

基于跨域学习的单样本目标检测方法

冯佳伟, 褚晶辉, 吕卫*

天津大学电气自动化与信息工程学院, 天津 300072

摘要 针对单样本目标检测样本量较少的问题, 提出了一种基于跨域学习的方法。该方法从数据增强的角度出发, 增加其他域的数据集作为辅助, 增强网络学习能力, 同时为解决不同域间存在差异的问题, 提出了一种基于图片尺度和实例尺度的跨域学习算法, 分别对输入的图片特征与检测网络的候选特征增加域分类器模型, 用于增强网络对跨域数据的背景和目标的域适应能力。在两个不同的跨域场景进行实验, 其中在 PASCAL VOC 数据集上与目前主流的单样本目标检测算法进行比较, 超过目前最好算法 2.8 个百分点, 从而证明了本文方法可以有效提高单样本目标的检测性能。

关键词 卷积神经网络; 目标检测; 单样本学习; 跨域学习

中图分类号 TP389.1

文献标志码 A

DOI: 10.3788/LOP212819

One-Shot Object Detection Based on Cross-Domain Learning

Feng Jiawei, Chu Jinghui, Lü Wei*

School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China

Abstract This paper proposes a method based on cross-domain learning to address the problem of small sample sizes in one-shot object detection. The proposed method begins with the aim of data enhancement and progresses with the addition of datasets in other domains as auxiliaries to enhance the network learning capabilities, simultaneously a cross-domain learning algorithm based on image and instance scales is proposed to solve the problem of differences between domains. A domain classifier model is added to the input image features and candidate features of the detection network to enhance the background of the network to cross-domain data and the target domain adaptability. Experiments for two different cross-domain scenarios are conducted, the PASCAL VOC dataset is compared with current mainstream one-shot object detection algorithms, and it presents an improvement of 2.8 percentage points on the current best algorithm. This proves that the proposed method can effectively improve the performance of one-shot object detection algorithm.

Key words convolutional neural network; object detection; one-shot learning; cross-domain learning

1 引言

目标检测(Object detection)作为计算机视觉的一个分支,在自动驾驶、遥感监测^[1-2]等领域发挥了巨大作用。目前在目标检测领域使用的深度学习方法都需要大量数据作为支撑,然而现实中很多场景无法获得大量标注数据,如对特殊标志的检测。当样本数量不足时,现有的高性能网络通常会表现很差,网络出现过拟合导致对未知样本没有足够的泛化能力。因此,从少量的样本中学习物体特征并用于分类或其他计算机视觉场景是具有重要意义的。

单样本学习(One-shot learning)是在小样本学习的基础上发展起来的,但是样本的数量限定为一个,导

致单样本学习相比小样本学习更具有挑战性。目前,单样本学习的研究主要分为两个方向:第一个方向是从数据入手,对少量数据通过先验知识进行数据增强,通过对数据量的增加转化为一般问题;第二个方向是从模型结构来考虑,通过先验知识来限制模型复杂度,降低假设空间的大小,减少需要考虑的维度,使得少量的数据依然可以满足当前空间。

单样本学习开始是基于分类任务进行研究的,因为分类任务相比检测任务较为容易。随着单样本学习^[3]理论不断完善与算法不断优化,一些学者从不同角度对单样本目标检测进行研究。单样本目标检测将单样本学习应用到目标检测领域,在分类任务的基础上进一步提高了任务的困难程度。Fu等^[4]提出 OSCD

