·图形图像处理·　　文章编号：1000-3428(2023)10-0202-10　　文献标志码：A　　中图分类号：TP391.41

基于渐进式训练的多判别器域适应目标检测

李惠森[1,2]，侯　进[1,2]，党　辉[1,2]，周宇航[1,2]

（1.西南交通大学 信息科学与技术学院 智能感知智慧运维实验室，成都 611756；
2.西南交通大学 综合交通大数据应用技术国家工程实验室，成都 611756）

摘　要：基于对抗训练的域适应目标检测的研究旨在不对新数据集进行额外标注的情况下，将检测模型应用于不同的数据集。但现有算法存在目标检测和域对齐任务难以平衡的问题，且一般的单判别器结构容易局限于数据的单个模式，导致域对齐的质量下降。提出一种基于渐进式训练的多判别器域适应目标检测算法，针对传统的单判别器结构对复杂结构数据进行域对齐时的局限性，在实例级的域适应头中引入多判别器结构，使其在学习域不变信息时考虑数据的多模结构，实现质量更高、更全面的域对齐。同时，为降低引入多判别器结构而增加的模型复杂度，设计基于Dropout技术的多判别器结构，对单个判别器参数进行重复利用，并创新性地引入渐进式训练策略，即随着训练的推进逐步增大域对齐任务的比重和难度，动态平衡目标检测和域对齐任务的权重。实验结果表明，所提算法在Cityscapes到Foggy Cityscapes的域适应场景下的平均检测精度为42.9%，相比近几年该领域的新算法提高了至少0.5个百分点。

关键词：目标检测；域适应；对抗训练；多判别器；渐进式训练策略

开放科学（资源服务）标志码（OSID）：

中文引用格式：李惠森，侯进，党辉，等. 基于渐进式训练的多判别器域适应目标检测[J]. 计算机工程，2023，49(10)：202-211，221.

英文引用格式：LI H S，HOU J，DANG H，et al. Domain adaptive multi-discriminator object detection based on progressive training[J]. Computer Engineering，2023，49(10)：202-211，221.

Domain Adaptive Multi-Discriminator Object Detection Based on Progressive Training

LI Huisen[1,2]，HOU Jin[1,2]，DANG Hui[1,2]，ZHOU Yuhang[1,2]

（1.Laboratory of Intelligent Perception and Smart Operation and Maintenance，School of Information Science and Technology，Southwest Jiaotong University，Chengdu 611756，China；
2.National Engineering Laboratory of Integrated Transportation Big Data Application Technology，Southwest Jiaotong University，Chengdu 611756，China）

【Abstract】The research on domain adaptive object detection based on adversarial training aims the deployment of detection models for use with different data sets without labeling new data sets. However，existing algorithms have difficulty in balancing the tasks of object detection and domain alignment. The general single discriminator structure is limited to single mode of data，resulting in degradation of domain alignment quality. This paper proposes a multi-discriminator domain adaptive object detection algorithm based on progressive training. Considering the limitations of the traditional single-discriminator structure in the domain alignment of complex structural data，a multi-discriminator structure is introduced into the instance-level domain-adaptive head to force it to consider multiple modes of data while learning the domain invariant information，which contributes to achieving higher quality and more comprehensive domain alignment. Meanwhile，to reduce the increased model complexity，a multi-discriminator structure designed to reuse the single discriminator parameters is introduced based on dropout technology. In this paper，an innovative progressive training strategy is introduced，whereby the proportion and difficulty of domain alignment are gradually increased with the progress in training，to dynamically balance the weight of object detection and domain alignment tasks. The experimental results indicate that the mean average precision of the algorithm in domain adaptation from Cityscapes to Foggy Cityscapes was 42.9%，which is an improvement of at least 0.5 percentage points compared to algorithms of recent years.

基金项目：国家重点研发计划（2020YFB1711902）。

作者简介：李惠森（1998—），男，硕士研究生，主研方向为计算机视觉、域适应、深度学习；侯　进（通信作者），副教授、博士；党　辉、周宇航，硕士研究生。

收稿日期：2022-09-22　　**修回日期：**2022-11-24　　**E-mail：**Rubosom@163.com

【Key words】object detection；domain adaptive；adversarial training；multi-discriminator；progressive training strategy
DOI：10. 19678/j. issn. 1000-3428. 0065821

0　概述

目标检测作为计算机视觉领域的重要任务之一，在生活中已得到广泛应用，如自动驾驶[1]、视频监控[2]、机器人视觉[3]、交通检测[4]等。得益于卷积神经网络（Convolutional Neural Network，CNN）强大的特征提取能力，基于卷积神经网络搭建的目标检测模型在测试数据集上获得了显著的效果。然而基于深度学习方法的实现是以训练数据集与测试数据集服从相同分布为前提，但是在真实场景中这一条件却往往不能满足。图像的拍摄角度、背景、成像质量甚至是采集设备参数的设置等方面的变化都会导致数据分布的差异。域适应领域上将内容相似但数据分布不完全相同的两个数据集称为两个不同的域，其差异则为域偏移。上述域偏移往往导致训练好的模型在新数据集上的测试性能大幅下降[5]。收集更多训练数据，建立更全面的大型数据集或针对特定场景定制新的数据集可以在一定程度上缓解该问题。但是对新的数据集进行标注所需要的人力和时间成本十分昂贵。水下场景等特殊的检测环境，也因为水下图像颜色失真、对比度低、细节模糊等原因存在标注困难的问题[6]。

为了解决上述问题，基于无监督的域适应方法被提出。无监督域适应方法主要通过将源域数据集的信息迁移到缺乏标注信息的目标域，使模型能适应目标域数据集。如此，模型在没有对目标域数据集进行标注的情况下也能较好地完成检测任务。近年来，基于对抗的域适应方法取得了较好的成果，其主要借鉴了生成对抗网络（Generative Adversarial Network，GAN）[7]中对抗训练的思想：首先引入一个域判别器用来区分特征来自源域还是目标域。在提升域判别器分类能力的同时，通过梯度反转层训练主干网络向增大判别器损失的方向更新参数，以欺骗判别器。通过对抗训练的方式得到的主干网络所提取的特征难以区分所属域，即具备域不变性。通过训练网络提取域不变特征可以减轻域偏移对性能造成的不利影响。真实场景下获取的图像数据通常为多模式分布结构，复杂性高，但目前基于对抗的域适应方法通常采用单判别器结构的域适应头。单个判别器进行域分类时容易局限于单个模式结构而进入局部最优解，进而导致域对齐的质量下降。此外，目前该领域的新方法大多从域对齐的角度进行改进，很少从整体层面考虑如何平衡目标检测和域对齐任务之间的比重。

