基于域内域间语义一致性约束的域自适应目标检测方法

钟安雨[1),3)] 王 蕊[1),2),3)] 张 华[1),3)] 邹 聪[1),3)] 荆丽桦[1),3)]

[1)](中国科学院信息工程研究所信息安全国家重点实验室 北京 100093)

[2)](之江实验室 浙江杭州 311100)

[3)](中国科学院大学网络空间安全学院 北京 100049)

摘 要 在目标检测任务中,当训练集和测试集来自不同应用场景时,通常存在检测性能下降问题,这源于不同场景的数据间存在域偏移(domain shift).收集不同场景的数据费时费力,且会增加模型部署成本,降低模型使用效率.针对这一问题,本文从强化特征的语义一致性以获得更好的域无关特征的思路出发,提出基于域内域间语义一致性约束的域自适应目标检测方法.首先,本文考虑了特征解耦过程中的特征域内一致性,提出了一种基于正交分离特征的正交关系一致性约束,该约束能够保留解耦前后特征中的语义信息,以此强化域内特征一致性,从而提升模型识别的准确率.进一步地,本文考虑了在不同域间解耦后特征的域间一致性,引入了基于伪标签的对比学习机制,将来自不同域间的实例级特征进行对齐,以此保证域间特征一致性来提升模型的跨域性能.为验证本文所提出的方法,在本领域常用的数据集Cityscapes-FoggyCityscapes上进行了测试,相对于基线方法本文所提出的方法取得了3.1%的平均准确率(mAP)提升,其中在部分特定子类上提升达到6%;相比较最新方法也有约1%的平均准确率提升.本文还在KITTI-Cityscapes和Sim10K-Cityscapes数据集上测试了所提方法,实验结果表明本方法在其它数据集上也能取得良好的域自适应效果.

关键词 域自适应;目标检测;深度学习;特征解耦;对比学习

中图法分类号 TP18 **DOI号** 10.11897/SP.J.1016.2023.00827

Consistency-aware Domain Adaptive Object Detection via Orthogonal Disentangling and Contrastive Learning

ZHONG An-Yu[1),3)] WANG Rui[1),2),3)] ZHANG Hua[1),3)] ZOU Cong[1),3)] JING Li-Hua[1),3)]

[1)](*State Key Laboratory of Information Security*, *Institute of Information Engineering*, *Chinese Academy of Sciences*, *Beijing*, 100093)

[2)](*Zhejiang Lab*, *Hangzhou*, 311100)

[3)](*School of Cyber Security*, *University of Chinese Academy of Sciences*, *Beijing*, 100049)

Abstract Traditional object detection methods suffer from performance degradation when the training and test data are from different domains, for example, photos from a sunny day and a cloudy day are two different domains, and an object detection model trained on a sunny day usually performance not well on a cloudy day. This is caused by the domain shift between two domains. Collecting data for every single domain is time-consuming and laborious, which will increase the cost of model deployment and reduce the efficiency of the model being used. Aiming at this

收稿日期:2022-07-20;在线发布日期:2022-10-03. 本课题得到国家自然科学基金面上项目(No. 62176253)、国家自然科学基金企业创新发展联合基金重点支持项目(No. U20B2066)和之江实验室开放课题(No. 2021KB0AB01)资助.**钟安雨**,硕士研究生,主要研究领域为计算机视觉以及深度学习,E-mail:zhonganyu@iie.ac.cn. **王 蕊**(通信作者),博士,研究员,中国计算机学会(CCF)会员,主要研究领域为计算机视觉以及深度学习,E-mail:wangrui@iie.ac.cn. **张 华**,博士,副研究员,中国计算机学会(CCF)会员,主要研究领域为计算机视觉以及深度学习. **邹 聪**,博士研究生,主要研究领域为计算机视觉以及细粒度识别. **荆丽桦**,博士研究生,工程师,主要研究领域为计算机视觉以及深度学习.

problem, the domain adaptive object detection method is proposed. Most domain adaptation methods eliminate domain shifts by finding domain-invariant feature representations in two domains. Although existing domain adaptation methods have achieved great success, there are still differences between the domain-invariant features extracted from the source domain and the target domain, which lead to poor performance when the model uses domain-invariant features from the target domain. Enlightened by the idea of strengthening the semantic consistency of features to obtain better domain-invariant features, this paper proposes a Consistency-aware Domain Adaptive object detection network (ConDA) with orthogonal disentangling and contrastive learning. Specifically, this paper first proposes an orthogonal relation consistency constraint based on the orthogonal disentangled features, which can better improve the intra-domain consistency and the transferability of the model. Orthogonal constraints are applied in the feature disentangling process to keep domain-invariant and domain-specific features different. Based on the orthogonal constraints, the relationship consistency loss can be applied by calculating the instance-level feature relationship consistency before and after feature disentangling and then constraining them to be the same. This loss can retain the semantic information during feature disentangling and strengthen the intra-domain consistency of the feature, thus improving the transferability of the model. Furthermore, in order to strengthen the inter-domain consistency of features, this paper proposes a contrastive learning branch with pseudo labels. This paper uses pseudo label on the detection results of the target domain with high confidence and then aligns instance-level features from different domains with contrastive learning, which can reduce the domain shift between same class instance-level features in different domains and align instance-level domain-invariant features from different domains, and improve the inter-domain consistency. In addition, this paper also adds a target-domain-like dataset generated by CycleGAN to the source domain dataset to reduce the domain shift between the source domain and the target domain, which helps improve the model detection results. To verify the method proposed in this paper, this paper tests the method on a pair of datasets that use Cityscapes as the source domain and FoggyCityscapes as the target domain, which is commonly used in this field. Compared with the baseline method Instance Invariant Domain Adaptive Object Detection (IIOD), this method has achieved a mean Accuracy Precision (mAP) improvement of 3.1%, of which the improvement on some specific subclasses is up to 6%; Compared with other latest methods in the field, the mAP is improved by about 1%. This paper also tests the method on two other pairs of datasets, and the results show that the method can also achieve a good result on other datasets.

Keywords domain adaptation; object detection; deep learning; feature disentangling; contrastive learning

1 引 言

目标检测是计算机视觉领域中的重要任务,在自动驾驶、图像分析、视频分析等领域中都扮演了重要角色.现有基于深度学习的目标检测模型的性能依赖训练集数据的多样性和标注质量,但现实中存在多种不同的扰动因素,比如天气和光照等,为每个应用场景准备充足的训练集需要消耗大量的人力物力,因此训练集中的数据无法完全覆盖实际应用中的所有场景.来自不同场景的数据存在分布上的差异,这种差异被称为域偏移[1],每一个场景被称为一个域.在模型训练的过程中,有标注数据的域被称为源域,实际应用场景所对应的没有标注数据的域被称为目标域.域偏移会对目标检测方法的性能产生负面影响.例如,通常自动驾驶数据集是在晴天

光照良好的情况下采集的，现有检测方法[2-5]在对应环境下检测效果良好，但在恶劣天气场景下会存在漏检错检的问题，检测效果下降.如果要针对所有目标场景都收集并标注数据，就会极大地增加成本，这限制了这些方法在实际应用场景下的部署应用.

为解决上述的域偏移的挑战，研究者提出了无监督域自适应目标检测(Unsupervised Domain Adaptation Object Detection)[6]方法.该方法能够帮助模型将从源域数据集中学习到的知识应用到目标域上，减轻域偏移的影响，并提升模型的泛化性能，在物体分类[7]、目标检测[8]等任务中有广泛应用.域自适应目标检测技术的关键在于对齐两个域间的特征分布，使得模型能从源域和目标域中提取出域无关特征，从而在目标域上完成检测任务.早期工作[9-13]使用统计特征对齐的思路，约束模型从源域和目标域中提取出具有相同统计性质的特征作为域无关特征.现在的主流方法[14-19]则通过在网络中添加域分类器和梯度反转层(Gradient Reverse Layer，GRL)[20]的方式引入对抗学习，使得特征提取器能够学习到域无关特征，实现特征对齐.

然而，上述方法在实现过程中只考虑了域无关特征，并没有对数据中的域相关特征做出约束.域相关特征能够反映出图像中和域相关的信息，在不添加额外约束的情况下，这些信息使得分离出的域无关特征受到了对分类没有帮助的信息的干扰，从而影响目标检测的效果.

不同于上述方法，Wu等人提出的方法[21]在域自适应目标检测任务中引入了特征解耦机制，在对抗学习的基础上使用了成对的带有域鉴别器的特征提取器，能够显式地将特征分离成域无关和域相关两部分.相比起其它方法，引入特征解耦的优势在于显式分离出的域无关特征语义信息保留得更多，能够更好地应用于小样本学习和域自适应等下游任务[22,23].

然而，上述工作引入特征解耦缺少对一致性的考虑，这使得其解耦出的域无关特征受到域相关信息的影响，如图1所示.图1的上半部分是缺少对特征一致性约束时的解耦过程，在缺少域内一致性约束的情况下，解耦前后的不同类别特征间的语义会产生不一致，这会影响将域无关特征并入同一空间时的特征对齐效果.此外，缺少了域间一致性约束会使得来自不同域的同类目标的特征不能被对齐，使得分类器的分类边界更难被确定，降低目标检测效果.

