题, 本文提出 MDMT 模型, 其结构如图 1 所示. 首先, MMT 模块为不同目标域生成相应的伪标签; 然后, SST 模块生成与目标域风格一致的类目标域风格图像, 提升学生网络对不同目标域的泛化能力; 最后, TDFP 模块对不同目标域动态调整过滤伪标签的阈值, 有效地提升教师网络针对不同目标域的自适应检测能力.

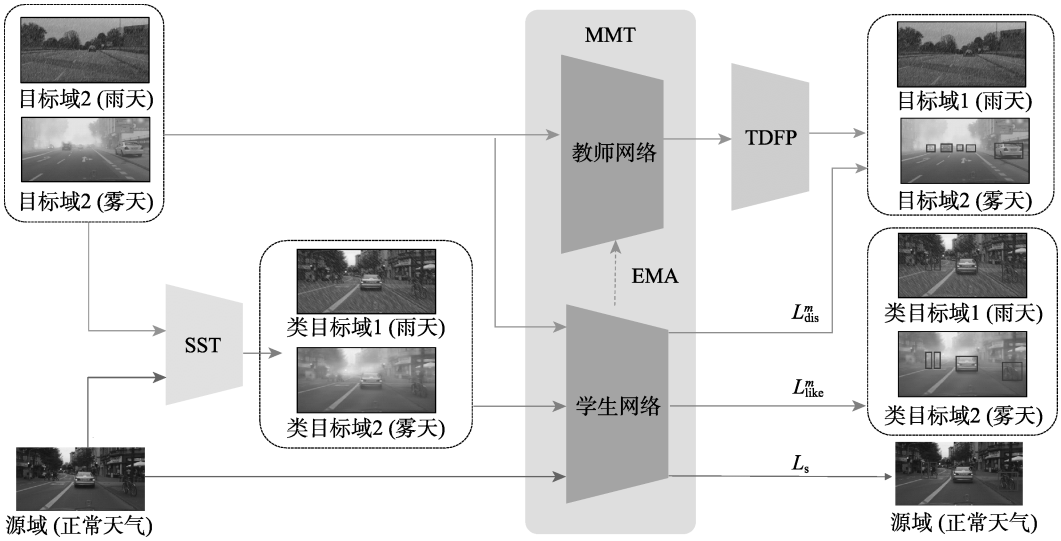

图 1 MDMT 模型结构

本文将源域表示为 $D_{\text{s}}=\{x_{\text{s}},y_{\text{s}}\}$, 其中, x_{s} 表示源域图像, 其标注 y_{s} 包含所有边界框的位置、大小和类别. 第 m 个目标域表示为 $D_{\text{t}}^{m}=\{x_{\text{t}}^{m}\}$, 其中, 目标域图像 x_{t}^{m} 不包含标注信息.

2.1 MMT 模块

利用 MMT 模块可以为无标注的目标域图像生成伪标签, 有效地解决域自适应问题中目标域缺少标注信息的问题. 如图 1 所示, MMT 模块由学生网络和教师网络这 2 个结构相同的网络组成. 学

生网络用于训练有标注的源域数据，使用随机梯度下降的方法更新网络参数. 教师网络则用于为无标注的目标域数据生成伪标签，通过伪标签指导学生网络学习无标注的目标域数据，使用学生网络的 EMA 更新网络参数. 最终使用训练好的学生网络进行不同天气场景下的目标检测与识别.

将有标注的源域数据 D_s 作为学生网络的输入，学生网络使用 Faster R-CNN 作为主干目标检测网络. 因此，直接使用 Faster R-CNN 的分类和回归损失作为学生网络的检测损失函数 L_s，可表示为

$$L_s = L_{RPN}\left(x_s, y_s\right) + L_{ROI}\left(x_s, y_s\right) \tag{1}$$

其中，L_{RPN} 表示 Faster R-CNN 中区域提议网络模块的损失；L_{ROI} 表示边界框回归和分类的预测损失，详见文献[16].

将无标注的目标域数据 D_t^m 同时输入教师网络和学生网络. 在训练过程中，教师网络首先根据 EMA 获得上述学生网络的参数，然后预测目标域数据 x_t^m 中目标边界框和类别作为伪标签 y_t^m. 为了充分地利用目标域数据，引入蒸馏损失 L_{dis}^m 指导学生网络，利用伪标签 y_t^m 进一步迁移学习目标域上的知识. 蒸馏损失 L_{dis}^m 可表示为

$$L_{dis}^m = L_{RPN}\left(x_t^m, y_t^m\right) + L_{ROI}\left(x_t^m, y_t^m\right) \tag{2}$$

其中，y_t^m 表示教师网络预测的第 m 个无标注目标域数据的伪标签.