针对上述问题，本文设计一种基于渐进式训练的多判别器域适应目标检测方法，在不使用额外标注信息的情况下，适应新数据集并保持更高的检测精度。针对单判别器结构域适应头容易局限于单一数据模式的问题，引入多判别器结构域适应头，使特征提取器在学习域不变信息时必须考虑数据的不同模式，同时降低判别器随机误差造成的影响，避免陷入局部最优。提出基于Dropout的多判别器结构，有效避免引入多判别器结构后网络模型参数量增加的问题。此外，从平衡目标检测任务和域对齐任务之间比重的角度考虑，设计渐进式训练策略，随着训练过程的推进，逐步增大域对齐任务的比重和难度，使收敛更加平缓稳定，提升模型的性能。

1　相关工作

1.1　目标检测

基于卷积神经网络的目标检测一般分为双阶段方法和单阶段方法两类。RCNN系列模型[8-10]为典型的双阶段方法。此类算法的第1阶段一般由区域候选网络（Region Proposal Network，RPN）生成粗略的候选区域，接着在第2阶段通过区域池化（Region-of-Interest Pooling，RoI Pooling）模块结合候选区域在主干特征提取网络所提取的特征上裁剪出目标区域特征，并送入全连接层（Fully Connected Layers）进行分类回归，得到最终结果。两阶段方法一般检测精度较高，但由于需要生成候选区域，其检测速度通常相对较慢。

单阶段方法舍弃了候选框提取的步骤，直接将物体分类和目标框的边界预测统一为回归问题，简化了目标检测的流程，大幅提升了检测速度，真正意义上满足了实时目标检测的需求。典型的单阶段方法有SSD系列[11]和YOLO系列[12-14]。尤其是YOLO系列经过近几年的更新，弥补了单阶段方法追求实时性所导致的精度损失，更好地兼顾了检测精度和实时性能，使之更适用于工程实践。

虽然单阶段方法在近几年的研究中取得了不错的进展，但双阶段方法因其优越的鲁棒性和可扩展性，仍然受到学者们的青睐。本文的研究重点是提高模型的域适应能力和检测精度，而不十分关注模型的实时性，所以本文所提方法选择以双阶段模型Faster R-CNN[8]为基础检测模型。

1.2　域适应目标检测

基于无监督的域适应方法旨在训练一个具备知识迁移复用能力和适应性的模型，该模型能从具备丰富标签信息的源域中学习有用信息，并迁移到新的没有标注信息的目标域中，从而适应新的数据集。在模型迁移的过程中，两个域之间的数据分布差异是导致模型性能下降的主要原因。因此，如何缩小域间分布差异或削弱该差异造成的影响是域适应研究领域的核心问题。在早期的研究中，比较经典的

方法是将衡量域间分布差异大小的指标作为衡量模型训练过程中损失的指标，并通过最小化该损失指导网络提取域不变特征。如文献[15]提出的深度适应网络(Deep Adaptation Network，DAN)在假设条件概率分布保持不变的前提下，计算了AlexNet[16]网络后3层全连接层输出的域间分布的最大均值差异(Maximum Mean Discrepancy，MMD)，并结合最优多核选择的方法最小化MMD。文献[17]则利用最大密度差异(Maximum Density Divergence，MDD)作为域间分布差异的度量。同时，文献[17]还将最小化MDD与对抗训练结合，提出了对抗紧密匹配域适应方法，充分结合了两种方法的优势。

目前域适应目标检测领域比较主流的方法是基于对抗训练的域适应方法，其基本思想是引入域鉴别器，在训练鉴别器判别输入特征所属域的同时训练特征提取网络混淆域鉴别器，从而隐式地缩小特征的域间差异。CHEN等[18]提出的Domain Adaptive Faster(DA Faster)R-CNN在Faster R-CNN中引入图像级别和目标实例级别的域鉴别器，分别在不同层次对齐特征，缩小特征的域间差异。文献[19]结合对局部相似特征的强对齐和全局特征的弱对齐实现更加精准的特征域对齐。文献[20]提出一种图诱导原型对齐(Graph-induced Prototype Alignment，GPA)框架，通过原型表示寻求类别级别的域对齐，同时设计一个类别加权对比损失来调整训练过程以缓解类别不平衡的负面影响。文献[21]以熵信息衡量特征区域和实例样本的不确定度，以此区分良好对齐的样本和尚未完成对齐的样本并据此采取不同的域对齐策略。

虽然机器学习领域的域适应研究很早就已经开始，但直到近几年，在目标检测领域才开始提出域适应方面的问题。现有算法的性能普遍较低，该领域仍然有较大的发展空间和研究价值。

2 本文算法

目标检测的域适应问题涉及两个域，具备完整标签信息的源域和仅有数据图像的目标域。将源域数据集形式化为：$d_{\mathrm{S}}=\left\{\boldsymbol{X}_{\mathrm{S}}^{i},\boldsymbol{Y}_{\mathrm{S}}^{i}\right\}_{i=1}^{N_{\mathrm{S}}}$，其中$\boldsymbol{X}_{\mathrm{S}}^{i}\in\mathbb{R}^{H\times W\times 3}$为源域数据集$N_{\mathrm{S}}$个数据样本中的第$i$个数据，其相应的标签信息$\boldsymbol{Y}_{\mathrm{S}}^{i}\in\mathbb{R}^{k\times 5}$包含图像中$k$个目标实例中定位框的4个坐标数据以及目标所属类别。

假设目标域数据集有N_{T}个样本，则可形式化为：$d_{\mathrm{T}}=\left\{\boldsymbol{X}_{\mathrm{T}}^{j}\right\}_{j=1}^{N_{\mathrm{T}}}$。

本文的目标是利用两个域的数据和源域的标签信息训练一个泛化性能良好的检测器，并最终能在目标域数据中完成目标检测任务。

2.1 框架概述

本文算法是在DA Faster的基础上进行改进设计的，整体框架如图1所示。算法的任务主要可分为检测任务和域对齐任务两大部分。针对目标检测任务，本文分别从源域和目标域中抽取一张图像作为输入，将ResNet-50[22]作为主干网络提取两张图像的特征$\boldsymbol{F}_{\mathrm{S}}$和$\boldsymbol{F}_{\mathrm{T}}$，随后送入RPN网络中生成候选区域。区域池化模块根据RPN网络的输出对特征$\boldsymbol{F}_{\mathrm{S}}$和$\boldsymbol{F}_{\mathrm{T}}$进行裁剪和池化操作得到尺寸相同的$\kappa_{\mathrm{S}}$个源域的目标区域实例级特征$\boldsymbol{f}_1,\boldsymbol{f}_2,\cdots,\boldsymbol{f}_{\kappa_{\mathrm{S}}}$以及$\kappa_{\mathrm{T}}$个目标域的实例级特征$\boldsymbol{f}_1,\boldsymbol{f}_2,\cdots,\boldsymbol{f}_{\kappa_{\mathrm{T}}}$。只将源域的实例级特征$\boldsymbol{f}_1,\boldsymbol{f}_2,\cdots,\boldsymbol{f}_{\kappa_{\mathrm{S}}}$送入到全连接层，得到最终的检测结果。针对域适应任务，本文在主干网络之后添加图像级域适应头模块，同时在区域池化层后引入实例级域适应头模块。考虑训练数据的复杂性和多模结构，设计了多判别器结构的域适应头，以充分学习不同模式的特征分布。此外舍弃DA Faster中对模型性能提升不大的一致性正则化模块，进一步减少计算量。