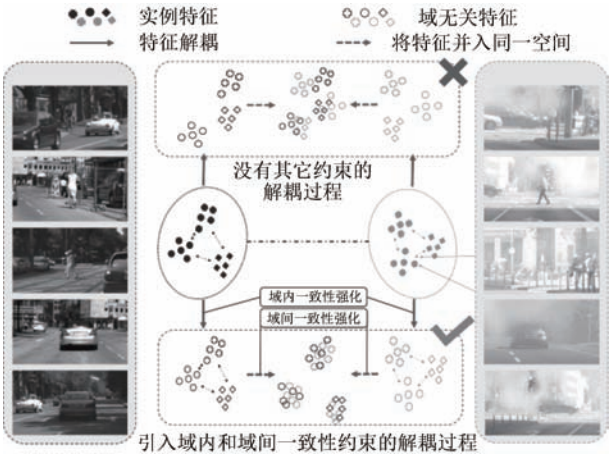

图1　考虑一致性约束前后特征解耦效果对比图

针对这两点问题，本文提出了基于正交解耦和对比学习的一致性强化域自适应目标检测框架(ConDA)，并对其做出了改进.

第一是域内一致性，即解耦前后的实例级特征间语义关系需要保持一致.缺少这种一致性表明解耦过程存在问题，影响到后续的检测与分类效果.针对这一点，本文设计了正交关系一致性损失.引入正交约束能够提高特征解耦过程中特征的质量[24]，并且增大这些特征间的距离，本文将正交约束添加到重建后的域相关和域无关特征上，能够减少解耦过程中域相关特征的干扰.在正交约束的基础上，设计了归一化关系一致性损失，能够保证解耦前后实例级域无关特征间的关系的一致性，保证特征的域内一致性.

第二是域间一致性，即来自不同域的解耦后特征应当保持一致.缺少这种一致性会使得模型受到有标签源域的影响，进而导致在目标域上的性能下降.针对这一点，受到对比学习[25]技术在表征学习领域取得成果的启发，本文在模型中添加了带有伪标签的对比学习分支.通过收集来自源域的实例级特征和来自目标域带有伪标签标注的特征，对比学习分支能够学习到每一个特定的类在源域和目标域上一致的表征，以此实现不同域间实例级特征的对齐，提高特征的域间一致性.

在保证了这两点一致性后，特征解耦过程就如同图1下半部分所示，来自不同域的不同类域无关特征能被很好地分开，从而提高目标检测的效果.

本文的主要贡献点总结如下：

(1)针对特征解耦过程缺少对特征域内一致性约束，使得解耦出的域无关特征被域相关信息影响，导致语义特征丢失的问题，本文在解耦后特征上添

加了正交约束，并提出了正交关系一致性损失，确保特征解耦前后语义特征完整.

(2)针对域自适应任务缺少对特征域间一致性约束，使得模型在目标域上出现错检漏检的问题，本文设计了带有伪标签机制的对比学习分支，能够提高特征的域间一致性并提升特征的可分辨性，减少错检漏检现象的发生.

(3)本文在本领域三个常用数据集上进行了实验，验证了本方法的有效性，并与最新方法进行了对比.结果表明，在本领域常用的测试场景Cityscapes-Foggy Cityscapes下，ConDA比基线方法在平均准确率上提升了3%；比最新的次优方法也有约1%的提升.

2 相关工作

2.1 域自适应目标检测

常用的目标检测框架可分为两阶段(如Fast-RCNN[2]、Faster-RCNN[3])和一阶段(如Yolo[14])，二者都已经取得了长足进展.目前的主流域自适应目标检测方法[8,14,15,19,26,27]大都基于两阶段目标检测框架，在这些框架的特征提取阶段上引入对抗学习方法获得域无关特征用于目标检测.较新的基于对抗学习方法的域自适应目标检测方法将对抗学习应用到模型的多个层级上，如Chen等人的工作[18]将对抗学习分为全局级别和实例级别两个等级，实现多级特征对齐.但这类方法欠缺对域相关特征的考虑，这些域相关特征会混合在提取到的域无关特征中，因此提取的域无关特征质量仍有提升空间.

此外，Wang等人[16]将模型优化过程分为了域相关方向和域无关方向，通过抑制模型在域相关方向的参数变化来实现域自适应，但这个域相关方向的选取依赖人工指定的规则，目前将模型训练过程中梯度最大的方向看做域相关优化方向，这不一定是最合适的方向；Zhang等人[28]在模型的区域选择网络(Region Proposal Network，RPN)中建立了一系列特征原型，提高该网络选择候选区域的准确率，减少漏检，但对特征提取部分没有改进；Deng等人[29]则在域自适应目标检测任务中引入了均值教师模型以及知识蒸馏的思想，将模型从源域学习到的知识迁移到目标域上，但这类方法同样未对特征提取部分进行改进，只是通过使用教师学生模型学习到域无关特征的提取.

本文选择两阶段目标检测框架Faster-RCNN[3]作为骨架网络，在此框架中引入特征解耦以及对比学习模块，以此提高特征的域内和域间一致性，提取更好的域无关特征.

2.2 特征解耦

特征解耦[30]是指从特征图中分解出语义上互不相关的特征的过程，它最早被用于生成对抗网络(Generative Adversarial Network，GAN)[31-33]中，如今在风格迁移[33]、图像生成[34,35]以及域自适应[21,36,37]等任务中都有应用.Wei等人[24]在解耦过程中添加了正交约束，并且取得了更好的解耦效果.对于域自适应目标检测任务而言，特征解耦过程能够减小域间差异并且帮助特征提取器提取更高质量的域无关特征.Wu等人[21]在实例级特征上进行了特征解耦并取得了良好的效果，但缺少对域相关信息的约束，使得解耦出的域无关特征中带有域相关信息.

本文在解耦过程中添加了基于正交化特征的一致性约束，能够减少域相关信息被错误解耦到域无关特征中，进而提高模型性能.

2.3 对比学习

对比学习[38,39]能够得到更易于分类的特征，进而提高下游任务的表现.对比学习算法将特征空间中不同类的特征进行聚类，在减少特征类内距离的同时增大特征的类间距离.相比于像素级特征，这些特征包含的语义信息更丰富，对分类任务更有用.对比学习在自监督和半监督任务中都有应用，并在小样本学习[40,41]、行人检测[42]、动作检测[43]等领域取得了良好效果.对比学习在域自适应物体检测中也有应用，例如Liu等人的工作[44]使用对比学习将使用CycleGAN生成的仿目标域实例级特征和源域的实例级特征按类进行聚类，辅助生成更好的域无关特征进行分类任务，取得了一定效果.但使用仿目标域实例级特征来替代真正的目标域特征并不合适，生成的仿目标域特征质量受CycleGAN训练质量影响，可能引入一些误差.

本文将对比学习应用到域自适应任务中以得到更好的表征，相比起已有的方法，本文使用伪标签从目标域中筛选可用于聚类的实例级特征，相比起使用生成的仿目标域特征能更好地体现目标域的特点，能提高特征的域间一致性，对齐来自不同域的域无关特征.

3 方法论述

本小节首先在3.1小节中整体介绍ConDA方

法的模型设计和数据处理流程，随后在3.2和3.3小节中分别介绍本文的核心模块基于正交关系一致性约束的域内一致性强化模块和基于对比学习和伪标签的域间一致性强化模块.最后在3.4小节中介绍ConDA的三阶段训练流程并对算法中涉及到的损失函数进行梳理.

为了后续表述的方便，首先对数学符号的使用做出约定.本文在对应的符号后使用下标s和t来分别表示来自源域和目标域的数据对象.

为了形式化地描述域自适应问题，本文用$\mathcal{D}_s=\{X_s, Y_s, B_s\}$来描述带有完整标签的源域，用$\mathcal{D}_t=\{X_t\}$来描述无标签的目标域.其中$X_s=\{x_s^i\}_{i=1}^m$是源域的图像，$Y_s=\{y_s^i\}_{i=1}^m$和$B_s=\{b_s^i\}_{i=1}^m$分别代表源域的类别标签和边界框(bounding box).

3.1　模型框架总览

本文提出方法的框架图如图2所示.在训练阶段，输入模型的图像先经过两组特征提取和特征解耦模块，再使用区域选择网络(RPN)模块从完成了特征解耦的特征图中选取候选区域，最后实现物体的定位与分类.图2中的梯形表示具有可学习参数的特征提取器或卷积层；“⊕”表示点加操作；“*RA*”、“*GRL*”和“*FC*”分别代表RoI对齐(RoIAlign)模块、梯度反转层以及全连接分支；“$\mathcal{L}_{CONT}$”、“$\mathcal{L}_{OPL}$”和“$\mathcal{L}_{ORC}$”分别代表有监督对比损失、正交约束损失和正交关系一致性损失.框架以Faster-RCNN[3]为基础，将其特征提取网络拆分成两部分，分别记作E_b^1和E_b^2，用于提取浅层和深层特征.区分浅层和深层特征并分别进行解耦的目的在于多进行的一轮解耦并将所得的域无关特征叠加在特征图的过程能够将更多的域无关信息引入到特征图中，对最后获得良好的域无关特征提供帮助.

在训练阶段，对于输入的一张图像x，使用浅层特征提取器提取浅层特征图$F_b^1=E_b^1(x)$，随后使用一对特征提取器对浅层特征进行特征解耦得到浅层域无关和域相关特征，记作$F_{di}^1=E_{di}^1(F_b^1)$和$F_{ds}^1=E_{ds}^1(F_b^1)$，这三个特征图的尺寸是一致的.

接下来，为了增加特征图中域无关特征的比例，将得到的浅层域无关特征图加在浅层特征图上.随后用深层特征提取器对该特征图进行特征提取，得到深层特征图，这个过程可以用$F_b^2=E_b^2(F_{di}^1+F_b^1)$表示.