MMT 模块的总体损失 L_{MMT} 为源域数据的检测损失与目标域数据的蒸馏损失之和，可表示为

$$L_{MMT} = L_s + \lambda L_{dis}^m \tag{3}$$

其中，λ 表示权衡参数，用来调节蒸馏损失在总体损失中的权重大小.

虽然 MMT 模块可以为无标注的目标域数据生成伪标签，但在多目标域任务中，源域和不同目标域之间的差异以及各个目标域之间的差异影响 MMT 模块的性能，因此，MDMT 模型还引入基于 SST 模块和 TDFP 模块，帮助 MMT 模块更好地适应多目标域目标检测任务.

2.2　SST 模块

在 MMT 模块中，教师网络的参数来自学生网络的 EMA. 为了提升 MMT 模块在多目标域上的性能，需要提升学生网络对不同目标域的泛化能力. 在多域任务中，源域与不同目标域之间的差异是导致学生网络泛化能力差的原因. 因此，本文提出 SST 模块，可以有效地减小源域与不同目标域之间的差异，提升学生网络对不同目标域的泛化能力.

为了实现这个目标，SST 模块采用 Choi 等[25]提出的生成对抗网络，将源域图像(正常天气)的风格迁移为目标域图像(恶劣天气，如雾天、雨天)风格，得到类目标域风格图像，如图 2 所示.

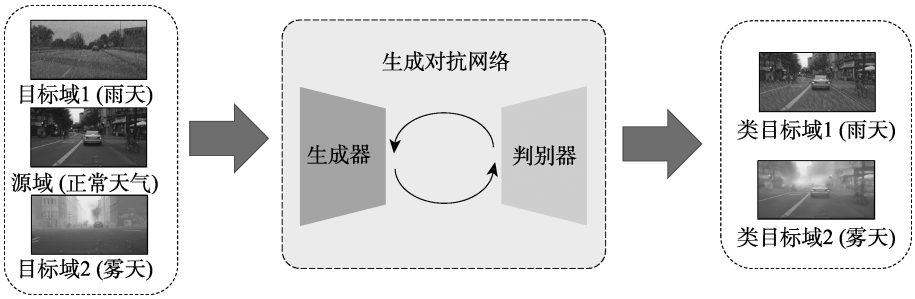

图 2　类目标域风格图像

由于图像风格转换不改变图像的内容，因此类目标域风格图像仍然可以使用其原始的标签信息，表示为 $D_{like}^m = \left\{x_{like}^m, y_s\right\}$. 其中，$x_{like}^m$ 表示与第 m 个目标域风格一致的类目标域风格图像，y_s 表示类目标域风格图像的标签. 类目标域图像作为额外的有标注数据输入学生网络，使学生网络从不同目标域风格图像上学习不同目标域风格的知识，有效地减小了源域与不同目标域之间的差异. 基于此，类似于式(1)，引入 Faster R-CNN 目标检测损失作为类目标域数据的检测损失 L_{like}^m，表示为

$$L_{like}^m = L_{RPN}\left(x_{like}^m, y_s\right) + L_{ROI}\left(x_{like}^m, y_s\right) \tag{4}$$

2.3　TDFP 模块

基于知识蒸馏的平均教师模型，为了获得高置信度的伪标签，教师网络通常需要设定一个固定的阈值来过滤掉置信度较低的伪标签[12]. 然而，对于多目标域任务，传统固定阈值的方法没有考虑不同目标域的差异，难以针对特定的目标域选

择合适的阈值，导致模型获得次优的结果. 为了解决这个问题，本文提出 TDFP 模块，可以根据某一时刻教师网络对不同目标域下不同类别的学习效果，动态地自适应调整阈值，使得教师网络在每个目标域下都能生成高质量的伪标签. TDFP 模块流程如图 3 所示.