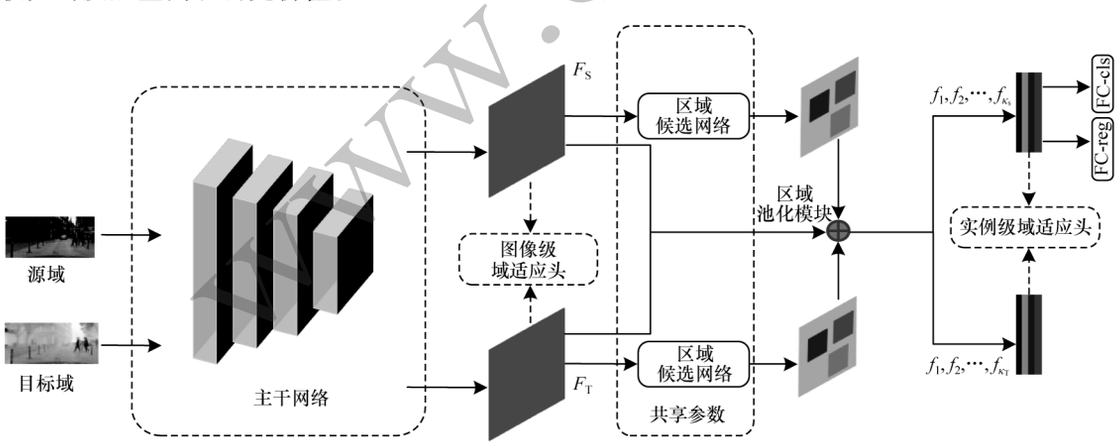

图1 本文算法的整体框架

Fig.1 Overall framework of the algorithm in this paper

2.2 域适应头

2.2.1 图像级域适应头

图像级域适应头以主干网络的输出$\boldsymbol{F}_{\mathrm{S}}$和$\boldsymbol{F}_{\mathrm{T}}$为输入，从图像层级进行域对齐。域适应头由一个域判别器和梯度反转层(Gradient Reverse Layer，GRL)[23]构成。域判别器对输入的特征进行二分类，

判别特征所属域。若域判别器得到充分训练，则其分类准确率可以作为域间特征数据分布差异的评估，准确率越低证明分布差异越小，即域不变性越强。基于该理论，本文引入对抗训练的思想，即在提升判别器分类准确率的同时训练特征提取器“欺骗”判别器，两者互相对抗优化，最终达到平衡时，特征提取器从不同域中提取的数据特征分布便可足够接近。具体操作是在域判别器之前连接一个梯度反转层。在训练网络的梯度反向传播过程中，计算域判别器损失的梯度并在判别器网络中反向传播，更新判别器参数以最小化损失。经过梯度反转层后梯度方向取反，使特征提取器参数向最大化域判别器损失的方向更新。

因为图像级域适应头的输入为卷积网络低层次的提取特征，保留了更多纹理、颜色、轮廓等细节信息，所以本文采取逐个像素对齐的策略，减轻由细节差异导致域偏移的影响，同时不破坏特征的整体语义信息。以卷积核大小为1的卷积层构造图像级域适应头的域判别器，输出一张与输入特征同尺寸的预测图，使每个像素的值为输入特征中对应区域所属域的预测值。图像级域适应头的具体结构如图2所示。假设输入特征的尺寸为 $H\times W$，D_i 为第 i 张训练图像的域标签，$D_i=0$ 表示图像来自源域，$D_i=1$ 代表图像来自目标域；$\chi_i^{(u,v)}$ 表示图像级域判别器第 i 张输出预测图中 (u,v) 位置的值，则图像级域适应头的域适应损失 L_{DAimg} 可以定义为：

$$L_{\mathrm{DAimg}}=-\sum_{i=1}^{N_S+N_T}\sum_{u=1,v=1}^{W\times H}\left[D_i\ln\chi_i^{(u,v)}+\left(1-D_i\right)\ln\left(1-\chi_i^{(u,v)}\right)\right] \tag{1}$$

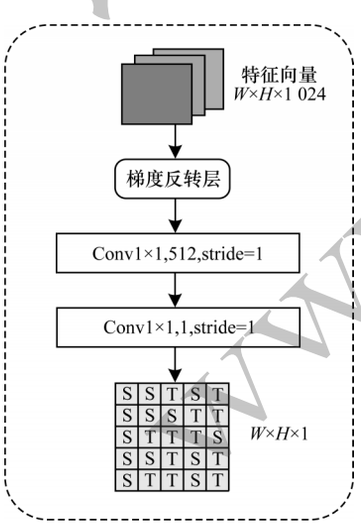

图2　图像级域适应头

Fig.2　Image level domain adaptive head

2.2.2　实例级域适应头

区域池化层的输出为网络的高层次提取特征，包含更加丰富的语义信息。所以本文将区域池化层输出的每个区域特征展开为1个一维向量作为实例级域适应头的输入，以保留每个实例的完整语义信息。实例级域适应头的结构与图像级域适应头相似，不同的是其域分类器由3层全连接层构成，具体结构如图3所示。将每个实例特征向量作为域分类器的输入，对应输出一个1×1的向量，表示对特征向量所属域的预测。假设一个批次的训练图像经过区域池化层得到 N 个实例级特征，则域分类器将输出 N 个预测结果 $\boldsymbol{\rho}=\left\{\boldsymbol{p}_1,\boldsymbol{p}_2,\cdots,\boldsymbol{p}_N\right\}$，令 D_i 为第 i 个实例特征的域标签，则实例级域适应损失 L_{DAins} 定义为：

$$L_{\mathrm{DAins}}=-\sum_{i=1}^{N}\left[D_i\ln\boldsymbol{p}_i+\left(1-D_i\right)\ln\left(1-\boldsymbol{p}_i\right)\right] \tag{2}$$