收稿日期: 2021-10-27; 修回日期: 2021-11-24; 录用日期: 2021-12-21; 网络首发日期: 2021-12-30

通信作者: *luwei@tju.edu.cn

网络,将单样本目标检测看作一种模板匹配,使用滑动窗口进行最佳匹配窗口的选择,主要利用了度量学习的方式来计算匹配关系的相似度。Chen等^[5]结合了多尺度检测提出一种两阶段检测框架,该框架在包含大量标注数据的源域数据集上训练模型,并在小规模目标域数据集上进行微调。Fan等^[6]提出了一种多关系模块检测器的小样本检测网络模型,该模型首先加入注意力机制,通过融合查询图像与支持图像的特征,在候选网络中引入支持数据,指导网络生成更多与查询图像相关的候选框并抑制无关的候选框;其次从局部、全局、互相关3个方面学习来自查询与支持图像的候选框关系。Karlinsky等^[7]从度量学习角度出发,提出一种可以在端到端的训练过程中同时学习主干网络参数、嵌入空间以及该网络中每个训练类别的多模态分布模型,并在小样本目标检测任务中验证了其有效性。Wang等^[8]提出了一种基于域适应的小样本目标检测算法,并应用到汽车目标检测场景中,使用对抗网络区分正负样本,以提高模型的鲁棒性。Fu等^[9]提出了一种通用的元学习目标检测框架Meta-SSD,该框架由一个元学习器和一个目标检测器构成,通过元学习器学习到一组通用的初始化参数,使用适当的参数更新策略,指导目标检测器学习新的类别从而实现小样本目标检测。

现存的单样本目标检测算法主要使用度量学习的方法,包括文献^[4-7]等。度量学习可以实现在单样本学习中对单一样本的充分利用,从而有效提高目标检测效果。在基于度量学习的框架上,可同时考虑从跨域学习角度对单样本目标检测任务进行增强。跨域学习使用的是其他任务下的真实数据集,与人为生成数据相比,通常有着更丰富的样本图片,但是由于任务类型不同,数据域的样本分布也不同,直接使用其他域的数据反而会因为域偏移误导网络的学习方向,从而降低网络的性能。跨域学习是使网络能够学习到不同数

据域之间的跨域信息,并对不同域之间的差异进行修正,使得网络能够更好地利用其他域的数据,提高目标检测等任务的准确率。

本文针对在单样本目标检测中使用跨域数据集带来的域偏移问题,提出了一种基于图片尺度和实例尺度的跨域单样本目标检测算法。在孪生网络的框架下,将目标图像和查询图像分别输入孪生网络框架进行特征提取,在图片尺度上构建目标图像和查询图像的域分类器,在图片尺度增强网络对于背景与目标的域适应能力,在实例尺度构建候选框域分类器,增强候选网络对类别偏差的鲁棒性。同时,为使得网络能准确地学习到不同域之间的偏移方向,使用了一种梯度反向传递的对抗学习方法,用于构建特征提取网络和域分类器之间的对抗关系,使特征提取网络能够对抗域分类器,从而提高域适应能力,增强网络对单样本目标的检测能力。

2 跨域单样本目标检测算法

提出了一种跨域单样本目标检测算法,将图片尺度和实例尺度的域适应方法应用到单样本目标检测框架上,具体的框架结构如图1所示,图中GRL为梯度反向层(Gradient reverse layer)。输入一张待检测目标图像 I 和一张仅带有类别信息的查询图像 Q ,然后经过特征提取网络分别提取目标特征 $\phi(I)$ 和查询特征 $\phi(Q)$,使用了ResNet-50^[10]作为基础特征提取网络。经过特征提取后,网络分为两个支路,一路使用候选框提取网络(Region proposal network, RPN)提取目标图像的候选框,并进行相应的池化和尺寸变化,将候选特征与查询特征 $\phi(Q)$ 进行融合并送入最终的分类网络和回归网络进行类别和检测框的预测。同时,网络的另一个分支用作域适应学习,包括对目标特征 $\phi(I)$ 和查询特征 $\phi(Q)$ 分别作图片尺度的域适应学习,并且对候选框特征进行实例尺度的域适应学习。