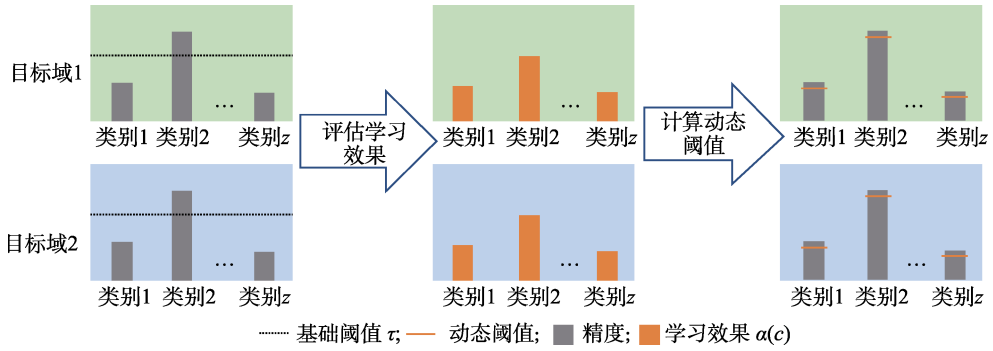

图 3 TDFP 模块流程

TDFP模块包括评估学习效果和计算动态阈值 2 个过程.

(1) 目标检测中第 c 个类别的平均精度(average precision, AP)表示为

$$\mathrm{AP}(c),\ c\in[1,z] \tag{5}$$

其中，z 表示数据集中的类别总数.

由于域自适应目标检测中类别的 AP 值普遍偏低，直接用 $\mathrm{AP}(c)$ 作为每个类别过滤伪标签的阈值会引入大量的错误伪标签[26]. 因此，本文提出归一化平均精度作为评估学习效果，公式为

$$M=\max\left(\mathrm{AP}(c)\right) \tag{6}$$

$$\alpha(c)=\frac{\mathrm{AP}(c)}{M} \tag{7}$$

其中，M 表示预测精度最高的类别的 AP; $\alpha(c)$ 表示归一化后第 c 类的预计学习效果，取值范围在 0~1.

(2) 训练过程中，在每个 epoch 学习过程中计算动态阈值. 第 c 个类别在第 k 个 epoch 时的动态阈值计算公式为

$$T_k(c)=\alpha_k(c)\times\tau \tag{8}$$

其中，$\alpha_k(c)$ 表示第 c 个类别在第 k 个 epoch 时的学习效果; τ 表示基础阈值，即传统方法中设置的高而固定的阈值，实验中将详细讨论不同的 τ 对模型性能的影响.

这种方法可以使学习效果较差的类别获得一个相对较低的阈值，从而使模型得到更多该类别的样本数据; 学习效果较好的类别继续保持较高的阈值; 同时，每个 epoch 更新一次，解决模型在不同的时刻对不同的类别学习效果不同的问题.

每个 epoch 实时更新动态阈值的方法，可以使得模型根据当前预测的结果分配相应的阈值，保证教师网络为不同目标域都可以生成高质量的伪标签.

教师网络通过蒸馏损失(式(2))和生成的伪标签，指导学生网络更有效地学习多目标域数据.

2.4 总体目标函数

MDMT 模型以端到端的方式进行训练，总体目标函数包括 3 部分: 源域数据检测损失 L_{s}，类目标域数据检测损失 L_{like}^{m} 和目标域数据蒸馏损失 L_{dis}^{m}. 其中，类目标域数据是离线生成的，然后通过端到端的方式优化所有损失. 形式上，训练阶段的总体优化目标函数为

$$L=L_{\mathrm{s}}+L_{\mathrm{like}}^{m}+\lambda L_{\mathrm{dis}}^{m} \tag{9}$$

其中，λ 表示权衡参数，用于调节蒸馏损失在总体损失中的权重大小.

3 实验与结果分析

3.1 数据集

3.1.1 Cityscapes 数据集

Cityscapes 是采集自不同城市的大规模城市街景数据集，包含 2975 张训练图片和 500 张验证图片，其中大部分为晴朗的天气场景，如图 4a 左图所示. FoggyCityscapes[27]是基于 Cityscapes 渲染生成的，它展示了雾天的街景，如图 4a 中间图所示. 为了测试在多种恶劣天气下的效果，利用 Photoshop 在 Cityscapes 的基础上构造 RainCityscapes，它展示了雨天的街景，如图 4a 右图所示. 在实验中，将

Cityscapes (正常天气)　FoggyCityscapes (雾天)　RainCityscapes (雨天)

a. Cityscapes 数据集

Daytime-sunny (晴朗天气)　Night-rain (夜晚雨天)　Dusk-rain (黄昏雨天)

b. BDD100K 数据集

图 4　数据集示例

Cityscapes 作为源域，FoggyCityscapes 与 RainCityscapes 共同作为复合目标域.

3.1.2　BDD100K 数据集

BDD100K 数据集由 10 万个驾驶视频抽取关键帧组成. 参照文献[22]的设置，将其分为 3 种不同的天气场景，如图 4b 所示: (1) Daytime-sunny, 包含日间晴朗场景; (2) Dusk-rain, 包含黄昏雨天场景; (3) Night-rain, 包含夜晚雨天场景. 在实验中，将 Daytime-sunny 作为源域，Dusk-rain 与 Night-rain 共同作为复合目标域.