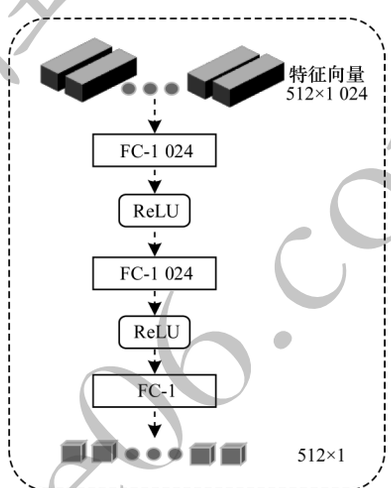

图3　实例级域分类器

Fig.3　Instance level domain classifier

2.2.3　多判别器域适应结构

文献[24]提出在图像分类的域适应问题中，一般的基于对抗训练的方法在使用单个域判别器完成域对齐任务时，往往没有考虑训练数据的复杂性和多模结构，导致特征提取器在训练中无法充分学习数据中不同模式的分布，在进行域对齐时容易忽略更多细节。受文献[25]的启发，本文在实例级域适应头中引入基于Dropout技术的多判别器域适应结构（Dropout based Multi-Discriminator architecture for Domain Adaptation，DMD^{2}A）。如图4所示，本文对图3所示的实例级域分类器进行修改，在网络的全连接层之间加入Dropout层。Dropout层在每次数据的前向传导过程中会进行一次Dropout操作，即随机隐蔽网络层中部分神经元。由于每次Dropout操作中所隐蔽的神经元是随机的，因此不同的Dropout操作后可以得到不同权重的网络。在单次迭代中，本文对实例级域分类器进行 K 次Dropout操作，得到 K 个权重不同的分类网络，如图5所示。本文重复地将上层网络输出的实例级特征输入到这 K 个不同的

网络中，即可得到K个预测结果$\{\boldsymbol{p}_1,\boldsymbol{p}_2,\cdots,\boldsymbol{p}_K\}$，则实例级域适应头的损失函数重新定义为：

$$L_{\mathrm{DAins}}=-\frac{1}{K}\sum_{u=1}^{K}\sum_{i=1}^{N}\left[D_i\ln\boldsymbol{p}_i^u+\left(1-D_i\right)\ln\left(1-\boldsymbol{p}_i^u\right)\right] \tag{3}$$

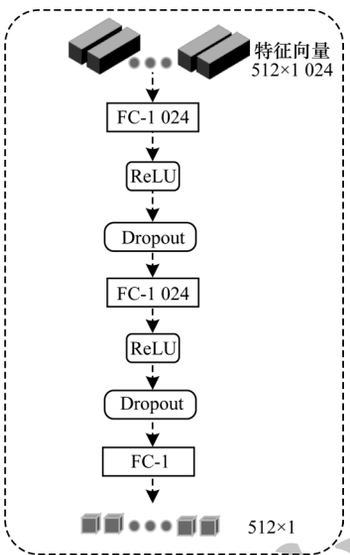

图4 基于Dropout的实例级域分类器

Fig.4 Instance level domain classifier based on Dropout

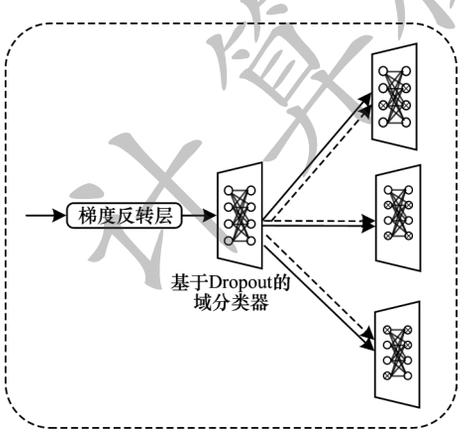

图5 基于Dropout的多判别器域适应结构

Fig.5 Multi-discriminator domain adaptation structure based on Dropout

图像级域适应头针对每个像素进行域对齐，图像级域判别器对特征值的空间分布更为敏感，引入Dropout层会降低分类器性能。因此图像级域适应头仍然采取单域判别器的结构。

2.3 渐进式训练策略

一般的域适应目标检测框架的损失函数都由检测损失L_{det}及域适应损失L_{DA}两部分组成，分别对应检测任务和域对齐任务。其中检测损失由RPN网络、最后全连接层的分类损失L_{cls}和回归损失L_{reg}共同构成，而域适应损失则分为图像级域适应损失和实例级域适应损失。网络的损失函数如式(4)~式(6)所示，参数λ一般作为平衡权重的超参数，控制目标检测损失和域适应损失的比例，平衡检测任务和域对齐任务。如何设置该参数的大小，将直接影响网络的整体性能。若λ设置过小，则域适应模块不能起到域对齐的效果，当λ减小到0时，就相当于传统的目标检测网络；若λ设置过大，则会导致网络过分注重域适应任务，所提取的特征表征能力和鉴别性太弱，从而使检测精度降低。另外，在对抗训练过程中，过大的λ参数也容易导致训练不稳定，甚至出现梯度爆炸或者梯度消失的问题。

$$L=L_{\mathrm{det}}+\lambda L_{\mathrm{DA}} \tag{4}$$

$$L_{\mathrm{det}}=L_{\mathrm{cls}}+L_{\mathrm{reg}} \tag{5}$$

$$L_{\mathrm{DA}}=L_{\mathrm{DAimg}}+L_{\mathrm{DAins}} \tag{6}$$

针对以上问题，本文提出渐进式的训练策略(Progressive Training Strategy，PTS)，将固定的λ参数修改为渐进增大的动态权重，将网络的损失重新定义为式(7)和式(8)。如式(7)所示，λ为随迭代次数τ变化的动态参数，在训练过程的前期呈线性增长，当迭代次数到达阈值τ_{TH}时λ不再增大。引入渐进式训练策略后，网络在训练前期会更关注检测任务，提高特征的表征能力；随着训练的推进，当网络的鉴别能力增加到一定程度时，便可通过逐渐增加λ参数，使网络的重心逐渐向域对齐任务偏移。如此便可更加平缓地推动训练的稳定进行，同时解决检测任务和域对齐任务难以平衡的问题。

$$\lambda(\tau)=\begin{cases}\lambda_{\mathrm{base}}\times(\tau/\tau_{\mathrm{TH}}),\tau\leqslant\tau_{\mathrm{TH}}\\ \lambda_{\mathrm{base}},\tau>\tau_{\mathrm{TH}}\end{cases} \tag{7}$$

$$L=L_{\mathrm{det}}+\lambda(\tau)L_{\mathrm{DA}} \tag{8}$$

另外，在DMD^{2}A集成判别器的数量上本文也采取渐进增加的策略。在DMD^{2}A中集成的判别器越多，对特征所属域的判别能力越强。在训练初期，网络提取特征的能力还未得到充分训练，所以网络暂时不考虑数据的多模结构，集成较少数量的判别器。随着训练进程的推进，主干网络的特征能力逐渐增强，其混淆域判别器的能力也不断提高，训练策略以此为依据动态增加判别器数量直至达到最大值N_{max}，从而在不丢失多模结构的前提下获取域不变特征，同时使训练过程更加平缓稳定。