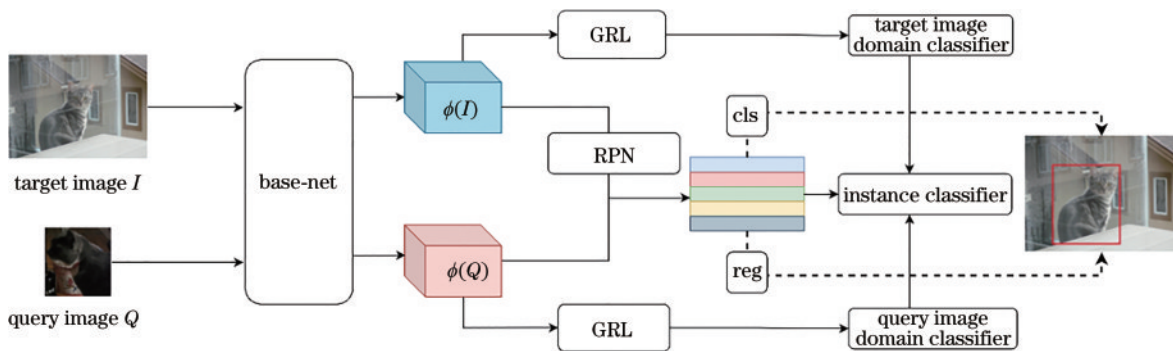


图1 跨域单样本目标检测算法框架

Fig. 1 Cross-domain one-shot object detection framework

2.1 域适应对抗训练方法

为将两个不同域的数据分布对齐,Ben-David等^[11]在2010年对不同域的一致性约束进行了一定研究,并

且使用一种 H 散度来定义源域 S 和目标域 T 之间的距离,对于来自源域的输入 x_S 和目标域的输入 x_T 建立一个域分类器 h , h 用来判断输入的目标是属于源域还是

目标域。将检测网络 f 与其结合,在保证 f 有较好的检测效果情况下,降低 H 散度,实现跨域学习的任务目标,具体可以表示为

$$\min_f D_H(S, T) = \max_f \min_h \{e_S[h(x_T)] + e_T[h(x_S)]\}, \quad (1)$$

式中: e_S 和 e_T 为域分类器 h 对于来自另一个域的分类

误差,即输入图片经过域分类器后,将其所属域判断错误的概率。式(1)代表的域对齐方式类似于对抗学习,即需要解决一个最大化最小值问题,在保证检测网络 f 最大的情况下,最小化 H 散度代表的域间距离,使得网络同时学习检测网络 f 和域分类器 h 所对应的参数。Ganin^[12]提出了 GRL 网络,将其与目标检测结合,实现跨域学习的目的,其结构如图 2 所示。

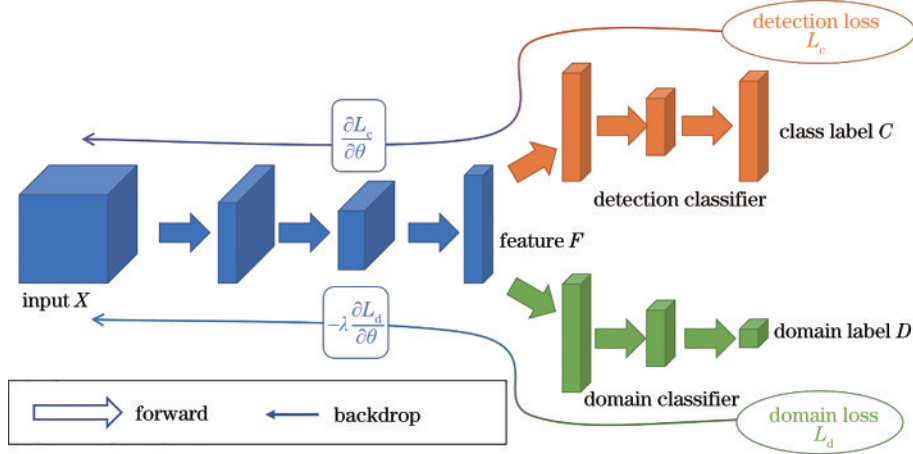


图 2 梯度反向网络结构

Fig. 2 Gradient reverse network

网络从输入图像 X 提取出特征 F ,分别送入检测器和域分类器,并分别使用检测损失函数 L_c 和域损失函数 L_d 作为目标函数。为了使网络有较好的检测效果,且对域标签的分类误差尽可能大,在反向传播时使用了不同的梯度下降方向,即对于类别分类器,使用正常的梯度传播偏导数 $\frac{\partial L_c}{\partial \theta}$,而对于域标签分类器,增加了一 λ 系数,其中 $\lambda \in (0, 1)$,使用 $-\lambda \frac{\partial L_d}{\partial \theta}$ 作为域分类器的反向传播偏导数。