3.2　实验设置

本文实验运行环境是 Intel i7-8700 CPU 以及单块 GeForce GTX1080Ti 显卡, 32 GB 运行内存, Linux 操作系统. 遵循文献[9]中的实验设置，MDMT 模型采用 Faster R-CNN 作为基本目标检测模型; 在 ImageNet[28]上预先训练的 ResNet-101[29]模型作为 Faster R-CNN 的主干网络; 在实现了带 ROI 对齐的 Faster R-CNN 之后，通过将图像短边设置为 600, 同时保持图像的横纵比来对图像进行缩放.

本文中，所有实验设置权衡参数 λ =0.001，基础置信阈值 τ =0.9. 在训练期间，学生网络采用随机梯度下降算法进行优化，初始学习率为 0.001, 学习率衰减率为 0.1; 教师网络的指数移动权重系数为 0.99. MDMT 模型是用 Pytorch[30]深度学习框架实现的.

为了评价本文方法，选取近年来的先进领域自适应目标检测方法与 MDMT 模型进行比较，包括 Source Only, DAF[9], CT[31], SCL[32], SWDA[10], ICCR[33], VDD[22]和 UMT[13]. 其中, Source Only 以 Faster R-CNN 模型仅用源域数据进行训练，然后直接对目标域数据进行测试的方法，将该方法作为所有对比方法的基准线. 为了保证实验的公平性和严谨性，本文对所有实验采用整体性能评估指标——类别平均精确度(mean average precision, mAP)，并在实验结果中展示各类别的 AP(average precision).

3.3　Cityscape 数据集实验

表 1 所示为 MDMT 模型与其他先进方法在 Cityscapes 数据集上的多域自适应评估结果. 可以看出，MDMT 模型在 mAP 上取得了最好的结果 40.3%，说明该模型可以在多种复杂的天气情况下实现精准的检测; 特别是对于在 Source Only 方法中学习效果最差的类别 train，MDMT 模型的检测精度仍有 34.9 个百分点的大幅提升，原因在于 MDMT 模型中的 TDFP 模块适当地调整了学习效果较差的类别的阈值，使学习效果较差的类别有了更多被学习的机会，最终提高了模型的检测精度. 表 2 和表 3 所示为在复合目标域上训练的模型，分别在 FoggyCityscapes 和 RainCityscapes 上进行评估的结果，可以看出, MDMT 模型在单独的目标域上仍然取得了最好的结果.

3.4　BDD100K 数据集实验

为了进一步评估模型在多种复杂天气下的检测性能，在 BDD100K 数据集上进行实验，表 4 所示为 MDMT 模型与其他方法的多域自适应检测结果. 可以看出，在图 4b 的场景下，光照条件较差且还有雨水的影响，给目标检测带来了很大的挑战，MDMT 模型在该数据集上取得了 31.4%的最优检测精度，进一步证明该模型在面对多种复杂的恶劣天气进行目标检测的有效性. 表 5 和表 6 所示为在复合目标域上训练的模型，分别在 Dusk-rain 和 Night-rain 上进行评估的结果，可以看出，MDMT 模型在单目标域上仍然取得了最好的结果.

表 1 6 种方法在 Cityscapes → FoggyCityscapes & RainCityscapes 数据集上的多域自适应评估结果 %

方法	bus	bicycle	car	motorcycle	person	rider	train	truck	mAP
Source Only	30.9	22.3	43.4	13.3	28.1	30.6	9.3	18.4	24.5
SWDA[10]	38.2	29.5	48.0	20.0	31.3	38.7	28.0	17.1	31.4
ICCR[33]	38.1	28.8	48.1	25.5	30.5	37.7	17.6	19.0	30.7
VDD[22]	43.2	32.0	44.5	23.8	31.0	39.5	27.1	25.2	33.3
UMT[13]	52.2	31.3	50.6	26.2	32.3	42.2	38.9	**30.1**	38.0
本文	**53.0**	**34.9**	**52.3**	**29.7**	**32.9**	**46.0**	**44.2**	29.5	**40.3**

注. 粗体表示最优值.