3 实验结果与分析

3.1 实验设置

本文实验以深度学习框架PyTorch1.7搭建实验环境，程序运行的硬件环境为Intel Core I7，NVIDIA GeForce GTX1080Ti。为了适应计算机显存，所有的输入图像都经过裁剪，裁剪后的图像长边尺寸不大于1 200像素，短边尺寸不小于600像素。网络中backbone的初始权重首先使用ImageNet进行预训练，然后对整个网络进行训练，总共进行70 000迭代。前50 000次迭代设置网络学习率为0.001，最后20 000次迭代的学习率降为0.000 1。训练过程中将batch_size设置为2，分别从源域和目标域数据集中选取一张图片作为输入。另外，实验中将动量设置

为0.9，权重衰减设为0.000 5。针对渐进式训练策略，设置λ_{base}为0.25，τ_{TH}为50 000，DMD²A的最大判别器数量N_{max}为8。图6为训练过程中λ参数随迭代次数的变化曲线。

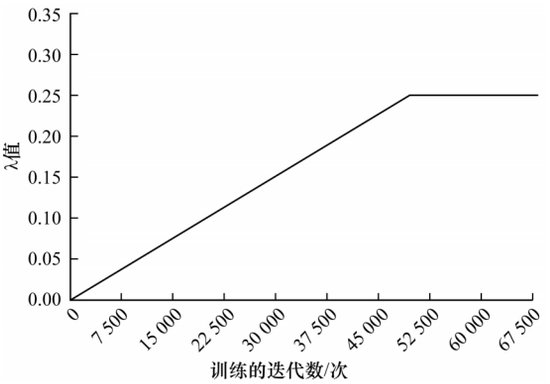

图6　λ参数随迭代次数的变化曲线

Fig.6　Variation curve of λ parameter with the iterations

3.2　数据集及场景设置

本文实验涉及3个数据集，具体如下：

1）CityScapes[26]数据集。CityScapes是一个城市场景的自动驾驶数据集，所有的数据图像都由车载摄像头拍摄。数据集包含50个城市在春、夏、秋3个季节不同时间段、不同场景、不同背景的街景图，提供2 975张训练图像和500张测试图像，其中具备实例标注信息的类别有行人、汽车、自行车等8类。

2）Foggy Cityscapes[27]数据集。Foggy Cityscapes数据集是一个合成雾化数据集，模拟真实的雾天情景。该数据集图像是在CityScapes的图像基础上添加雾噪声合成，其标注信息也直接继承CityScapes数据集得来。

3）KITTI[28]数据集。KITTI数据集是由德国卡尔斯鲁厄理工学院和丰田美国技术研究院联合制作的。数据集包含训练集7 481张图像，测试集7 518张图像，目标所含类别8个。KITTI数据集也是自动驾驶数据集，其图像场景和天气情况均与CityScapes数据集相似。但两个数据集收集图像的摄像设备不同。

为验证本文算法的有效性，实验设置了两个不同的域适应场景对网络进行测试。每个场景涉及两个数据集，分别作为源域数据集和目标域数据集。在训练过程中，从源域数据集和目标域数据集的训练图片中各随机抽取一张图像作为每次迭代的输入，同时只使用源域数据集的标注信息。在测试阶段则使用目标域的图像进行测试，并根据目标域数据集的标注信息计算模型的相关指标。两个域适应场景分别如下：

1）场景A。CityScapes到Foggy Cityscapes，设置CityScapes数据集为源域，Foggy Cityscapes数据集为目标域数据集。该场景用于模拟检测模型从良好天气到雾天的适应情况。在真实的自动驾驶应用中，目标检测系统需要面对各种不同的复杂天气情况，所以检测模型具备适应不同天气的能力至关重要。场景A即针对不同天气下图像数据的域偏移进行测试。

2）场景B。CityScapes到KITTI，在真实的应用场景中，两个数据集的图像即使拍摄的天气情况和场景相似，也会因为采像设备的不同而导致成像质量、分辨率、曝光度等方面的差异，进而呈现数据分布的差异。CityScapes数据集所使用的采像设备为COMS 2MP传感器（OnSemi AR0331型号），图片分辨率为1 024×2 048像素，帧率设为17 Hz，相机设置的基线高度为22 cm。KITTI数据集使用PointGrey Flea2录像机采集图像，设备的分辨率为1 392×512像素，帧率为10 Hz，设备的基线高度为54 cm。两个数据集所使用的采像设备及其设置参数都有所差异，使用这两个数据集设置场景B可测试模型对不同成像设备差异的适应能力。

3.3　量化结果与分析

为了验证本文提出的域适应目标检测算法的有效性，本文在不同的场景中进行实验，并与本领域其他优秀的算法进行对比。实验选取交并比（Intersection over Union，IoU）阈值为0.5的情况下计算得到的平均精度均值（mean Average Precision，mAP）作为实验的评价指标[29]。mAP为各个类别平均精确度（Average Precision，AP）的均值，具体计算公式如式（9）~式（12）所示：

$$P=\frac{T_P}{T_P+F_P}\times 100\% \tag{9}$$

$$R=\frac{T_P}{T_P+F_N}\times 100\% \tag{10}$$

$$A_{AP}=\int_0^1 P(r)\mathrm{d}r \tag{11}$$

$$m_{mAP}=\frac{1}{N}\sum_{i=1}^{N}A_{AP,i} \tag{12}$$

其中：N表示待检测的类别个数；P即模型的精确率（Precision）；R表示召回率（Recall）；T_P指准确检测的目标个数；F_P表示将背景误判为目标的数目；F_N则指未检测出的目标或判别为目标但分类错误的实例个数。具体的实验结果如下：

1）场景A。在该场景下，两个数据集图像中的目标位置以及目标类别都相同，但是Foggy Cityscapes数据集图像经过雾化处理后在视觉上有较大的差异，且部分目标变得模糊不清，造成模型性能严重下降。在源域数据集训练好的Faster R-CNN[8]在Foggy Cityscapes数据集中测试的mAP仅为26.9%。表1为场景A的测试结果，表中加粗数字表示该组数据最大值。相较于没有引入域适应方法而只使用源域数据集进行训练的Faster R-CNN，本文提出的基于域适应的目测检测算法的mAP有很大幅度提升，超出Faster R-CNN算法16个百分点。为进一步证明本文算法的优越性能，本文与该领域其