GRL 梯度反向传递使网络可以在两个方向上进行优化,形成一种存在对抗关系的神经网络。在网络中加入 GRL 对抗学习,使得网络能够在保证检测效果的同时削弱对不同域的区别能力,让特征提取网络能够使两个不同域的数据有一定程度的趋近。

2.2 基于图片尺度和实例尺度的跨域学习方法

在图片尺度上使用两个不同的域分类器分别作用于 $\phi(I)$ 和 $\phi(Q)$,在进入域分类器之前,先经过一个 GRL,使 GRL 与特征提取网络之间进行对抗,使得图片尺度的域分类器能够最大化对域标签的分类。

使用 D_i 表示第 i 对训练输入的域标签,其中 $D_i = 0$ 代表输入来自源域, $D_i = 1$ 代表输入来自目标域。同时将目标图像的域分类输出表示为 p_i^1 ,将查询目标图像的域分类输出表示为 p_i^0 ,进而定义目标图像损失函数 L_1 和查询图像的损失函数 L_0 ,并使用交叉熵损失函数作为目标函数,计算方法如下:

$$L_1 = - \sum_i [D_i \text{lb} p_i^1 + (1 - D_i) \text{lb}(1 - p_i^1)], \quad (2)$$

$$L_0 = - \sum_i [D_i \text{lb} p_i^0 + (1 - D_i) \text{lb}(1 - p_i^0)]. \quad (3)$$

目标图像域分类器和查询图像域分类器都需要经过一个 GRL 网络进行反向梯度传播,从而使域分类器和分类检测网络产生有效的对抗效果,并使特征提取网络能够结合二者的融合目标函数得到最终的特征提取器。通过这两个域分类器的联合学习,即可以对源域和目标域的背景信息进行域迁移和域适应,并在检测未知类别时,能够使特征提取网络具有更好的类无关泛化能力。

对于实例尺度的域适应设计,对检测网络 RPN 输出的候选特征,使用域分类器进行实例尺度的域迁移学习,可以减少局部实例在不同域之间的差异。与图片尺度上的域适应设置一致,将第 i 张图片的第 j 个候选特征输入域分类器,得到的分类结果表示为 $p_{i,j}$,定义实例尺度上的损失函数为 L_R ,具体表示为

$$L_R = - \sum_{i,j} [D_i \text{lb} p_{i,j} + (1 - D_i) \text{lb}(1 - p_{i,j})]. \quad (4)$$

此外,设计了一个联合约束损失函数 L_C ,在保证各自分类器效果的同时,能在图片和实例之间产生一定一致性,使得域分类器可以在不同的尺度增强跨域学习的鲁棒性,具体表示为

$$L_C = \sum_{i,j} \|p_i^0 - p_{i,j}\|_2 + \sum_{i,j} \|p_i^1 - p_{i,j}\|_2, \quad (5)$$

式中:使用 l_2 范数计算图像和实例分类器的距离。本

文的网络框架使用基于 Faster R-CNN^[13] 的检测损失函数 L_D , 其包括了类别损失和检测框损失, 网络的总损失函数 L_{total} 表示为

$$L_{total} = L_D + \mu(L_1 + L_Q + L_R + L_C), \quad (6)$$

式中: μ 为超参数, 用于调整在检测网络和域适应分类器之间的平衡。

本文的域适应分类器的设置, 可以实现整个网络的端到端训练; 在反向传播时, GRL 网络用于改变反向传播时梯度下降的方向, 实现有效的学习。

使用 ResNet-50 前 3 层作为特征提取网络, 最后一层作为目标检测网络, 用于回归目标的类别与检测框位置, 域分类器使用了 3×3 卷积核与 1×1 卷积核的组合, 共 4 层卷积组合层, 最后输入全连接层得到域分类结果。