表 2 6 种方法在 Cityscapes → FoggyCityscapes 数据集上的单域自适应评估结果 %

方法	bus	bicycle	car	motorcycle	person	rider	train	truck	mAP
Source Only	33.2	26.2	42.7	16.6	29.5	34.9	6.0	24.9	26.8
SWDA[10]	39.1	33.4	44.6	21.4	33.1	43.9	32.6	19.7	33.5
ICCR[33]	41.4	32.4	48.7	29.2	32.4	43.6	27.3	22.6	34.7
VDD[22]	36.5	31.4	39.9	21.4	30.4	38.7	18.0	22.4	29.8
UMT[13]	50.0	29.4	44.7	23.4	32.0	42.5	35.2	27.9	35.6
本文	**53.0**	**36.2**	**50.4**	**30.9**	**34.1**	**48.1**	**43.9**	**29.9**	**40.8**

注. 粗体表示最优值.

表 3 6 种方法在 Cityscapes → RainCityscapes 数据集上的单域自适应评估结果 %

方法	bus	bicycle	car	motorcycle	person	rider	train	truck	mAP
Source Only	31.6	18.3	46.5	13.6	27.0	27.9	11.3	13.6	23.7
SWDA[10]	21.6	26.5	50.4	20.7	29.1	32.9	32.8	15.6	28.7
ICCR[33]	37.6	23.0	48.5	24.3	29.1	31.8	17.4	15.5	28.4
VDD[22]	48.4	32.4	51.9	28.3	31.5	40.6	43.1	28.5	38.1
UMT[13]	**57.5**	**32.4**	52.7	25.7	32.0	42.3	40.3	**33.4**	39.5
本文	55.9	32.2	**53.0**	**33.2**	**34.7**	**44.2**	**44.0**	32.1	**41.2**

注. 粗体表示最优值.

表 4 8 种方法在 Daytime-sunny → Dusk-rain & Night-rain 数据集上的多域自适应评估结果 %

方法	bus	bike	car	motor	person	rider	truck	mAP
Source Only	35.1	19.3	44.0	8.8	17.5	12.8	37.7	25.0
DAF[9]	35.9	18.3	44.2	10.1	22.0	17.9	39.9	26.9
CT[31]	31.3	15.4	41.7	8.4	19.1	15.3	32.3	23.4
SCL[32]	32.7	19.7	44.9	10.5	22.9	18.5	38.3	26.8
SWDA[10]	36.9	20.7	45.1	6.6	23.1	16.9	41.5	27.3
ICCR[33]	38.8	20.4	44.6	11.7	24.7	15.4	41.6	28.2
VDD[22]	40.9	25.7	48.9	11.3	25.9	**20.2**	43.4	31.0
本文	**42.5**	**26.9**	**49.8**	**13.7**	**26.8**	17.0	**43.4**	**31.4**

注. 粗体表示最优值.

表 5 8 种方法在 Daytime-sunny → Dusk-rain 数据集上的单域自适应评估结果 %

方法	bus	bike	car	motor	person	rider	truck	mAP
Source Only	38.6	21.5	51.7	12.0	19.7	13.6	40.9	28.3
DAF[9]	39.5	21.0	51.6	12.6	24.8	20.5	42.7	30.4
CT[31]	34.9	17.6	49.8	11.6	21.9	17.9	35.6	27.0
SCL[32]	35.7	22.3	50.7	14.8	25.3	19.9	40.1	29.8
SWDA[10]	39.2	24.6	49.6	9.2	25.5	19.3	43.7	30.1
ICCR[33]	42.0	21.9	51.5	16.5	27.2	16.8	44.1	31.4
VDD[22]	43.5	27.6	52.3	17.5	28.6	**21.1**	**46.7**	33.9
本文	**45.8**	**31.5**	**52.9**	**20.0**	**29.9**	18.6	46.5	**35.0**

注. 粗体表示最优值.

3.5　消融实验分析

在复合目标域 FoggyCityscapes 和 RainCityscapes 数据集上进行消融性实验，以验证本文方法中各个模块的有效性，表 7 所示为消融实验的结果对比.

表 6　8 种方法在 Daytime-sunny → Night-rain 数据集上的单域自适应评估结果　%

方法	bus	bike	car	motor	person	rider	truck	mAP
Source Only	23.4	13.3	31.8	1.5	10.2	10.9	23.2	16.3
DAF[9]	24.2	11.0	32.4	**4.6**	12.7	11.9	27.7	17.8
CT[31]	19.5	9.7	29.0	1.1	9.9	9.1	17.6	13.7
SCL[32]	22.9	12.8	35.8	0.9	14.8	**15.0**	30.2	18.9
SWDA[10]	29.6	10.4	37.9	0.7	15.0	11.1	**31.6**	19.5
ICCR[33]	28.4	**16.5**	33. 6	0.9	16.4	12.2	30.3	19.7
VDD[22]	27.9	15.3	35.9	2.2	13.5	10.8	30.5	19.4
本文	**31.7**	15.6	**39.3**	1.7	**16.5**	13.9	30.0	**21.2**

注. 粗体表示最优值.