他经典算法进行对比，包括DA Faster[18]、SWDA[19]、GPA[20]、UaDAN[21]、EPM[30]、SFA[31]、DDF[32]。在场景A中，本文提出的算法测试的mAP比其他算法高出至少0.5个百分点；相较于其借鉴的DA Faster，本文算法提升了10.9个百分点。由于场景A中源域和目标域数据集具有相同的目标类别，为更全面地评估所提算法的综合性能，实验同时测试了数据集中各个类别的AP并进行对比。可以看到除了person、rider、car和bicycle类别以外，其他类别的AP均达到了最高水平。

表1 CityScapes到Foggy Cityscapes的域适应结果

Table1 Result of adaptation from Cityscapes to Foggy Cityscapes %

算法	AP								mAP
	person	rider	car	truck	bus	train	motorcycle	bicycle	
Faster R-CNN[8](Source only)算法	26.9	38.2	35.6	18.3	32.4	9.6	25.8	28.6	26.9
DA Faster[18]算法	29.2	40.4	43.4	19.7	38.3	28.5	23.7	32.7	32.0
SWDA[19]算法	29.9	42.3	43.5	24.5	36.2	32.6	30.0	35.3	34.3
GPA[20]算法	32.9	46.7	54.1	24.7	45.7	41.1	32.4	38.7	39.5
UaDAN[21]算法	36.5	46.1	53.6	28.9	49.4	42.7	32.3	38.9	41.1
EPM[30]算法	44.2	46.6	58.5	24.8	45.2	29.1	28.6	**44.0**	39.0
SFA[31]算法	**46.3**	**48.6**	**62.6**	22.1	43.4	24.3	29.9	43.1	41.3
DDF[32]算法	37.6	45.5	56.1	30.7	50.4	47.0	31.1	39.8	42.3
本文算法	37.4	45.2	54.7	**30.9**	**52.1**	**50.6**	**35.0**	38.8	**42.9**

2）场景B。在该场景下，本文只考虑源域和目标域数据集都包含且具备标注信息的car类别。同样在场景B设置下测试car类别的AP，并与其他经典算法进行对比，量化结果如表2所示。与场景A相比，该场景下设置的源域和目标域两个数据集图像的场景更相似，数据分布的域偏移更小，且只考虑单个类别的AP，故Faster R-CNN测试结果更好。而本文算法在训练的反馈阶段，特征提取器的参数更新的同时受多个不同判别器的损失指导，很大程度上削弱了单个判别器的误差和局限性造成的影响，可以更加稳健地提取域不变特征。所以面对场景B中2个数据集数据分布的微小差异时，本文算法的优势更为明显，实验测试的结果比对比算法高出至少0.8个百分点。

表2 CityScapes到KITTI的域适应结果

Table2 Result of domain adaptation from Cityscapes to KITTI %

算法	AP for car
Faster R-CNN[8](Source only)算法	53.5
DA Faster[18]算法	64.1
RPDA[33]算法	61.3
DDF[32]算法	75.0
本文算法	75.8

3.4 P-R曲线

在目标检测任务中，精确率和召回率两个指标往往是矛盾的存在：设置更高的置信度阈值，能提高模型的精确率，但也会增加目标被漏检的风险，导致召回率降低；相反，如果降低置信度阈值，将会有更多目标被划分为正例，召回率会相应提高，但也会损失精度。单独以精确率或者召回率作为衡量指标都不能全面地评估模型的性能。P-R曲线上的各点表示模型在不同置信度下精确率和召回率的关系，反映了模型对两个指标的平衡。

本节给出了本文算法在场景A下的P-R曲线，更直观地体现了算法改进的优越性。如图7所示，没有加入域适应结构的Faster RCNN实验得到的P-R曲线所展现的性能较差。DA Faster引入了域适应头，得到的P-R曲线包含了更大的曲线下面积（Area Under Curve，AUC）。而本文算法相较于DA Faster算法，其P-R曲线的AUC显著提升，完全包裹住了Faster R-CNN和DA Faster的曲线。

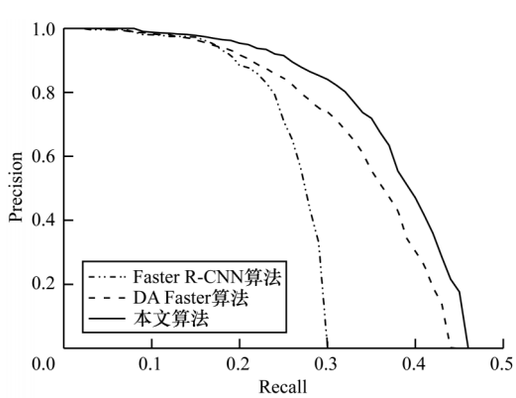

图7 不同算法的P-R曲线

Fig.7 P-R curves of different algorithms

3.5 定性结果与分析

为更直观地体现本文算法的性能，本文随机抽取了图像样本进行测试，实验的定性结果如图8所示（彩色效果见《计算机工程》官网HTML版），其中每一行为一张样本图像的结果，图8（e）为真实标签

在Foggy Cityscapes数据集图像上的实验结果；图8(a)为未引入域适应方法的Faster R-CNN的实验结果，实验中仅使用源域数据集图像和标注信息进行训练并在同一数据集上进行测试。可以看到经典的目标检测算法得到充分训练后，在同一数据上进行检测的结果已十分精确。图8(b)图像则是使用图8(a)图像实验中训练得到的相同模型在目标域数据集上测试的结果。可以看到，把在源域数据集中训练好的模型放到目标域数据集中进行测试，预测结果明显下降，出现大量漏检和误检的现象。如在图8第1行的样本图像中，虽然在近处的行人和车辆仍然可以检测到，但是道路远处的车辆出现大量漏检。而基于域适应方法的DA Faster在没有借助目标域标注信息的情况下，在很大程度上削减了域偏移带来的负面影响，在源域数据集上训练好的模型仍然能很好地迁移到目标域数据集，并在目标域数据集保持良好的检测性能，其测试结果如图8(c)所示。但是，针对雾天情况下道路远处一些异常模糊的目标，DA Faster仍存在漏检和误检的情况，如第1行样本图像中，虽然DA Faster能检测出部分道路远端的模糊目标，但是在尽头的小目标车辆依然无法检测，且道路旁边有部分重叠的行人目标也没有完全检测出来。图8(d)为本文算法的预测结果，可以看到算法能准确检测道路尽头异常模糊的车辆目标以及重叠的行人小目标，表现出优越的检测性能。