3 实验结果分析

实验环境: 软件使用 Python 3.6 和 PyTorch 1.6 框架搭建神经网络; 硬件使用 NVIDIA RTX 3090 显卡。实验中, 选用平均精度 (Mean average precision, mAP) 作为网络性能的评价指标。实验参数方面, 规定目标图像的尺寸大小为 608×608 , 查询图像的尺寸大小为 128×128 。在优化器选择方面, 使用了 Adam 梯度优化方法, 学习率在前 20000 次迭代时设置为 0.001, 在后 5000 次迭代时设置为 0.0002。在网络的

超参数方面, 式(6)中的 μ 设置为 0.3。

网络框架需要两张图像作为输入: 一张是查询图像; 一张是目标图像。查询图像需要只包含该类的目标, 在实验时需要数据集进行预处理, 使用已有的标签从完整图像中裁剪出查询图像。在测试时对不可见类的单个样本同样将目标裁剪为查询图像, 从而实现单样本的学习及利用。实验时使用数据集中的可见类作为训练集, 测试时使用不可见类作为测试集。

为了验证本文的跨域学习相比直接进行数据集增强的方法有明显的改进, 本节设置了两组数据集进行实验验证, 分别验证本算法对数据集环境差异的域偏移和虚拟图片与真实图片之间的域偏移的跨域学习能力。共使用 4 个数据集, 包括自动驾驶场景数据集 Cityscapes^[14]、Foggy Cityscapes^[15] 和 2 个通用类型数据集 PASCAL VOC^[16]、Clipart^[17]。

3.1 数据环境差异的域适应学习验证

使用 Cityscapes、Foggy Cityscapes 用于数据集环境差异的域适应验证。Cityscapes 是一个关于城市街景的语义级理解图片数据集, 包含来自 50 个不同城市的街景, 共 8 个类别, 其训练集包含 2975 张图片, 验证集包含 500 张图片。Foggy Cityscapes 同样是一个城市街景数据集, 在 Cityscapes 的基础上增加了天气信息, 使图片呈现出一种雾天的效果, 2 个数据集的图片效果对比如图 3 所示。



图 3 数据集示例图。(a) Cityscapes; (b) Foggy Cityscapes

Fig. 3 Diagrams of example datasets. (a) Cityscapes; (b) Foggy Cityscapes

将 Cityscapes 设置为源域 S , 将 Foggy Cityscapes 设置为目标域 T , 由于单样本目标检测的任务特殊性, 需要将数据集中的类别划分为可见类与不可见类, 将 8 个类别中的人、小汽车、火车作为不可见类, 其余 5 类作为可见类加入训练, 在训练时将目标域和源域的可见类作为训练数据送入网络, 分别为 1927 张和 3854 张, 并使用目标域中的不可见类作为测试数据, 共包含 418 张图片。

为了验证本文设计的图片尺度和实例尺度的跨域分类器能够有效地提高网络的域适应能力, 分别使用 L_1 、 L_Q 、 L_R 和 L_C 验证每个部分对网络性能的提升效果。网络不增加任何域适应方法进行训练, 即不对数据集域之间的差异做任何处理, 只是简单将源数据集和目标数据集作为联合数据集加入训练。

跨域学习在环境变化上关于不可见类别的域适应能力结果如表 1 所示。由表 1 中可知, 使用了式(6)的

表 1 跨域分类器在环境差异下的消融实验

Table 1 Cross-domain classifier ablation experiment under environmental differences

No.	L_1	L_Q	L_R	L_C	Person	Car	Train	mAP / %
1					28.6	44.2	28.6	33.8
2	✓				32.5	55.2	32.1	39.9
3		✓			33.6	53.1	34.2	40.3
4			✓		31.3	53.5	31.5	38.8
5	✓	✓			34.8	56.7	35.9	42.5
6	✓	✓	✓		35.1	57.6	36.3	43.0
7	✓	✓	✓	✓	35.8	57.1	36.9	43.3