接着，使用另一对结构相同但是具体参数不同的特征提取器对深层特征图进行解耦，以此得到深层域无关特征和域相关特征，分别用F_{di}^2和F_{ds}^2表示.RPN模块会从深层域无关特征图中寻找一组候选区域$P=\{p^i\}_{i=1}^k$.接下来，RoI对齐模块会使用这些候选区分别从深层域无关特征图、深层域相关特征图和深层特征图中截取三组实例级特征，记作

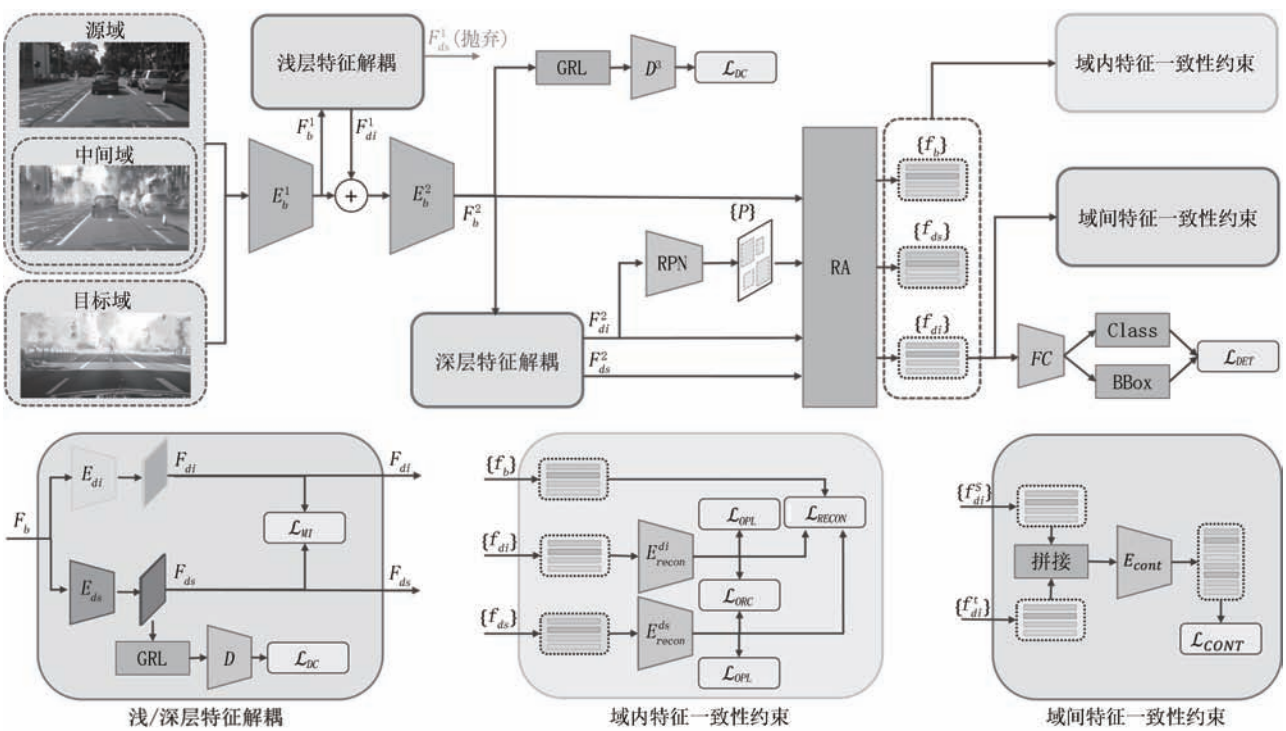

图2　本文提出方法ConDA的框架图

$\{f_{di}^i\}_{i=1}^k$、$\{f_{ds}^i\}_{i=1}^k$和$\{f_b^i\}_{i=1}^k$.为了保证解耦过程中不损失信息,本文使用重建损失来约束域无关和域相关的实例级特征在重建后和解耦前的深度特征一致,该重建损失如下所示:

$$\mathcal{L}_{RECON}=\sum_{i=1}^{k}|f_b^i-\left(E_{recon}^{di}(f_{di}^i)+E_{recon}^{ds}(f_{ds}^i)\right)|_2^2 \quad (1)$$

其中E_{recon}^{di}和E_{recon}^{ds}是两个用于将解耦后的特征恢复到解耦前的重建模块.

此外,特征解耦模块的域相关特征解耦器E_{ds}和深层特征提取器E_{di}后都添加了带有梯度反转层的域鉴别器.引入域鉴别器采用了对抗学习的思路,当域鉴别器难以鉴定特征来自哪个域时,说明两个域间的特征已经足够相似,域差距已经被成功消除.当域相关特征被成功解耦后,解耦模块另一分支得到的特征就是所需的域无关特征.浅层、深层特征解耦模块中的域相关特征提取器以及最终得到的深层特征提取器产出的特征都带有域相关特征,其得到的特征应当能被用于区分源域和目标域,因此有必要使用域鉴别器对其增加约束.本文将域鉴别器分类的过程视作二分类过程,并使用Focal Loss损失[45]来约束这三个域鉴别器,使其能够正确分辨特征来自源域或是目标域.为了后文简洁,将这三个损失合并为一项,如下所示:

$$\mathcal{L}_{DC}=\mathcal{L}_{FL}(D^1(F_{ds}^1))+\mathcal{L}_{FL}(D^2(F_{ds}^2))+\mathcal{L}_{FL}(D^3(F_b^2)) \quad (2)$$

其中D^1、D^2和D^3分别是三个域鉴别器,$\mathcal{L}_{FL}$则是Focal Loss损失函数.

为了进一步增大域无关特征F_{di}和域相关特征F_{ds}间的差异,需要最小化域相关特征和域无关特征的互信息.本文使用神经网络互信息估计器(mutual information neural estimator)[46]结合蒙特卡洛积分[36]来计算互信息损失.

本文使用的互信息损失如下所示:

$$\mathcal{L}_{MI}=\frac{1}{n}\sum_{j=1}^{n}T(\mathcal{F}_{di},\mathcal{F}_{ds},\theta)-\log\left(\frac{1}{n}\sum_{j=1}^{n}\exp(T(\mathcal{F}'_{di},\mathcal{F}_{ds},\theta))\right) \quad (3)$$

其中$\mathcal{F}_{di}$表示域无关特征,$\mathcal{F}_{ds}$表示域相关特征,$\mathcal{F}'_{di}$则是打乱顺序后的域无关特征.$\mathcal{F}_{di}$和$\mathcal{F}_{ds}$中的特征是成对的,相同维度特征对应同一个实例或者对应同一个维度,而经过了打乱顺序后的域无关特征$\mathcal{F}'_{di}$和$\mathcal{F}_{ds}$中的特征则不成对.本文对深层特征和浅层特征都进行了互信息损失的计算,其中对于深层域无关特征,本文使用实例级特征替代完整的深层特征图,这能够使得网络更关注于实例特征,同时也能减少网络的计算量,加快网络训练速度.具体而言,式中的$T(\theta)$是一个以θ为参数的全连接神经网络,负责从成对特征$(\mathcal{F}_{di},\mathcal{F}_{ds})$中采样作为联合分布,再从不成对特征$(\mathcal{F}'_{di},\mathcal{F}_{ds})$中采样作为边缘分布,以此计算互信息损失.

最后,后续的检测分支会对来自源域的图像继续进行常规的回归与分类,并根据标注计算相应的损失,这部分损失函数记作$\mathcal{L}_{DET}$.

3.2 基于正交关系一致性的域内一致性约束

提高域内特征一致性的目的是保持实例级特征在解耦前后的语义关系一致,这能够防止域相关信息被错误地解耦到域无关特征中去,提高域无关特征的质量.本文的基线方法IIOD[21]提出了一致性损失来约束实例级特征在解耦前后的相似度邻接矩阵保持一致,但由于解耦前的特征中包含有域相关信息,直接约束解耦前后特征的关系一致会使得得到的域无关特征受到这部分信息的影响.

为了解决上述缺陷,本文首先对特征解耦过程添加了正交约束,随后,在正交解耦特征的基础上提出了归一化关系一致性损失.Wei等人的工作[24]已经验证了在解耦过程中添加正交约束能够得到更好的解耦效果,受其启发,本文在重建后的域无关特征和域相关特征上添加了正交约束[47].这个约束同时也是后续归一化关系一致性约束的基础.

本小节中使用r_{di}和r_{ds}分别表示重建后的特征$E_{recon}^1(f_{di})$和$E_{recon}^2(f_{ds})$.具体来说,正交约束包含两个部分,首先是用于保证同属于域相关或者同属域无关的特征处于同一个特征空间,这一部分可以写作:

$$L_s=\frac{1}{2k^2-2k}\sum_{i=1,j=1,i\neq j}^{k}<r_{di}^i,r_{di}^j>+<r_{ds}^i,r_{ds}^j> \quad (4)$$

第二部分则是保证域相关和域无关特征之间相互正交,这部分可以写作:

$$L_d=\frac{1}{k^2}\sum_{i=1}^{k}\sum_{j=1}^{k}<r_{di}^i,r_{ds}^j> \quad (5)$$

将二者结合在一起就是完整的正交约束损失,如下所示:

$$\mathcal{L}_{OPL}=(1-L_s)+|L_d|$$

其中$<\cdot,\cdot>$是向量点乘操作.

重建后的域相关和域无关特征应当能够匹配原有解耦前的特征,可以记作:

$$f_b^i = r_{di}^i + r_{ds}^i$$

关系邻接矩阵能够用于描述实例级特征间的关系.对于来自深度特征图的实例级特征$\{f_b^i\}_{i=1}^k$,将其关系邻接矩阵记作A_b,其中的元素$A_b(i,j)$用于表示f_b^i和f_b^j两个实例特征间的关系.本文使用余弦相似度来计算上述关系,结合正交约束定义式,A_b中的元素可按下式来计算:

$$A_b(i,j) = f_b^i \cdot f_b^j = (r_{di}^i + r_{ds}^i)\cdot(r_{di}^j + r_{ds}^j) \quad (6)$$

由于重建后的域相关和域无关特征已经进行了正交约束,因此对于任意i和j,总有$r_{di}^i \cdot r_{ds}^j = 0$.