表 7　MDMT 模型在 Cityscapes → FoggyCityscapes 和 RainCityscapes 数据集上的消融实验结果　%

方法	bus	bicycle	car	motorcycle	person	rider	train	truck	mAP
MMT	40.5	29.6	44.5	20.8	30.6	41.1	30.3	20.2	32.2
MMT+SST	52.3	32.1	51.6	24.3	32.4	43.6	37.1	27.7	37.6
MMT+SST+TDFP	**53.0**	**34.9**	**52.3**	**29.7**	**32.9**	**46.0**	**44.2**	**29.5**	**40.3**

注. 粗体表示最优值.

从表 7 可以看出，仅使用 MMT 模块并不能取得良好的效果，与表 1 中目前最先进结果的 mAP 相差 5.8 个百分点；在保存 MMT 模块结构不变的基础上，加入 SST 模块后，mAP 增长了 5.4 个百分点，提升明显，验证了通过类目标域风格图像，可以有效地拉近源域与不同目标域之间的距离，使得模型在不同的目标域上都有良好的性能；在 MMT 模块上加入 SST 模块的基础上，继续添加 TDFP 模块后，所有类别上的 AP 值均有提升，特别是在训练效果最差的类别 motorcycle 上，取得了比 MMT+SST 模块 5.4 个百分点的提升，验证了 TDFP 模块能够对学习效果较差的类别有更好的学习效果.

消融实验结果表明，依次增加各个模块之后，整体性能评估指标的 mAP 都有一定的提升，充分表明了 MDMT 模型的有效性和合理性.

图 5 所示为 MDMT 模型在复合目标域 FoggyCityscapes 和 RainCityscapes 数据集上设置不同基础阈值 τ 对检测精度的影响. 可以看出，当 τ =0.7 时，mAP 值较低，说明过低的阈值会引入错误伪标签，从而影响整体的性能；逐渐增大 τ 至 0.9 时，模型获得最佳性能；继续增大 τ，则 mAP 值下降，说明过于严格的阈值标准会导致模型性能下降.

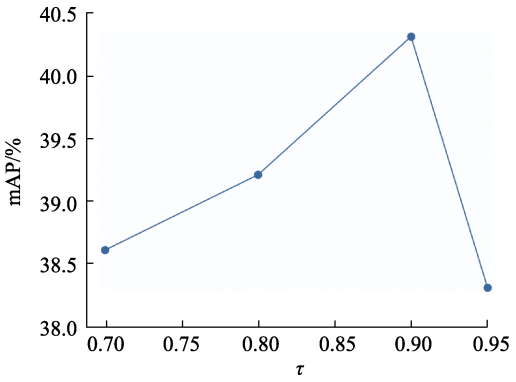

图 5　不同 τ 下 mAP 对比

3.6　复杂度分析

为了分析 MDMT 模型的计算消耗，在复合目标域 FoggyCityscapes 和 RainCityscapes 数据集上进行了复杂度实验，从检测速度和模型大小 2 个指标进行对比，结果如表 8 所示. 可以看出，与目前的领域自适应方法相比，MDMT 模型最有最小的模型容量(360.5 MB)，表明该模型具有更好的实际应用迁移能力；与目前最快的方法 SWDA 相比，MDMT 模型虽然在检测速度上慢 0.6 帧/s，但是该模型具有更少的参数，即在模型大小上减小了 43.2 MB，可能的原因是 MDMT 模型需要对每个类别计算动态阈值，该方法虽然参数量小，但带来了一定的计算复杂度，影响了检测速度. 结合表 1 中的结果，综合考虑精度、检测速度和模型大小 3 种指标，MDMT 模型仍是获得最优的结果.

表 8 6 种方法在 Cityscapes → FoggyCityscapes & RainCityscapes 数据集上的复杂度实验结果

方法	检测速度/(帧·s^{-1})	模型大小/MB
Source Only	4.7	360.5
SWDA[10]	**5.2**	403.7
ICCR[33]	5.4	429.9
VDD[22]	5.1	405.8
UMT[13]	4.8	363.3
本文	4.6	**360.5**

注. 粗体表示最优值.