设置相比基础网络的 mAP 指标提高了 9.5 个百分点,其中仅使用目标图像域分类器和查询图像域分类器分别能提升 6.1 和 6.5 个百分点,而仅使用实例尺度的域分类器能够提升 5.0 个百分点,说明经过本文设计的域分类器对抗学习,能够有效提升特征提取网络对于不同数据的域适应能力。同时在增加 L_C 后,相比不增加约束的网络平均提高了 0.3 个百分点,使用了图像和实例尺度上的联合约束,能够增强特征候选网络对类别的鲁棒性,即在 RPN 输出时,能将不同域间的类别差异信息进行一定程度的修正,提高输出候选框的

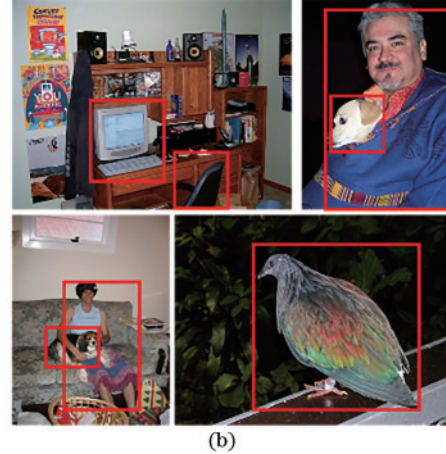
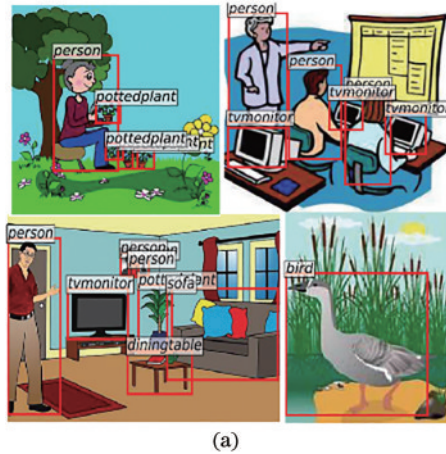


图 4 数据集示例图。(a) Clipart; (b) PASCAL VOC

Fig. 4 Diagrams of example datasets. (a) Clipart; (b) PASCAL VOC

首先针对 4 个不可见类别进行消融实验验证,结果如表 2 所示。本文提出的网络最好的平均精度达到了 71.0%,相比基础网络提高了 10.9 个百分点,在单独使用其中一个域分类器时,平均能提高 8 个百分点左右,说明域分类器能够有效地对不可见类进行域适应学习,同时使用 L_C 平均精度提高 0.3 个百分点。

在 PASCAL VOC 上与其他单样本目标检测算法进行比较,包括 SiamFC^[18]、SiamRPN^[19]、OSCD^[4]和 COAE^[20],实验结果如表 3 所示。可以看出本文使用跨域学习的方法,相比目前最好的算法平均提高了 2.8 个百分点。

准确性,使其能够更加接近真实目标。

3.2 虚拟图片与真实图片的域适应学习验证

随着深度网络和计算机图像学的进步,使用虚拟图片训练卷积神经网络已经逐步被应用到各个算法中。但计算机生成的虚拟图片相比真实环境中的图像较为简单,没有过多复杂的场景,且目标形状都较为规则,对于深度学习网络泛化真实环境中的图片是一个挑战。将本文设计的跨域学习分类器应用到虚拟图片和真实图片之间的域适应方向上,可有效地改善上述问题。

本文使用艺术图像数据集 Clipart 作为源域 S 以及 PASCAL VOC 作为目标域 T 。Clipart 数据集中的图片均为绘画风格,如图 4(a) 所示,其物体较为规则并且相比现实情况差距较大。PASCAL VOC 作为目标检测最通用的数据集之一,包含真实情况下的 20 类图片,如图 4(b) 所示。相比艺术图片的数据集 Clipart, PASCAL VOC 包含更多的信息,也会导致背景信息对目标检测任务的干扰。使用与文献[4]、文献[20]同样的实验设置,选择飞机、猫、牛、羊作为不可见类,将其余 16 类作为可见类。在实验时保证对于不可见类,无论是源域还是目标域,其样本均不会在训练时出现,确保不可见类是网络没见过的类别,其中训练集共 14056 张图片,测试集共 1357 张图片。