此时关系邻接矩阵可以被化简为:

$$A_b(i,j) = r_{di}^i \cdot r_{di}^j + r_{ds}^i \cdot r_{ds}^j \quad (7)$$

同样用余弦相似度来表示特征间的关系,域无关特征的关系邻接矩阵A_{di}中的元素可以按下式计算:

$$A_{di}(i,j) = r_{di}^i \cdot r_{di}^j \quad (8)$$

为了最小化域相关信息带来的影响,本文用每个关系邻接矩阵减去其自身的均值.令$\omega_{di}^{(i,j)} = r_{di}^i \cdot r_{di}^j$,$\omega_{ds}^{(i,j)} = r_{ds}^i \cdot r_{ds}^j$,就能得到减去均值后的两个关系邻接矩阵,记作:

$$\widehat{A_b}(i,j) = A_b(i,j) - \frac{1}{k^2}\sum_{i=1}^{k}\sum_{j=1}^{k}\left(\omega_{di}^{(i,j)} + \omega_{ds}^{(i,j)}\right) \quad (9)$$

$$\widehat{A_{di}}(i,j) = A_{di}(i,j) - \frac{1}{k^2}\sum_{i=1}^{k}\sum_{j=1}^{k}\omega_{di}^{(i,j)} \quad (10)$$

如果令$\widehat{A_b} = \widehat{A_{di}}$,此时每个$\widehat{A_b}$中的元素和它在$\widehat{A_{di}}$所对应的另一个元素间的差距可写作:

$$\epsilon(i,j) = \omega_{ds}(i,j) - \frac{1}{k^2}\sum_{i=1}^{k}\sum_{j=1}^{k}\omega_{ds}(i,j) \quad (11)$$

这就是深度特征图中域相关信息带来的影响.将每个元素的差距相加,可得$\sum_{i=1}^{k}\sum_{j=1}^{k}\epsilon(i,j) = 0$,证明此时从整体来看域无关信息带来的影响能够相互抵消.

综上所述,本文新设计的基于正交约束的归一化关系一致性损失优于基线模型中的关系一致性损失,解决了基线模型中损失没有考虑到域相关信息带来影响的问题.这个归一化的关系一致性损失如下所示:

$$\mathcal{L}_{ORC} = \sum_{i=1}^{k}\sum_{j=1}^{k}\widehat{A_b} - \widehat{A_{di}} \quad (12)$$

3.3　基于对比学习和伪标签的域间一致性约束

除了保证解耦前后域内特征的一致性,保证不同域间解耦后特征的一致性也很重要.对比学习是一种减小特征类内距离,扩大类间距离的有效手段,能够使得特征更容易被分类[25].受其启发,本文将来自不同域的同类特征视为一个大类,使用对比学习来缩小它们间的差距,以此实现域间一致性的强化.同时,域间距离的增大也能降低分类器分类的难度,进一步提高模型性能.

本文在RoI对齐模块后添加了对比映射分支[39],该分支与分类和回归分支并行.本文中的对比映射分支包含两个全连接层加激活层的结构,并在最后带有一个归一化层.对比映射分支的输入是实例级域无关特征$\{f_{di}^i\}_{i=1}^k$,它会将这些输入进行重新编码,得到一组维度更小的向量$\{z^i \in \mathbb{R}^{C_{cont}}\}_{i=1}^k$用于对比损失的计算.

由于有监督对比损失效果优于无监督对比损失[48],在源域有标注可用的情况下,本文采取有监督对比损失InfoNCE[49]作为对比学习的损失函数.计算该损失需要两组编码后的特征以及它们的标签$S_s = \{z_s^i, y_s^i\}_{i=1}^M$和$S_t = \{z_t^i, y_t^i\}_{i=1}^N$,其中$y$表示特征$z$所对应的列表标注.由于一轮训练过程中的两张图像不成对,M和N通常是不同的.对于来自源域的图像而言,标注$\{y_s\}$是已知的;而对于来自目标域的图像,本文使用伪标签法用网络生成的预测结果作为标签.根据本文在前期实验中的观察,基线模型在目标域出现的漏检多于错检,证明在区域选择模块选择了候选区域的前提下,模型给出的分类结果多数情况下是可以信任的.因此,本文设置了一个阈值λ来筛选模型给出的分类结果,当模型给出的置信度高于阈值时,就用模型预测结果作为对应特征的伪标签$\{y_t\}$.

为了提高特征的域间一致性,本文将得到的两组特征S_s和S_t合并到一个更大的集合$S = \{z^i, y^i\}_{i=1}^{M+N}$中,并按下式来计算对比损失:

$$\mathcal{L}_{CONT} = \frac{1}{M+N}\sum_{i=1}^{M+N} L_{z_i}, \quad (13)$$

$$L_{z_i} = -\frac{1}{N_{S_i^+}}\log\frac{\sum_{j=1}^{N_{S_i^+}}\exp\left(\mathrm{sim}\left(z_i, z_j^+\right)/\tau\right)}{\sum_{k=1}^{N_{S_i^-}}\exp\left(\mathrm{sim}\left(z_i, z_k^-\right)/\tau\right)} \quad (14)$$

其中τ是能够调整正样本对和负样本对重要程度的超参数[38,39],$S_i^+ = \{z_j\}_{j=1}^{N_{S_i^+}}$是与$S$中除$z_i$本身外所有类别为$y_i$的特征构成的集合,$S_i^- = \{z_j\}_{j=1}^{N_{S_i^-}}$则是$S$中所有类别不为$y_i$的特征构成的集合.$N_{S_i^+}$和$N_{S_i^-}$分别代表集合$S_i^+$和$S_i^-$中的元素数量.$\mathrm{sim}(\cdot)$表示对比

损失的相似度计算方式，在此处使用余弦相似度.

3.4 三阶段优化策略

为了将损失函数应用在模型中的合适部位以获得更好的优化效果，本文仿照基线模型将训练过程分成了三阶段，以端到端的方式训练，在每一阶段中没有被提及的模块中的参数都被固定不发生变动.在训练阶段的每一次迭代中，一对分别来自于源域和目标域的图像$\{x_s, x_t\}$会被送入模型三次，每次都会计算不同的损失并优化模型中不同部分的参数.上标s和t分别用于表示用源域图像x_s和目标域图像x_t计算得出的损失函数，下标p_1、p_2和p_3则表示三个不同阶段的损失函数.

采取这种设计会在训练阶段带来额外的计算量，但是从基线模型[21]的实验中可以看出针对不同模块添加损失函数能够获得更好的实验结果，计算量的提升是可以接受的.此外，采取三阶段优化策略并不会影响模型在测试阶段时的计算量，因此对于实际应用场景下的模型没有影响.如果条件不允许，本方法也可以将三阶段的损失合并在一个阶段中，进行端到端训练.

第一阶段的目标是优化整个模型中的所有参数.本文选择在这一阶段计算检测损失、域分类器损失以及对比损失，这些损失都和所有的特征提取器有关.第一阶段的整体损失函数如下所示：

$$\mathcal{L}_{p_1}=\mathcal{L}_{DC}^{s}+\mathcal{L}_{DET}^{s}+\mathcal{L}_{DC}^{t}+\mathcal{L}_{CONT} \tag{15}$$

第二阶段的目标则是强化解耦效果，增加解耦后的特征对之间的差异，并保证解耦前后语义信息的一致性，尽量保留域无关信息.本文选择在这一阶段计算互信息损失，正交约束损失和归一化关系一致性损失，这些损失都与特征解耦相关.需要优化的参数包括模型的基础特征提取器以及两对用于解耦的特征提取器.第二阶段的整体损失函数如下所示：

$$\mathcal{L}_{p_2}=\mathcal{L}_{MI}^{s}+\mathcal{L}_{OPL}^{s}+\mathcal{L}_{ORC}^{s}+\mathcal{L}_{MI}^{t}+\mathcal{L}_{OPL}^{t}+\mathcal{L}_{ORC}^{t} \tag{16}$$

第三阶段的目标是保证实例化级别的解耦过程中不出现信息损失，因此只需要对深层特征解耦部分的特征计算重建损失，需要优化的参数也只有对应的两个特征提取器E_{ds}^{2}和E_{dt}^{2}.这一阶段的整体损失函数如下所示：

$$\mathcal{L}_{p_3}=\mathcal{L}_{RECON}^{s}+\mathcal{L}_{RECON}^{t} \tag{17}$$

最后，上述的三阶段训练流程可以用算法1所示的流程来描述，本文训练过程中使用的损失函数及其作用的总结可以总结为表1中的内容.要额外提及的是，由于这些损失函数的数量级基本一致，因此在训练过程中没有对其进行额外的权重调整.