3.7 检测可视化

表 9 所示为 MDMT 模型与其他先进方法在 FoggyCityscapes & RainCityscapes 数据集上的检测可视化结果. 可以看出, Source Only 方法不能很好地检测目标对象, 出现了大量的漏检; 目前最先进的 UMT 方法虽然较 Source Only 方法出现漏检的情况有所改善, 但出现了较多的误检, 雨天情况下在没有目标的区域预测出了 person, 雾天情况下出现将交通指示杆误检为 person 的情况. MDMT 模型误检和漏检的情况较少, 即使在有遮挡的情况下, 仍然可以精确地检测出目标对象.

表 10 所示为 MDMT 模型与其他先进方法在 Dusk-rain & Night-rain 数据集上的检测可视化结果. 可以看出, 黄昏雨天的场景受雨水的影响, 图像比较模糊; 夜晚雨天的场景中图像亮度非常低, 同时存在雨水的干扰, 进一步加大了检测的困难; Source Only 方法在 2 种场景下均出现了大量的误检, 如将 car 预测为 bus; 与 Source 方法相比, 目前最先进的 VDD 方法检测结果有了大幅提升, 但在目标对象有重叠时出现了大量的误检情况; 在更具有挑战性的黄昏雨天及夜晚雨天的场景下, MDMT 模型依然可以准确地定位和识别图像中存在的目标对象, 表现出了良好的检测性能.

表 9 不同方法在 FoggyCityscapes & RainCityscapes 数据集上检测可视化结果

数据集	真实标注	Source Only	UMT[13]	本文
RainCityscapes(雨天)				
FoggyCityscapes(雾天)				

表 10 不同方法在 Dusk-rain & Night-rain 数据集上检测可视化结果

数据集	真实标注	Source Only	UMT[13]	本文
Dusk-rain(黄昏雨天)				
Night-rain(夜晚雨天)				

4 结 语

为了在多种复杂天气场景下进行跨域目标检测任务, 本文提出多目标域自适应动态平均教师模型, 包括 MMT 模块,SST 模块和 TDFP 模块, 可以同时在多目标域跨域任务上取得最佳的检测结果. 在 Cityscapes 和 BDD100K 数据集上的大量实验结果表明, MDMT 模型不仅在多个目标域上表现最佳, 而且在每个单目标域上也取得最好性能. 与雾天和雨天场景下的检测结果相比, MDMT 模型在夜晚场景下的检测精度略低. 未来将引入对比学习方法, 进一步优化夜晚场景下进行的跨域目标检测.

参考文献(References):

[1] Peng Jiyao, Liu Ziyang, Feng Chuanxu, *et al*. Scale-aware bilateral feature pyramid networks for traffic sign detection[J].

Journal of Computer-Aided Design & Computer Graphics, 2022, 34(1): 133-141(in Chinese)
(彭济耀, 刘子杨, 冯传旭, 等. 面向交通标志检测的尺度感知双向特征金字塔网络[J]. 计算机辅助设计与图形学学报, 2022, 34(1): 133-141)

[2] Wang Wei, Tang Xinyao, Zhang Chaoyang, *et al*. Spatial positioning method of vehicle in cross-camera traffic scene[J]. Journal of Computer-Aided Design & Computer Graphics, 2021, 33(6): 873-882(in Chinese)
(王伟, 唐心瑶, 张朝阳, 等. 跨相机交通场景下的车辆空间定位方法[J]. 计算机辅助设计与图形学学报, 2021, 33(6): 873-882)

[3] Xie Jin, Cai Zixing, Deng Haitao, *et al*. Traffic sign classification based on deep learning of image invariant feature[J]. Journal of Computer-Aided Design & Computer Graphics, 2017, 29(4): 632-640(in Chinese)
(谢锦, 蔡自兴, 邓海涛, 等. 基于图像不变特征深度学习的交通标志分类[J]. 计算机辅助设计与图形学学报, 2017, 29(4): 632-640)

[4] Liu Y Y, Liu Z Y, Fang F, *et al*. Hierarchical domain-consistent network for cross-domain object detection[C] //Proceedings of the IEEE International Conference on Image Processing. Los Alamitos: IEEE Computer Society Press, 2021: 474-478

[5] Li Jingjing, Meng Lichao, Zhang Ke, *et al*. Review of studies on domain adaptation[J]. Computer Engineering, 2021, 47(6): 1-13(in Chinese)
(李晶晶, 孟利超, 张可, 等. 领域自适应研究综述[J]. 计算机工程, 2021, 47(6): 1-13)

[6] Long M S, Cao Y, Wang J M, *et al*. Learning transferable features with deep adaptation networks[C] //Proceedings of the 32nd International Conference on International Conference on Machine Learning. New York: ACM Press, 2015: 97-105