表 2 跨域分类器针对虚拟与现实图片的消融实验

Table 2 Cross-domain classifier ablation experiment for virtual and real pictures

No.	L_1	L_Q	L_R	L_C	Aero	Cat	Cow	Sheep	mAP / %
1					35.1	64.7	76.2	64.5	60.1
2	✓				46.2	79.2	80.1	69.2	68.7
3		✓			46.6	78.3	80.6	67.1	68.2
4			✓		45.1	78.9	79.5	68.9	68.1
5	✓	✓			49.5	80.5	81.2	69.8	70.3
6	✓	✓	✓		48.2	81.0	83.1	70.3	70.7
7	✓	✓	✓	✓	48.6	81.7	83.5	70.1	71.0

表 3 不同算法在 PASCAL VOC 数据集下的性能对比
Table 3 Comparison of different algorithms under the PASCAL VOC dataset

Method	Aero	Cat	Cow	Sheep	mAP / %
SiamFC	12.4	31.6	6.8	2.3	13.3
SiamRPN	3.5	21.7	15.9	15.7	14.2
OSCD	25.3	47.9	75.3	60.0	52.1
COAE	46.2	75.6	83.9	67.1	68.2
Ours	48.6	81.7	83.5	70.1	71.0

为了验证本文提出的图片尺度和实例尺度的跨域

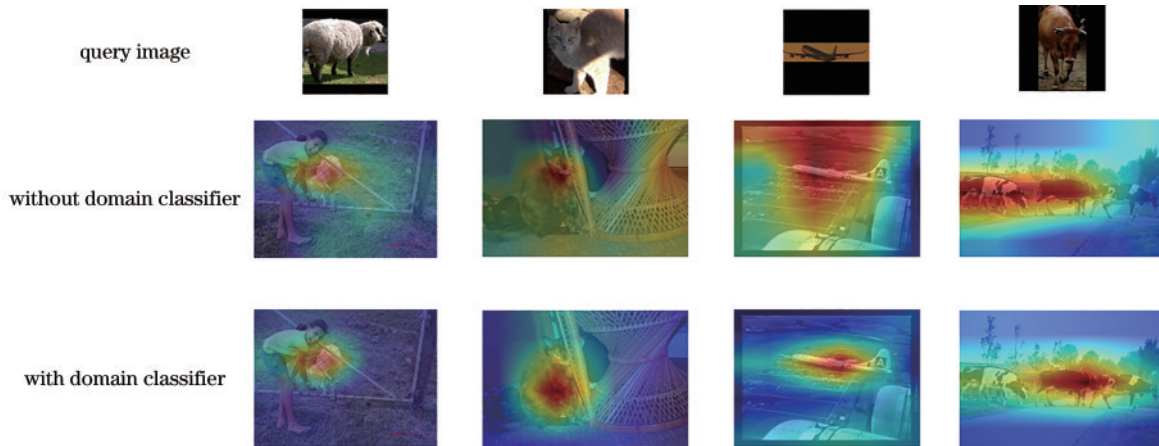


图 5 域分类器对网络候选框提取能力影响的对比

Fig. 5 Effect of domain classifier on network's feature extraction ability

4 结 论

从数据增强的角度出发,提出了一种基于图片尺度和实例尺度的跨域学习算法,并将其应用到单样本目标检测算法中。通过增加源域和目标域数据来增强单样本目标检测的表现;在图片尺度和实例尺度分别对不同域之间存在的域偏移进行域适应学习,图片尺度的域分类器增强网络对跨域数据背景与目标的适应能力,实例尺度的域分类器增强候选网络的鲁棒性,将两种域分类器同时应用到单样本目标检测中,提高对单样本的目标检测效果。在两个不同的跨域场景进行了实验验证,结果表明本文提出的跨域学习算法可以有效增强网络在单样本目标检测任务中对域偏移的适应能力。

参 考 文 献

- [1] 黎经元, 厉小润, 赵辽英. 融合空频域特征的光学遥感图像舰船目标检测[J]. 激光与光电子学进展, 2021, 58(4): 0415005.
Li J Y, Li X R, Zhao L Y. Ship target detection in optical remote sensing images based on spatial and frequency features[J]. Laser & Optoelectronics Progress, 2021, 58(4): 0415005.
- [2] 徐志京, 丁莹. 自适应旋转区域生成网络的遥感图像舰