表1 本文所用的损失函数的总结

损失函数名	作用
检测损失 $\mathcal{L}_{DET}$	完成目标检测任务
域分类损失 $\mathcal{L}_{DC}$	使得特征提取器能在不同域上提取出域无关特征
对比损失 $\mathcal{L}_{CONT}$	减小不同域的域无关特征的类内距离，增大类间距离，确保域间一致性
互信息损失 $\mathcal{L}_{MI}$	扩大域无关域相关特征的差异
正交约束损失 $\mathcal{L}_{OPL}$	增强特征解耦的效果，作为关系一致性损失的基础
关系一致性损失 $\mathcal{L}_{ORC}$	约束解耦前后特征间关系一致，确保域内一致性
重建损失 $\mathcal{L}_{RECON}$	减少解耦过程中信息的损失

算法1 ConDA三阶段训练算法

输入：有标注的源域数据集$\{X_s, Y_s, B_s\}$，无标注目标域数据集$\{X_t\}$，预训练特征提取器E_b^1和E_b^2以及其对应的特征解耦模块，学习率lr，最大训练代数e_{max}，伪标签阈值λ

输出：训练完毕的模型M

1. FOR $e\in\{0, 1, 2, \cdots, e_{max}\}$ DO
2. 从源域数据集随机抽取一张图片$\{x_s, y_s, b_s\}$，从目标域数据集随机抽取一张图片$\{x_t\}$

// 第一阶段训练，优化所有参数

3. 将x_s输入模型M中，利用标注数据y_s, b_s计算目标检测损失$\mathcal{L}_{DET}^{s}$，并根据式(2)计算域分类损失$\mathcal{L}_{DC}^{s}$
4. 将x_t输入模型M中，根据设置的伪标签阈值λ将模型预测结果作为伪标签y_t, b_t，并根据式(2)计算域分类损失$\mathcal{L}_{DC}^{t}$
5. 根据标注数据$\{y_s, b_s\}$和伪标签$\{y_t, b_t\}$从模型特征图中提取实例特征，并输入对比分支，得到编码后特征集合$S_s=\{z_s^i, y_s^i\}_{i=1}^{M}$和$S_t=\{z_t^i, y_t^i\}_{i=1}^{N}$
6. 根据式(13)和式(14)计算对比损失$\mathcal{L}_{CONT}$
7. 根据式(15)计算第一阶段整体损失$\mathcal{L}_{p_1}$，并基于该损失优化更新模型M中的所有参数

// 第二阶段训练，强化解耦效果，优化特征提取器参数

8. 将x_s输入模型M中，根据式(3)计算互信息损失$\mathcal{L}_{MI}^{s}$，根据式(4)和式(5)计算正交约束损失$\mathcal{L}_{OPL}^{s}$，根据式(12)计算归一化关系一致性损失$\mathcal{L}_{ORC}^{s}$
9. 将x_t输入模型M中，仿照步骤8计算目标域图像的互信息损失$\mathcal{L}_{MI}^{t}$、正交约束损失$\mathcal{L}_{OPL}^{t}$和归一化关系一致性损失$\mathcal{L}_{ORC}^{t}$
10. 根据式(16)计算第二阶段整体损失$\mathcal{L}_{p_2}$，并更新模型M中的特征提取器和特征解耦模块参数

// 第三阶段训练，保证解耦过程不出现信息丢失

11. 将x_s输入模型M中，根据式(1)计算重建损失$\mathcal{L}_{RECON}^{s}$
12. 将x_t输入模型M中，仿照步骤11计算目标域图像的重建损失$\mathcal{L}_{RECON}^{t}$
13. 根据式(17)计算第三阶段整体损失$\mathcal{L}_{p_3}$，并更新模型中的深层特征解耦模块E_{ds}^{2}和E_{di}^{2}的参数

14. END
15. 返回训练完毕的模型M

4　实验与分析

4.1　数据集

首先介绍本文在实验中用到的四个数据集以及生成“中间域”数据扩展训练集的过程.

Cityscapes：Cityscapes[50]是一个高分辨率城市街景图片数据集.它包含了在良好天气环境下从50个不同城市街道上收集的共计2975张训练集图像和500张验证集图像.

FoggyCityscapes：FoggyCityscapes[51]是一个基于Cityscapes[50]数据集人工合成的数据集，它包含了一组人工合成的不同浓度雾气下的街景图像.每张Cityscapes数据集中的图像都有三个不同浓度雾气的版本，本文在后续实验过程中均使用雾气最重版本的图像.

KITTI：KITTI[52]是一个包含了城市街景、高速路和农村场景图片的数据集，涵盖了能够用于多种下游任务的数据格式.本文工作中使用KITTI数据集的2D检测任务子集，该子集包含7481张训练集图像和7518张验证集图像.

Sim10K：Sim10K[53]是一个使用GTAV(Grand Theft Auto V)游戏引擎人工合成的图像数据集，它有三个不同体量的版本，分别包含1万、5万和20万张图片.本文实验中使用包含1万张图片的子集用于训练.

中间域数据集生成：为了对数据集进行扩展，同时尽量消除源域和目标域间的域差距，本文仿照Chen等人的工作[18]使用CycleGAN[54]生成“中间域”数据，对于每一个“源域-目标域”的组合训练一个CycleGAN模型，并用源域的训练集图像仿照目标域风格生成仿目标域的源域图像，这些图像将在训练过程中被添加到源域训练集中.

4.2　实现细节

本文使用ResNet-101[55]作为特征提取器的骨干网络，其参数在ImageNet上进行过预训练.根据实验测试的结果，对比学习分支的超参数τ设为0.2，伪标签阈值λ设为0.8，映射后的特征维度C_{cont}设为128.对源域是Cityscapes的情况，本文使用SGD作为优化器并将起步学习率设为0.001，动量参数设为0.9，每经过五个epoch，学习率将下降十倍.epoch大小设置为10 000，共计训练15个epoch，并从其中选择最好的结果作为最终结果.

对于KITTI或是Sim10K数据集作为源域的情况，由于这些数据集只有一个目标类，为了减轻过拟合现象并加快收敛速度，此时的起步学习率设为0.0005，并每经过3个epoch就降低一次学习率.

和本领域其它工作一样，本文以0.5阈值下的平均准确率(mean Average Precision，mAP)作为评价指标.

4.3　定量实验

本文将实验结果与其他已发表文献中的实验结果在三种不同的域自适应条件下进行比较，分别是不同天气下的适应、虚拟到现实的适应以及跨成像设备的适应.

另外，本文在实验过程中另外训练了两个使用ResNet-101[55]骨干网络的Faster-RCNN[3]模型：一个只用有标注的源域数据进行训练并直接在目标域上进行测试，它的结果能用于展示在没有添加域自适应方法时常规目标检测方法的域自适应能力，可以作为供参考的下限，在实验表格中用“S-Only”表示；另一个则直接用有标注的目标域数据进行训练并测试，它的结果能用做域自适应方法上限的参考.一般来说，所得结果越接近这项设置下的结果，就说明方法越接近域自适应方法的极限，其结果在实验表格中用“Oracle”来表示.

4.3.1　不同天气适应

在不同天气条件下的域自适应是域自适应任务最流行的任务之一，它能够在自动驾驶、城市监控等很多领域得到应用.训练集中未出现过的恶劣的天气条件可能对检测结果带来负面影响，影响自动驾驶车辆的安全.在这组设置下，本文使用Cityscapes数据集作为源域，FoggyCityscapes数据集作为目标域.如同其它工作，例如Chen等人在本数据集对上进行测试的实验设置[8]，本文从Cityscapes数据集中选择了八个类的物体用于训练和测试.

不同天气条件下的适应结果展示在表2中.综合所有子类，ConDA取得了41.7%的平均准确率.

从表2中能够看出，ConDA在公交、火车、卡车

表2 从Cityscapes迁移到FoggyCityscapes上的平均准确率(%)

方法	公交	单车	轿车	摩托	行人	骑行者	火车	卡车	mAP
S-Only[3]	26.4	27.9	33.1	18.0	24.7	31.9	9.2	11.0	22.8
DAF[8]	35.3	27.1	40.5	20.0	25.0	31.0	20.2	22.1	27.6
SCDA[19]	39.0	33.6	48.5	28.0	33.5	38.0	23.3	26.5	33.8
CRDA[27]	45.1	34.6	49.2	30.3	32.9	43.8	36.4	27.2	37.4
HTCN[18]	47.4	37.1	47.9	32.3	33.2	47.5	40.9	31.6	39.8
IIOD[21]	46.1	35.3	49.6	29.9	32.8	44.4	38.0	33.0	38.6
CRDA+VDD[56]	52.0	36.8	51.7	34.2	33.4	44.0	34.7	33.9	40.0
CRDA+FAA[17]	46.5	35.3	48.3	32.3	37.4	46.4	39.3	32.6	39.8
KTNet[57]	41.2	38.8	**60.6**	30.7	**46.4**	43.2	40.4	25.8	40.9
DSS[16]	49.2	**41.8**	53.6	**36.2**	42.9	**51.2**	18.9	33.6	40.9
HTCN+RPA[28]	45.5	36.8	49.6	35.7	33.6	43.8	46.0	32.9	40.5
Oracle[3]	49.2	38.8	52.7	35.3	37.2	48.2	48.5	35.2	43.5
ConDA	**52.1**	34.8	52.1	32.7	32.7	45.8	**47.3**	**37.8**	**41.7**

三个类别上分别取得了52.1%、47.3%、37.8%的平均准确率，高于其余的方法.在之前的工作中，这些类很很容易被错误地分类为“轿车”或其它载具，本文通过提高域间一致性的方式增大了这些类别的类间距离，从而提高了模型对这些类分类的准确程度.

在单车、轿车、摩托、行人、骑行者五个子类上，ConDA也取得了相对良好的效果.和KTNet[57]以及DSS[16]相比，ConDA缺少对特定子类的针对性优化，因此在某些子类上效果稍差于上述方法.但在综合权衡所有类别后，ConDA在总体平均准确率上仍有0.8%的提升，在整体性能上优于表中其他方法.

此外，ConDA在多数子类上比同样使用特征解耦方法的基线模型IIOD[21]有较大提升，在总准确率上也有3.1%的提升，说明基于正交约束的归一化关系一致性损失以及带有伪标签的对比学习分支在提高特征一致性上起到了作用，能够提高解耦后域无关特征的质量，进而提升检测效果.