图8　不同算法的检测结果对比

Fig.8　Detection results comparison of differernt algorithms

3.6　消融实验

为进一步验证算法中各个模块的作用，本文基于场景A设计了消融实验。本文将多判别器域适应结构(DMD^{2}A)引入图像级判别器中可得到图像级多判别器结构域适应头($MDDA_{img}$)，引入实例级判别器得到实例级多判别器结构域适应头($MDDA_{ins}$)。将$MDDA_{img}$、$MDDA_{ins}$以及渐进式训练策略(PTS)3个改进模块组成不同的组合加入DA Faster中，同时去掉一致性正则化模块，其他部分与DA Faster保持一致，将DA Faster作为对照组。实验共设置5组对照组合，结果如表3所示，其中“√”表示使用该模块，“—”表示不使用该模块。对比DA Faster，单独在模型的图像级和实例级域适应头中引入多判别器结构能较大程度地提高算法性能。在引入多判别器域适应结构的基础上使用渐进式训练策略可以进一步提升模型性能，mAP均超过了40%。而同时引入实例级多判别器结构域适应头和渐进式训练策略可以得到5个方案中的最佳结果，mAP为42.9%。值得注意的是，如果将图像级和实例级域适应模块都引入多判别器结构，实验结果反而有所降低。本文认为导致模型性能下降的原因是图像级域分类器针对特征的每个像素进行分类，对特征的空间分布较为敏感，Dropout操作会直接破坏分类结果，进而影响域对齐效果。

表3　消融实验结果

Table 3　Result of ablation experiment　%

算法	PTS	MDDAimg	MDDAins	mAP
DA Faster[18]	—	—	—	32.0
本文算法	—	√	—	37.1
	—	—	√	40.8
	√	√	—	40.5
	√	√	√	41.4
	√	—	√	42.9

3.7 判别器数量对模型性能的影响

为进一步验证多判别器域适应结构对模型性能的提升，本文设计了对比实验探究判别器数量对模型性能的影响。实验基于场景A设置，改变域判别器的最大值N_{max}进行对比实验。同时，实验设置了对照组。在对照组的模型训练中不加入渐进式训练策略，即每次实验模型的判别器数量为固定值，实验结果如图9所示。相较于单判别器结构的模型，多判别器结构模型性能有大幅提高，且随着N_{max}的增大而上升，当N_{max}为8时模型性能达到最佳，N_{max}进一步增大会导致模型性能有所下降。而对照组也呈现相同趋势，当判别器数量在整个训练过程固定时，增加其数量同样可以提升模型性能。两组实验充分验证了多判别器结构对模型性能的提升，且实验表明加入渐进式训练的模型整体比对照组的更好，这进一步验证了渐进式训练策略能较好地提升模型性能。

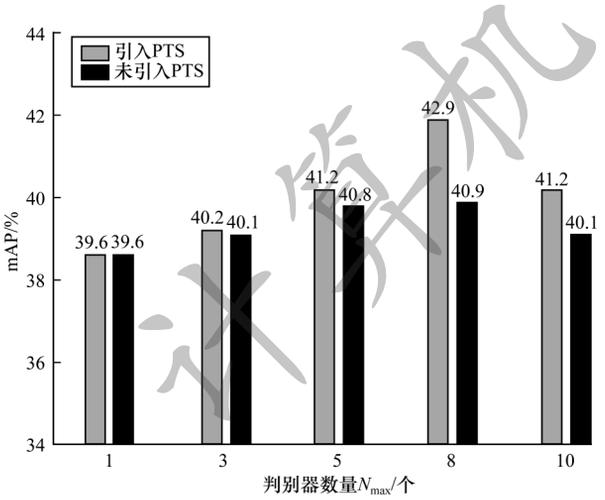

图9 判别器数量对模型性能的影响

Fig.9 Influence of the number of discriminators on model performance

3.8 λ参数对模型性能的影响

本节进一步探究参数λ对模型性能的影响。实验基于场景A设置，DMD^{2}A中的判别器数量固定为8个，改变参数λ的大小进行对比实验。另外，设置对照组实验，对照组在模型中加入渐进式训练策略，将λ_{base}设为0.25，DMD^{2}A的最大判别器数量N_{max}设为8，实验结果如图10所示。当参数设置固定时，λ设为0.1时实验结果最好，mAP为41.8%，当λ大于或者小于0.1时，模型的性能都会下降；当参数为1时，模型的性能骤降。而加入渐进式训练策略后（即图10中“contrast group”对应的实验数据），模型的性能超过了其他固定参数设置的模型，这进一步验证了渐进式训练策略的有效性。

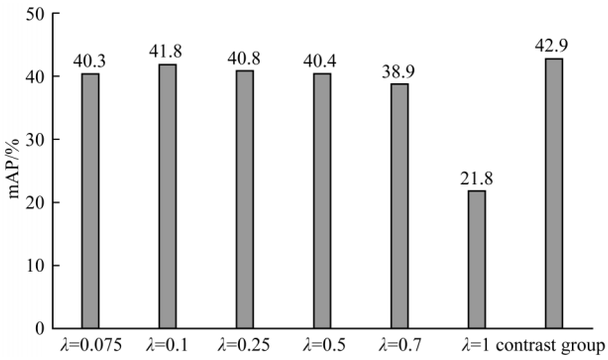

图10 λ参数对模型性能的影响

Fig.10 Influence of λ parameter on model performance

3.9 模型复杂度

本节探究引入多判别器域适应结构对模型复杂度的影响。由于针对模型的改进只会改变实例级判别器的结构，所以本文将实例级判别器的参数量N_{max}、模块大小以及浮点运算数作为模型复杂度的衡量指标。同时，本文以传统的方式使用多个不同参数的判别器组合成多判别器结构（Multi-Discriminator Domain Adaptation，MDDA），并以MDDA替换DMD^{2}A后得到的模型为对比模型，结果如表4所示。可以看到，引入DMD^{2}A结构后，衡量模型复杂度的各项指标不会随着判别器数量的增加而上升；而使用MDDA结构的模块随着判别器数量的增加，其参数量和模型大小会急剧上升。实验结果表明，使用Dropout技术进行改进的多判别域适应结构的引入在提升模型性能的同时不会增加模型的复杂度，相对于传统的多个判别器简单叠加的方法具备明显的优势。

表4 实例级域适应头模块的复杂度

Table 4 Complexity of instance level domain adaptation header module

结构	判别器数量/个	参数量/10^6	模型大小/10^6	浮点运算数/(10^9 frame·s^{-1})
单判别器	1	3.14	0.01	1.61
DMD^{2}A	3	3.14	0.01	1.61
	5			
	8			
MDDA	3	9.45	0.02	4.83
	5	15.74	0.04	8.06
	8	25.19	0.06	12.89

4 结束语

本文提出一种基于渐进式训练的多判别器域适应目标检测算法。该算法在Faster R-CNN模型的基础上进行改进，分别针对图像级别和实例级别采用不同的域对齐方法。针对一般的基于对抗的域适应方法无法充分考虑数据的复杂性和多模分布结构的

问题，引入基于Dropout的多判别器域适应头结构，完成更细致、更全面的域对齐。另外，本文从如何平衡检测任务和域对齐任务的角度出发，创新性地提出渐进式训练策略，即随着训练的推进逐渐增大域对齐任务的比重和难度，使模型的收敛更加平滑稳定，进一步提高模型性能。但是，针对参数λ的设置，本文仅采用了简单的线性增长方式，下一步将探究合适的增长策略，探索更高效的渐进式训练策略。

参考文献

[1] LI P L，CHEN X Z，SHEN S J. Stereo R-CNN based 3D object detection for autonomous driving[C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition . Washington D. C. ，USA：IEEE Press，2020：7636-7644.