学习能够提升在跨域数据集中对背景和目标的检测能力,在 PASCAL VOC 数据集中,将网络生成的候选框进行叠加,并转换成热力图展示在原图上,用于表示候选框提取网络对于待检测目标的提取能力。

是否使用域分类器对网络的跨域学习能力影响的对比如图 5 所示。由图 5 可知,使用跨域分类器后的网络对于候选框的选择更集中于所选类别。比如对羊类别的目标检测过程中,不使用跨域分类器的网络容易将人作为待检测目标进行标记,而使用了跨域分类器,能够使网络学习到跨域数据中不同类别之间的差别,使得检测框集中在羊上。

船目标检测[J]. 激光与光电子学进展, 2020, 57(24): 242805.

Xu Z J, Ding Y. Ship object detection of remote sensing images based on adaptive rotation region proposal network[J]. Laser & Optoelectronics Progress, 2020, 57(24): 242805.

- [3] Vinyals O, Blundell C, Lillicrap T, et al. Matching networks for one shot learning[C]//Advances in Neural Information Processing Systems 29, December 5-10, Barcelona, Spain. [S.l.: s.n.], 2016: 3630-3638.
- [4] Fu K, Zhang T F, Zhang Y, et al. OSCD: a one-shot conditional object detection framework[J]. Neurocomputing, 2021, 425: 243-255.
- [5] Chen H, Wang Y L, Wang G Y, et al. LSTD: a low-shot transfer detector for object detection[EB/OL]. (2018-03-05)[2021-04-08]. <https://arxiv.org/abs/1803.01529>.
- [6] Fan Q, Zhuo W, Tang C K, et al. Few-shot object detection with attention-RPN and multi-relation detector [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 4012-4021.
- [7] Karlinsky L, Shtok J, Harary S, et al. RepMet: representative-based metric learning for classification and few-shot object detection[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York:

- IEEE Press, 2019: 5192-5201.
- [8] Wang T, Zhang X P, Yuan L, et al. Few-shot adaptive faster R-CNN[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 7166-7175.
- [9] Fu K, Zhang T F, Zhang Y, et al. Meta-SSD: towards fast adaptation for few-shot object detection with meta-learning[J]. IEEE Access, 2019, 7: 77597-77606.
- [10] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.
- [11] Ben-David S, Blitzer J, Crammer K, et al. A theory of learning from different domains[J]. Machine Learning, 2010, 79(1/2): 151-175.
- [12] Ganin Y, Lempitsky V. Unsupervised domain adaptation by backpropagation[C]//ICML'15: Proceedings of the 32nd International Conference on International Conference on Machine Learning-Volume 37, July 6-11, 2015, Lille, France. Cambridge: JMLR, 2015: 1180-1189.
- [13] 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.
- [14] Cordts M, Omran M, Ramos S, et al. The cityscapes dataset for semantic urban scene understanding[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 3213-3223.
- [15] 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.
- [16] Everingham M, van Gool L, Williams C K I, et al. The pascal visual object classes (VOC) challenge[J]. International Journal of Computer Vision, 2010, 88(2): 303-338.
- [17] Inoue N, Furuta R, Yamasaki T, et al. Cross-domain weakly-supervised object detection through progressive domain adaptation[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 5001-5009.
- [18] Bertinetto L, Valmadre J, Henriques J F, et al. Fully-convolutional Siamese networks for object tracking[M]//Hua G, Jégou H. Computer vision-ECCV 2016 workshops. Lecture notes in computer science. Cham: Springer, 2016, 9914: 850-865.
- [19] Li B, Yan J J, Wu W, et al. High performance visual tracking with Siamese region proposal network[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 8971-8980.
- [20] Hsieh T I, Lo Y C, Chen H T, et al. One-shot object detection with co-attention and co-excitation[EB/OL]. (2019-11-28)[2021-10-27]. <https://arxiv.org/abs/1911.12529v1>.