最后，在两个特定类别“卡车”和“公交”上，ConDA取得了高于Oracle设置下这两类的准确率.这是由于这两类物体在画面中通常比较靠前，受雾气影响较轻，同时占画面面积较大易于检测，使得这两类物体作为伪标签数据的准确率很高.这相当于在训练时同时使用了源域和目标域的标签信息，因此能够得到高于仅使用目标域标注的Oracle的实验结果.

4.3.2 虚拟到现实的适应

相比于从现实生活中采集数据，虚拟合成数据在收集和标注上的成本很低.使用虚拟合成的数据集来替代现实采集的数据集能够极大降低收集数据集的成本.对于从虚拟到现实的域自适应任务，本文使用Sim10K作为源域，Cityscapes作为目标域.由于两个数据集类别的交集较少，在后续实验中只对“轿车”类别进行检测.

这组设置下的实验结果展示在表3的第二列.本文提出的方法在轿车分类上取得了43.1%的平均准确率，比最新方法[19]有0.1%的提升.

表3 从Sim10K和KITTI迁移到Cityscapes的准确率(%)

方法	S→C	K→C
S-Only[3]	34.3	30.2
DAF[8]	38.9	38.5
SWDA[58]	40.1	37.9
ATF[14]	42.8	42.1
SCDA[19]	43.0	**42.5**
Oracle[3]	60.0	62.7
ConDA	**43.1**	42.4

4.3.3 跨摄像机的适应

摄像机内参和外参的不同会导致采集到图像在质量、尺寸、视角上的不同，这些不同同样会带来跨域的性能损失，需要使用域自适应方法解决.对于跨摄像机的域自适应任务，本文使用KITTI作为源域，Cityscapes作为目标域，且同样只对“轿车”类别进行检测.

这组设置下的实验结果展示在表3的第三列.本文提出的方法在轿车分类上取得了42.4%的平均准确率，能够展示出ConDA在其它域自适应情况下同样能取得稳定良好的结果.

4.4　消融实验

4.4.1　模块有效性实验

本文通过从基线模型上逐渐添加所设计的模块的方式进行消融实验.消融实验中使用Cityscapes作为源域,FoggyCityscapes作为目标域,结果如表4所示.其中"*c*"表示只对来自源域的实例级特征使用对比学习,"*c**"表示使用本文设计的完整的伪标签和对比学习,"*n*"表示使用正交约束和归一化一致性损失,"*data*"表示在源域训练集中添加由CycleGAN生成的中间域数据.

表4　ConDA的消融实验结果(%)

方法	公交	单车	轿车	摩托	行人	骑行者	火车	卡车	mAP
基线[21]	46.1	35.3	49.6	29.9	32.8	44.4	38.0	33.0	38.6
基线+*c*	45.8	35.5	51.1	33.6	34.3	44.1	34.6	35.7	39.3
基线+*c**	46.7	36.6	50.8	32.9	32.3	43.2	47.6	31.2	40.2
基线+*n*	46.7	35.6	51.4	33.5	32.1	42.5	45.8	33.1	40.1
基线+*c**+*n*	51.8	33.2	51.9	31.7	32.4	45.3	47.0	34.1	40.9
基线+*data*	51.6	36.0	51.8	32.8	34.1	45.1	42.2	35.6	41.2
ConDA(完整)	**52.1**	34.8	**52.1**	32.7	32.7	**45.8**	47.3	**37.8**	**41.7**

首先,在只添加简单的对比学习分支的情况下,ConDA相比基线模型取得了0.7%的平均准确率提升(38.6%→39.3%).这说明对比学习分支能够起到增大实例级特征的域间距离,减少域内距离的作用,对于分类任务的准确性很有帮助.

伪标签能够进一步地提升对比学习分支的效果,说明在添加了伪标签的情况下,来自目标域的特征可以参与到对比学习的过程中,使得来自目标域的特征能够减少和源域特征的差距,提升域间一致性并进一步提高分类准确性.

接着分析正交约束和归一化一致性损失的效果.相比于基线模型,ConDA取得了1.5%的平均准确率提升(38.6%→40.1%),这说明本文所提出的正交约束和归一化一致性损失成功地减轻了基线模型解耦过程中域相关信息带来的影响,保证了特征的域内一致性,提高了解耦后域无关特征的质量,实现了更好的特征解耦.

最后测试了将两种方法合并在一起后的实验结果,该结果表明这两种方法分别在两个不同的方向上起到作用,二者带来的提升是可以叠加的.

此外,近期提出的方法[18,29]都在训练过程中使用了中间域数据以取得更好的结果.为了单独验证中间域数据带来的影响,本文测试了加入中间域数据的基线模型的表现,如表4的第6行所示.能够看出,中间域数据能给基线模型带来很大的提升,这是因为中间域数据减少了源域和目标域之间的差距,减轻了域偏移,从而降低了域自适应的难度.在添加了中间域数据后,ConDA同样取得了一定的提升.

4.4.2　超参数取值实验

为了确认本文模块中部分超参数的取值,本文对这些超参数的取值进行了测试,实验结果如表5、表6、表7所示.

表5　不同温度系数τ对结果的影响(%)

τ取值	0.1	0.2	0.4
mAP	38.6	**39.3**	37.8

表6　对比分支不同映射维度C_{cont}对结果的影响(%)

C_{cont}取值	128	512	1024
mAP	**39.3**	39.1	**39.3**

表7　不同伪标签系数λ对结果的影响(%)

λ取值	0.5	0.7	0.8	0.9
mAP	39.2	39.7	**40.2**	38.2

表5和表6是对比学习模块相关参数的消融实验,这些实验是仅在基线模型上添加对比学习模块下进行的.在表5实验中,映射维度C_{cont}取1024维,在表6实验中,温度系数τ取0.2.

温度系数τ能够影响正样本对和负样本对的重要程度,过大代表正样本对重要程度更高,模型倾向于将样本都聚类到一起,但会提高错检概率;过小则代表负样本对重要程度更高,模型倾向于保持不同类样本间的间距,但会提高漏检概率.因此,保持适中的温度系数是最合适的,根据表5中的实验结果,本文取$\tau=0.2$进行后续实验.

对比分支映射维度C_{cont}影响映射后向量保留信息的完整程度,对于较为复杂的任务来说取更大的值比较合适.根据表6中的结果,几个不同的映射维度

取值对最后结果影响不明显，考虑到减少计算量的因素，选取较小的映射维度$C_{cont}=128$进行后续实验.

表7是伪标签阈值λ选取的消融实验，如果该参数过大，就不能产生足量的伪标签数据供模型训练，影响最终效果，该参数过小则会导致模型训练过程中产生的错误结果被误判为正确标签，在模型训练过程中引入更多误差.根据表7中的实验结果，本文取$\lambda=0.8$进行后续实验.

4.5 定性实验

在这一小节，本文将用三种可视化的方法展示本文所述方法的表现.

4.5.1 检测结果

图3展示了ConDA和基线模型检测结果的直观对比.第一行图片展示了典型的错检情况.对于画面中的骑行者，基线模型将其识别为了“行人”，而ConDA正确地完成了检测.第二行图片展示了漏检情况，在画面右侧有数个行人，但是基线模型只检测到其中一个，ConDA则全都检测到了.第三行图片是错检和漏检混合的情况，同基线模型相比，ConDA成功地检测并区分了“轿车”和“摩托车”.

图3 ConDA和基线模型方法的检测结果对比，不同颜色深度的框体表示不同的类

错检和漏检通常意味着模型鉴别能力(discriminability)和迁移能力(transferability)的不足，而通过对域内一致性和域间一致性的强化，ConDA在这两项能力上都有提升.

4.5.2 CAM结果对比

CAM(Class activation mapping，类激活映射)图[59]通常被用于可视化表示图像中分类依据区域的热度图.通过在ConDA中引入对比学习，模型中的特征提取器能提取到更易于区分的特征，这意味着分类器能够关注到图像中更合适的部分并做出决策，这个过程可以被CAM可视化.需要特别提到的一点是由于两个模型中的区域选择模块参数不同，即使是对同一个目标给出的候选区也不相同，因此本文直接将标注数据中的边界框数据输入到RoI对齐模块来提取实例级特征并用于分类和CAM图绘制.

图4的第一行展示了错检情况下的CAM图.基线模型将其识别为了“行人”，而ConDA模型正确将其分类为“骑行者”.从CAM图4(b)和4(c)的对比中可以看出，ConDA模型的分类器将注意力分配给了自行车的车轮部分，同时相对于基线模型的热度图，ConDA的热度图关注区域(图中深色区域)更集中，且在车轮处也有一些关注区域，这使得本文模型能将画面中的人分类为骑行者.图4的第二行则展示了漏检情况下的CAM图.基线模型没有检测到画面中的物体，而ConDA模型则正确检测到了画面

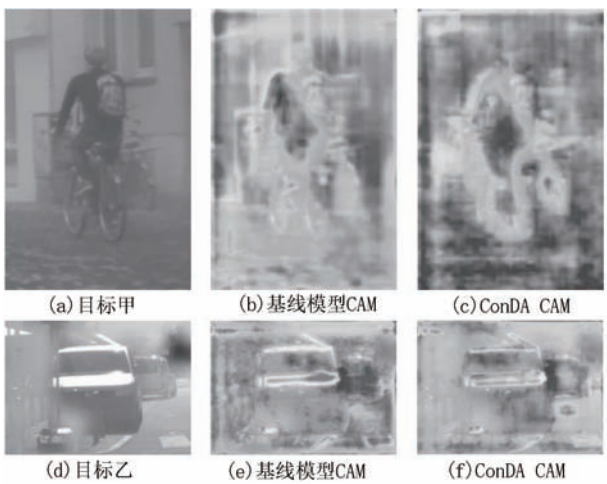

图4 ConDA和基线方法对同一目标的CAM图对比

中的轿车.从CAM图4(e)和4(f)的对比中可以看出ConDA模型将注意力集中在了车体上,实现了正确检测.