[7] Gretton A, Sriperumbudur B, Sejdinovic D, *et al*. Optimal kernel choice for large-scale two-sample tests[C] //Proceedings of the 25th International Conference on Neural Information Processing Systems. New York: ACM Press, 2012: 1205-1213

[8] Goodfellow I, Pouget-Abadie J, Mirza M, *et al*. Generative adversarial networks[J]. Communications of the ACM, 2020, 63(11): 139-144

[9] Chen Y H, Li W, Sakaridis C, *et al*. Domain adaptive faster R-CNN for object detection in the wild[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2018: 3339-3348

[10] 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. Los Alamitos: IEEE Computer Society Press, 2019: 6949-6958

[11] French G, Mackiewicz M, Fisher M. Self-ensembling for visual domain adaptation[OL]. [2022-07-05]. https://arxiv.org/abs/1706.05208

[12] Cai Q, Pan Y W, Ngo C W, *et al*. Exploring object relation in mean teacher for cross-domain detection[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2019: 11449-11458

[13] Deng J H, Li W, Chen Y H, *et al*. Unbiased mean teacher for cross-domain object detection[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2021: 4089-4099

[14] 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. Los Alamitos: IEEE Computer Society Press, 2016: 3213-3223

[15] Yu F, Chen H F, Wang X, *et al*. BDD100K: a diverse driving dataset for heterogeneous multitask learning[OL]. [2022-07-05]. https://arxiv.org/abs/1805.04687

[16] 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

[17] 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. Los Alamitos: IEEE Computer Society Press, 2016: 779-788

[18] Liu W, Anguelov D, Erhan D, *et al*. SSD: single shot multiBox detector[C] //Proceedings of the 14th European Conference on Computer Vision. Heidelberg: Springer, 2016: 21-37

[19] Carlucci F M, Porzi L, Caputo B, *et al*. AutoDIAL: automatic domain alignment layers[C] //Proceedings of the IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2017: 5077-5085

[20] Nguyen-Meidine L T, Belal A, Kiran M, *et al*. Unsupervised multi-target domain adaptation through knowledge distillation[C] //Proceedings of the IEEE Winter Conference on Applications of Computer Vision. Los Alamitos: IEEE Computer Society Press, 2021: 1338-1346

[21] Gholami B, Sahu P, Rudovic O, *et al*. Unsupervised multi-target domain adaptation: An information theoretic approach[J]. IEEE Transactions on Image Processing, 2020, 29: 3993-4002

[22] Wu A M, Liu R, Han Y H, *et al*. Vector-decomposed disentanglement for domain-invariant object detection[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2021: 9322-9331

[23] Hinton G, Vinyals O, Dean J. Distilling the knowledge in a neural network[OL]. [2022-07-05]. https://arxiv.org/abs/1503.02531

[24] Tarvainen A, Valpola H. Mean teachers are better role models: weight-averaged consistency targets improve semi-supervised deep learning results[C] //Proceedings of the 31st International Conference on Neural Information Processing Systems. Cambridge: MIT Press, 2017: 1195-1204

[25] Choi Y, Choi M, Kim M, *et al*. StarGAN: unified generative adversarial networks for multi-domain image-to-image translation[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2018: 8789-8797

[26] Zhang B W, Wang Y D, Hou W X, *et al*. Flexmatch: boosting semi-supervised learning with curriculum pseudo labeling[C] //Proceedings of the 35th International Conference on Neural Information Processing Systems. Cambridge: MIT Press, 2021:

18408-18419

[27] Sakaridis C, Dai D X, Van Gool L. Semantic foggy scene understanding with synthetic data[J]. International Journal of Computer Vision, 2018, 126(9): 973-992

[28] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90

[29] He K M, Zhang X Y, Ren S Q, *et al*. Deep residual learning for image recognition[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2016: 770-778

[30] Paszke A, Gross S, Chintala S, *et al*. Automatic differentiation in pyTorch[OL]. [2022-07-05]. https://openreview.net/pdf?id=BJJsrmfCZ

[31] Zhao G L, Li G B, Xu R J, *et al*. Collaborative training between region proposal localization and classification for domain adaptive object detection[C]//Proceedings of the 16th European Conference on Computer Vision. Heidelberg: Springer, 2020: 86-102

[32] Shen Z Q, Maheshwari H, Yao W C, *et al*. SCL: towards accurate domain adaptive object detection via gradient detach based stacked complementary losses[OL]. [2022-07-05]. https://arxiv.org/abs/1911.02559

[33] Xu C D, Zhao X R, Jin X, *et al*. Exploring categorical regularization for domain adaptive object detection[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2020: 11721-11730