[2] HATTORI H，LEE N，BODDETI V N，et al. Synthesizing a scene-specific pedestrian detector and pose estimator for static video surveillance[J]. International Journal of Computer Vision，2018，126(9)：1027-1044.

[3] SCALISE R，THOMASON J，BISK Y，et al. Improving robot success detection using static object data[C]//Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems. Washington D. C. ，USA：IEEE Press，2020：4229-4235.

[4] 邹慧海，侯进. 改进SSD算法的道路小目标检测研究[J]. 计算机工程，2022，48(5)：281-288.
ZOU H H，HOU J. Research on road small target detection with improved SSD algorithm[J]. Computer Engineering，2022，48(5)：281-288. (in Chinese)

[5] GOPALAN R，LI R N，CHELLAPPA R. Domain adaptation for object recognition：an unsupervised approach[C]//Proceedings of International Conference on Computer Vision. Washington D. C. ，USA：IEEE Press，2012：999-1006.

[6] 李莉，王新强，银珊. 基于衰减补偿与直方图拉伸的水下图像增强算法[J]. 计算机工程，2022，48(6)：222-227.
LI L，WANG X Q，YIN S. Underwater image enhancement algorithm based on attenuation compensation and histogram stretching[J]. Computer Engineering，2022，48(6)：222-227 (in Chinese)

[7] GOODFELLOW I J，POUGET-ABADIE J，MIRZA M，et al. Generative adversarial nets[C]//Proceedings of the 27th International Conference on Neural Information Processing Systems. New York，USA：ACM Press，2014：2672-2680.

[8] 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.

[9] CAI Z W，VASCONCELOS N. Cascade R-CNN：delving into high quality object detection[C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C. ，USA：IEEE Press，2018：6154-6162.

[10] HE K M，GKIOXARI G，DOLLÁR P，et al. Mask R-CNN[C]//Proceedings of IEEE International Conference on Computer Vision. Washington D. C. ，USA：IEEE Press，2017：2980-2988.

[11] LIU W，ANGUELOV D，ERHAN D，et al. SSD：single shot MultiBox detector[C]//Proceedings of European Conference on Computer Vision. Berlin，Germany：Springer，2016：21-37.

[12] REDMON J，FARHADI A. YOLO9000：better，faster，stronger[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. ，USA：IEEE Press，2017：6517-6525.

[13] REDMON J，FARHADI A. YOLOv3：an incremental improvement[EB/OL]. [2022-08-08]. https://arxiv. org/abs/1804. 02767.

[14] BOCHKOVSKIY A，WANG C Y，LIAO H Y M. YOLOv4：optimal speed and accuracy of object detection[EB/OL]. [2022-08-08]. https://arxiv. org/abs/2004. 10934.

[15] 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 Machine Learning. New York，USA：ACM Press，2015：97-105.

[16] KRIZHEVSKY A，SUTSKEVER I，HINTON G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM，2017，60(6)：84-90.

[17] LI J J，CHEN E P，DING Z M，et al. Maximum density divergence for domain adaptation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2021，43(11)：3918-3930.

[18] CHEN Y H，LI W，SAKARIDIS C，et al. Domain adaptive faster R-CNN for object detection in the wild[C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C. ，USA：IEEE Press，2018：3339-3348.

[19] SAITO K，USHIKU Y，HARADA T，et al. Strong-weak distribution alignment for adaptive object detection[C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C. ，USA：IEEE Press，2020：6949-6958.

[20] XU M H，WANG H，NI B B，et al. Cross-domain detection via graph-induced prototype alignment[C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C. ，USA：IEEE Press，2020：12352-12361.

[21] GUAN D Y，HUANG J X，XIAO A R，et al. Uncertainty-aware unsupervised domain adaptation in object detection[J]. IEEE Transactions on Multimedia，2022，24：2502-2514.

[22] HE K M，ZHANG X Y，REN S Q，et al. Identity mappings in deep residual networks[C]//Proceedings of European Conference on Computer Vision. Berlin，Germany：Springer，2016：630-645.

[23] GANIN Y，LEMPITSKY V. Unsupervised domain adaptation by backpropagation[EB/OL]. [2022-08-08]. https://arxiv. org/abs/1409. 7495.

[24] PEI Z Y，CAO Z J，LONG M S，et al. Multi-adversarial domain adaptation[C]//Proceedings of AAAI Conference on Artificial Intelligence. Menlo Park，USA：AAAI Press，2018：3934-3941.

（下转第221页）

（上接第211页）

[25] KURMI V K, BAJAJ V, SUBRAMANIAN V K, et al. Curriculum based dropout discriminator for domain adaptation[EB/OL]. [2022-08-08]. https://arxiv. org/abs/1907. 10628.

[26] CORDTS M, OMRAN M, RAMOS S, et al. The cityscapes dataset for semantic urban scene understanding [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA. IEEE Press, 2016: 3213-3223.

[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] GEIGER A, LENZ P, URTASUN R. Are we ready for autonomous driving? The KITTI vision benchmark suite[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA: IEEE Press, 2012: 3354-3361.

[29] PADILLA R, NETTO S L, DA SILVA E A B. A survey on performance metrics for object-detection algorithms[C]// Proceedings of International Conference on Systems, Signals and Image Processing. Washington D. C. , USA: IEEE Press, 2020: 237-242.

[30] HSU C C, TSAI Y H, LIN Y Y, et al. Every pixel matters: center-aware feature alignment for domain adaptive object detector [C]//Proceedings of European Conference on Computer Vision. Berlin, Germany: Springer, 2020: 733-748.

[31] WANG W, CAO Y, ZHANG J, et al. Exploring sequence feature alignment for domain adaptive detection transformers [C]//Proceedings of the 29th ACM International Conference on Multimedia. New York, USA: ACM Press, 2021: 1730-1738.

[32] LIU D N, ZHANG C Y, SONG Y, et al. Decompose to adapt: cross-domain object detection via feature disentanglement[J]. IEEE Transactions on Multimedia, 2023, 25: 1333-1344.

[33] ALQASIR H, MUSELET D, DUCOTTET C. Region proposal oriented approach for domain adaptive object detection[C]//Proceedings of International Conference on Advanced Concepts for Intelligent Vision Systems. Berlin, Germany: Springer, 2020: 38-50.

编辑　赖玉玲