4.5.3　t-SNE结果对比

t-SNE[60]是一种能够直观有效地展示高维空间中不同类别样本点可区分性的方法.本文在图5中对比了ConDA和基线模型在分类阶段使用的实例级域无关特征的可区分性,其中被框出的部分是表现更好的地方.图中的每一个点都是一个来自于FoggyCityscapes测试集的实例级特征的映射.同样的,由于两个模型参数不同导致对候选框的选择不同,本文在这项实验中也采用标注的包围框作为RoI对齐模块提取实例级特征的标准.根据结果,能够看出本文提出方法的t-SNE图比起基线模型的有更大的类间距离,这意味着分类器能够以更高的准确率实现分类.

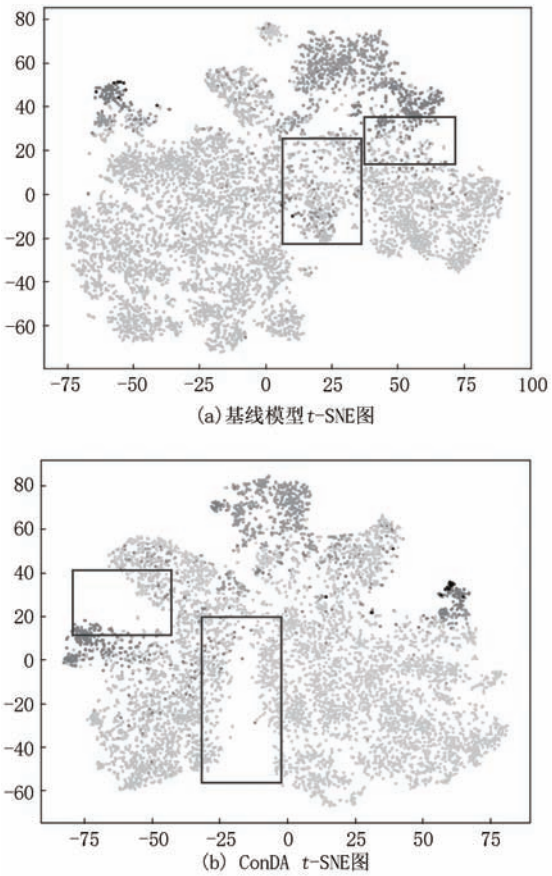

图5　ConDA和基线模型方法的t—SNE图对比

5　结　论

本文提出了一种基于域内域间语义一致性约束的域自适应目标检测框架,通过约束并增强特征的域内一致性和域间一致性,提升了模型的鉴别能力和迁移能力.本文首次提出了正交约束和归一化关系一致性损失结合的方法,这能够约束实例级特征的关系邻接矩阵在解耦前后不发生变化,同时减轻域相关数据带来的影响,提高解耦的效果.本文还设计了带有伪标签机制的对比学习分支,能够减小来自不同域的同类物体特征的类内间距并扩大它们的类间距离,提高特征的域间一致性.最后,本文在三个数据集上进行了实验,与其它方法进行对比并验证了本文提出方法的有效性.

参　考　文　献

[1] Torralba Antonio, Efros Alexei A. Unbiased look at dataset bias//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Colorado Springs, USA, 2011: 1521-1528

[2] Girshick Ross B. Fast R-CNN//Proceedings of the IEEE International Conference on Computer Vision. Santiago, Chile, 2015: 1440-1448

[3] Ren Shaoqing, He Kaiming, Girshick Ross B., Sun Jian. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149

[4] Redmon Joseph, Santosh Kumar Divvala, GirshickRoss B., Farhadi Ali. You Only Look Once: Unified, Real-Time Object Detection//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA, 2016: 779-788

[5] He Kaiming, Gkioxari Georgia, Dollár Piotr, Girshick Ross B. Mask R-CNN//Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy, 2017: 2980-2988

[6] Sinno Jialin Pan, Qiang Yang. A Survey on Transfer Learning. IEEE Transactions on Knowledge and Data Engineering, 2010, 22(10): 1345-1359

[7] Dai Hong, Hao Xuan Ting, Sheng Li Jie, Miao Qi Guang. Domain Adaptation Algorithm for Few-Shot Classification Task. Chinese Journal of Computers, 2022, 45(05): 935-950 (in Chinese)
(戴宏,郝轩廷,盛立杰,苗启广. 面向小样本约束的域适应分类算法. 计算机学报,2022,45(05):935-950)

[8] Chen Yuhua, Wen Li, Sakaridis Christos, Dai Dengxin, Luc Van Gool. Domain Adaptive Faster R-CNN for Object Detection in the Wild//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA, 2018: 3339-3348

[9] Gong Boqing, Yuan Shi, Fei Sha, Grauman Kristen. Geodesic flow kernel for unsupervised domain adaptation//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Providence, USA, 2012: 2066-2073

[10] Fernando Basura, Habrard Amaury, Sebban Marc, Tuytelaars Tinne. Unsupervised Visual Domain Adaptation Using Subspace Alignment//Proceedings of the IEEE International Conference on Computer Vision. Sydney, Australia, 2013:2960-2967

[11] Long, and Cao Mingsheng, Yue and Wang, Jianmin and Jordan, Michael. Learning Transferable Features with Deep Adaptation Networks//Proceedings of the International Conference on Machine Learning. Lille, France, 2015:97-105

[12] Long Mingsheng, Han Zhu, Wang Jianmin, Jordan Michael I. Deep Transfer Learning with Joint Adaptation Networks//Proceedings of the International Conference on Machine Learning. Sydney, Australia, 2017:2208-2217

[13] Zellinger Werner, Grubinger Thomas, Lughofer Edwin, Natschläger Thomas, Saminger-PlatzSusanne. Central Moment Discrepancy (CMD) for Domain-Invariant Representation Learning//Proceedings of the International Conference on Learning Representations. Toulon, France, 2017:1-13

[14] He Zhenwei, Lei Zhang. Domain Adaptive Object Detection via Asymmetric Tri-Way Faster-RCNN//Proceedings of the European Conference on Computer Vision. Glasgow, UK, 2020: 309-324

[15] Hsu Cheng-Chun, Tsai Yi-Hsuan, Lin Yen-Yu, Yang Ming-Hsuan. Every Pixel Matters: Center-Aware Feature Alignment for Domain Adaptive Object Detector//Proceedings of the European Conference on Computer Vision. Glasgow, UK, 2020:733-748

[16] Yu Wang, Rui Zhang, Zhang Shuo, Miao Li, Xia Yangyang, ZhangXishan, LiuShaoli. Domain-Specific Suppression for Adaptive Object Detection//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. virtual, 2021: 9603-9612

[17] Huang Jiaxing, Guan Dayan, Xiao Aoran, Lu Shijian. RDA: Robust Domain Adaptation via Fourier Adversarial Attacking//Proceedings of the IEEE International Conference on Computer Vision. Montreal, Canada, 2021:8968-8979

[18] Chen Chaoqi, Zheng Zebiao, Ding Xinghao, Yue Huang, Qi Dou. Harmonizing Transferability and Discriminability for Adapting Object Detectors//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Seattle, USA, 2020:8866-8875

[19] Zhu Xinge, Pang Jiangmiao, Yang Ceyuan, Shi Jianping, Lin Dahua. Adapting Object Detectors via Selective Cross-Domain Alignment//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Long Beach, USA, 2019: 687-696

[20] Ganin, Yaroslav and Lempitsky, Victor. Unsupervised Domain Adaptation by Backpropagation//Proceedings of the International Conference on Machine Learning. Lille, France, 2015:1180-1189

[21] Wu, and HanAming, and ZhuYahong, and YangLinchao, Yi. Instance-Invariant Domain Adaptive Object Detection Via Progressive Disentanglement. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022,44(8):4178-4193

[22] LocatelloFrancesco, BauerStefan, LucicMario, RätschGunnar, GellySylvain, SchölkopfBernhard, BachemOlivier. Challenging Common Assumptions in the Unsupervised Learning of Disentangled Representations//Proceedings of the International Conference on Machine Learning. Long Beach, USA, 2019: 4114-4124

[23] DoKien, TranTruyen. Theory and Evaluation Metrics for Learning Disentangled Representations//Proceedings of the International Conference on Learning Representations. Addis Ababa, Ethiopia, 2020:1-30

[24] Wei Yuxiang, Shi Yupeng, Xiao Liu, Ji Zhilong, Yuan Gao, Wu Zhongqin, Zuo Wangmeng. Orthogonal Jacobian Regularization for Unsupervised Disentanglement in Image Generation//Proceedings of the IEEE International Conference on Computer Vision. Montreal, Canada, 2021:6701-6710

[25] Hadsell Raia, Chopra Sumit, Le CunYann. Dimensionality Reduction by Learning an Invariant Mapping//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York, USA, 2006:1735-1742

[26] Peng Su, Wang Kun, Zeng Xingyu, Tang Shixiang, Chen Dapeng, Di Qiu, Wang Xiaogang. Adapting Object Detectors with Conditional Domain Normalization//Proceedings of the European Conference on Computer Vision. Glasgow, UK, 2020:403-419

[27] XuChang-Dong, ZhaoXing-Ran, Xin Jin, WeiXiu-Shen. Exploring Categorical Regularization for Domain Adaptive Object Detection//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Seattle, USA, 2020:11721-11730

[28] ZhangYixin, WangZilei, MaoYushi. RPN Prototype Alignment for Domain Adaptive Object Detector//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. virtual, 2021:12425-12434

[29] DengJinhong, Wen Li, ChenYuhua, DuanLixin. Unbiased Mean Teacher for Cross-Domain Object Detection//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. virtual, 2021:4091-4101

[30] Bengio Yoshua, Courville Aaron C., Vincent Pascal. Representation Learning: A Review and New Perspectives. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013,35(8):1798-1828

[31] Xi Chen, Yan Duan, Houthooft Rein, Schulman John, Sutskever Ilya, Abbeel Pieter. Info GAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets//Proceedings of the Advances in Neural Information Processing Systems. Barcelona, Spain, 2016:2172-2180

[32] Lin Zinan, Kiran Koshy Thekumparampil, Fanti Giulia, Oh Sewoong. InfoGAN-CR and ModelCentrality: Self-supervised Model Training and Selection for Disentangling GANs//Proceedings of the International Conference on Machine Learning. virtual, 2020:6127-6139

[33] Shen Yujun, Gu Jinjin, Tang Xiaoou, Zhou Bolei. Interpreting the Latent Space of GANs for Semantic Face Editing//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Seattle, USA, 2020:9240-9249

[34] Peebles William S., Peebles John, Zhu Jun-Yan, Efros Alexei

A., Torralba Antonio. The Hessian Penalty: A Weak Prior for Unsupervised Disentanglement//Proceedings of the European Conference on Computer Vision. Glasgow, UK, 2020:581-597

[35] Yujun Shen and Bolei Zhou. Closed-Form Factorization of Latent Semantics in GANs//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. virtual, 2021:1532-1540

[36] Peng Xingchao, Huang Zijun, Sun Ximeng, Saenko Kate. Domain Agnostic Learning with Disentangled Representations//Proceedings of the International Conference on Machine Learning. Long Beach, USA, 2019:5102-5112

[37] Yang Junlin, Dvornek Nicha C., Fan Zhang, Chapiro Julius, Ming De Lin, Duncan James S. Unsupervised Domain Adaptation via Disentangled Representations: Application to Cross-Modality Liver Segmentation//Proceedings of the Medical Image Computing and Computer Assisted Intervention. Shenzhen, China, 2019:255-263

[38] He Kaiming, Fan Haoqi, Wu Yuxin, Xie Saining, Girshick Ross B. Momentum Contrast for Unsupervised Visual Representation Learning//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Seattle, USA, 2020:9726-9735

[39] Ting Chen, Kornblith Simon, Norouzi Mohammad, Hinton Geoffrey E. A Simple Framework for Contrastive Learning of Visual Representations//Proceedings of the International Conference on Machine Learning. virtual, 2020:1597-1607

[40] Bo Sun, Li Banghuai, Cai Shengcai, Ye Yuan, Chi Zhang. FSCE: Few-Shot Object Detection via Contrastive Proposal Encoding//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. virtual, 2021:7352-7362

[41] Han Zongyan, Fu Zhenyong, Chen Shuo, Jian Yang. Contrastive Embedding for Generalized Zero-Shot Learning//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. virtual, 2021:2371-2381

[42] Hao Chen, WangYaohui, LagadecBenoit, DantchevaAntitza, BrémondFrançois. Joint Generative and Contrastive Learning for Unsupervised Person Re-Identification//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. virtual, 2021:2004-2013

[43] Singh Ankit, Chakraborty Omprakash, Varshney Ashutosh, Panda Rameswar, FerisRogério, SaenkoKate, DasAbir. Semi-Supervised Action Recognition With Temporal Contrastive Learning//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. virtual, 2021:10389—10399

[44] Feng Liu, Zhang Xiaoxong, Fang Wan, Ji Xiangyang, Ye Qixiang. Domain Contrast for Domain Adaptive Object Detection. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(12): 8227-8237.

[45] Lin Tsung-yi, Goyal Priya, Girshick Ross B., He Kaiming, Dollár Piotr. Focal Loss for Dense Object Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020,42(2):318-327

[46] Mohamed Ishmael Belghazi, Baratin Aristide, Rajeswar Sai, Ozair Sherjil, Bengio Yoshua, R. Devon Hjelm, Aaron C. Courville. Mutual Information Neural Estimation//Proceedings of the International Conference on Machine Learning. Stockholm smassan, Stockholm, Sweden, 2018:530-539

[47] Ranasinghe Kanchana, Naseer Muzammal, Hayat Munawar, Khan Salman H., Fahad Shahbaz Khan. Orthogonal Projection Loss//Proceedings of the IEEE International Conference on Computer Vision. Montreal, Canada, 2021:12313-12323

[48] Khosla Prannay, Teterwak Piotr, Chen Wang, SarnaAaron, Tian Yonglong, Isola Phillip, Maschinot Aaron, Liu Ce, Krishnan Dilip. Supervised Contrastive Learning//Proceedings of the Advances in Neural Information Processing Systems. virtual, 2020:1-13

[49] Aäron van den Oord, Li Yazhe, Vinyals Oriol. Representation Learning with Contrastive Predictive Coding. http://arxiv.org/abs/1807.03748, 2018, 07, 02

[50] Cordts Marius, Omran Mohamed, Ramos Sebastian, Rehfeld Timo, Enzweiler Markus, Benenson Rodrigo, Franke Uwe, Roth Stefan, Schiele Bernt. The Cityscapes Dataset for Semantic Urban Scene Understanding//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA, 2016:3213-3223

[51] Dai Dengxin, Sakaridis Christos, Hecker Simon, Luc Van Gool. Curriculum Model Adaptation with Synthetic and Real Data for Semantic Foggy Scene Understanding. International Journal of Computer Vision, 2020,128(5):1182-1204

[52] Geiger Andreas, Lenz Philip, Urtasun Raquel. Are we ready for autonomous driving? The KITTI vision benchmark suite//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. RI, USA, 2012:3354-3361

[53] Johnson-Roberson Matthew, Barto Charles, Mehta Rounak, Sharath Nittur Sridhar, Rosaen Karl, Vasudevan Ram. Driving in the Matrix: Can virtual worlds replace human-generated annotations for real world tasks? //Proceedings of the IEEE International Conference on Robotics and Automation. Singapore, 2017:746-753

[54] Zhu Jun-yan, Park Taesung, Isola Phillip, Efros Alexei A. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks//Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy, 2017: 2242-2251

[55] He Kaiming, Zhang Xiangyu, Ren Shaoqing, Jian Sun. Deep Residual Learning for Image Recognition//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA, 2016:770-778

[56] Wu Aming, Rui Liu, Han Yahong, Zhu Linchao, Yi Yang. Vector-Decomposed Disentanglement for Domain-Invariant Object Detection//Proceedings of the IEEE International Conference on Computer Vision. Montreal, Canada, 2021: 9322-9331

[57] Tian Kun, Zhang Chenghao, Ying Wang, Xiang Shiming, Pan Chunhong. Knowledge Mining and Transferring for Domain Adaptive Object Detection//Proceedings of the IEEE International Conference on Computer Vision. Montreal, Canada, 2021: 9113-9122

[58] Saito Kuniaki, Ushiku Yoshitaka, Harada Tatsuya, Saenko Kate. Strong-Weak Distribution Alignment for Adaptive Object Detection//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Long Beach, USA, 2019: 6956-6965

[59] Zhou Bolei, Khosla Aditya, Lapedriza Àgata, Oliva Aude, Torralba Antonio. Learning Deep Features for Discriminative Localization//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA, 2016: 2921-2929

[60] Laurens van der Maaten and Geoffrey Hinton. Visualizing Data using t-SNE. Journal of Machine Learning Research, 2008, 9 (86): 2579-2605

ZHONG An-Yu, M. S. candidate. His main research interests include deep learning and computer vision.

WANG Rui, Ph. D., professor. Her main research interests include deep learning and computer vision.

ZHANG Hua, Ph. D., associate professor. His main research interests include deep learning and computer vision.

ZOU Cong, Ph. D. candidate. His current research interests include computer vision and fine-grained recognition.

JING Li-Hua, Ph. D. candidate, engineer. Her main research interests include deep learning and computer vision.

Background

Object detection is an important task of computer vision, which plays a critical role in many fields closely related to daily life. However, the performance of existing object detection methods will decrease when the training data set is different from the actual application scenarios. Collecting and annotating data for each target scenario will greatly increase the cost, which limits the deployment and application of these methods in actual application scenarios.

Unsupervised domain adaptation object detection is developed to solve this problem. Its key is to align the feature distribution between the two domains so that the model can extract domain invariant features from the source domain and the target domain, so as to complete the detection task in the target domain. Early work used statistical feature alignment, witch constraint model extracted features with the same statistical properties from the source domain and the target domain as domain invariant features. The current method introduces adversarial learning by adding a domain classifier and gradient reverse layer to the network so that the feature extractor can learn domain invariant features and realize align the feature.

In this paper, we introduced feature disentangling into the domain adaptive object detection task. By improving the intra-domain consistency and inter-domain consistency of features, the discriminability and transferability of the model are improved. Compared with the baseline method, the mean Accuracy Precision (mAP) is improved by 3.1%, of which the improvement on some specific subclasses is up to 6%; Compared with the latest method, the mAP is improved by about 1%.

This work is supported in part by the National Natural Science Foundation of China Under Grants No. 62176253 and No. U20B2066, and Open Research Projects of Zhejiang Lab (NO. 2021KB0AB01). The opinions, findings, conclusions, and recommendations expressed in this paper are those of authors and do not necessarily reflect the views of the funding agencies or the government.