

基于无监督域适应的低空海面红外目标检测

宋子壮*, 杨嘉伟, 张东方, 王诗强, 张越

北京遥感设备研究所, 北京 100854

摘要 提出一种基于无监督域适应的低空海面红外目标检测方法。首先利用图像翻译网络将源域图像翻译为目标域图像并共享标签。其次在 YOLOv5s 目标检测网络中使用梯度反转层优化网络提取特征的域间适应性。此外利用最大均值差异损失进一步缩小从网络中提取的不同红外探测器图像的特征分布。最后采用 AdamW 异步更新优化算法进一步提高模型在训练过程中的稳定性与检测精度。将所提方法在不同红外探测器采集的低空海面红外船只与无人机数据集中进行实验。实验结果表明,相较于传统有监督学习方法,所提方法有效降低了人工标注成本,且源域检测精度提高 6.56 个百分点,目标域检测精度提高 2.62 个百分点,有效提升目标检测模型在不同红外探测器间的泛化能力。

关键词 机器视觉; 红外探测器; 无监督域适应; 梯度反转层; 稳定训练; 目标检测

中图分类号 TP391.4

文献标志码 A

doi: 10.3788/AOS202242.0415001

Low-Altitude Sea Surface Infrared Object Detection Based on Unsupervised Domain Adaptation

Song Zizhuang*, Yang Jiawei, Zhang Dongfang, Wang Shiqiang, Zhang Yue

Beijing Institute of Remote Sensing Equipment, Beijing 100854, China

Abstract A low-altitude sea surface infrared object detection method based on unsupervised domain adaptation is proposed. First, the source domain images are translated into target domain images by image translation network, and the labels are shared. Second, the gradient reversal layer is used in YOLOv5s object detection network to optimize the inter-domain adaptability of feature extraction. In addition, the maximum mean discrepancy loss is used to further narrow the feature distribution of different infrared detector images extracted from the network. Finally, AdamW asynchronous update optimization algorithm is adopted to further improve the training stability and detection accuracy. The proposed method is tested on low-altitude sea surface infrared ships and unmanned aerial vehicles collected by different infrared detectors. Experimental results show that compared with the traditional supervised learning method, the proposed method effectively reduces the cost of manual labeling, and detection accuracy of source domain and target domain are improved by 6.56 and 2.62 percentage points respectively, which effectively improves the generalization ability of the object detection model between different infrared detectors.

Key words machine vision; infrared detector; unsupervised domain adaptation; gradient reversal layer; stable training; object detection

1 引言

深度学习网络因其具有强大的特征提取能力而被广泛应用,但其易受噪声的干扰,甚至微小的噪声扰动都会影响最终效果^[1-2]。与可见光探测器不同,

红外探测器因受生产工艺等限制而存在更加明显的非均匀性差异,则非均匀噪声所带来的影响更需要关注,故对目标检测网络的泛化能力提出了更高的要求。同时,在实际项目应用中,为了保证深度学习网络推理的实时性,通常会使用轻量级网络^[3-4],但

收稿日期: 2021-07-02; 修回日期: 2021-08-07; 录用日期: 2021-08-20

通信作者: *strongerzz@163.com

其参数量和网络层数较重量级网络少,且提取出的特征表达能力有限。此外,低空海面的光照条件与环境温度时刻变化,致使深度学习网络在不同红外探测器中应用对目标特性的不确定性泛化能力不足。上述情况使得由单一红外探测器所采集的图像训练出的目标检测模型,仅在对应探测器中具有良好的检测效果。为了提高模型在更多红外探测器中的泛化能力,需要对每套红外探测器所采集的图像进行人工标签标注,并通过训练对模型参数进行优化。当多套探测器面临相同检测任务时,这会造成许多不必要的人工标注成本。对此,如何对网络已学到的红外探测器图像特征进行迁移、深入挖掘网络与不同红外探测器数据之间的联系、减少人力标注成本、提升检测网络对不同红外探测器间的泛化能力以及在低空海面背景下获得良好的目标检测效果至关重要。

迁移学习是解决源域数据集与目标域数据集之间存在分布差异的方法,根据目标域有无标签可以分为有监督域适应和无监督域适应^[5]。域适应是一种学习域间样本的共同特征而进行迁移学习的方法,用于提高模型在不同域间的泛化能力。与传统预训练一个模型,后输入新数据来微调网络参数的迁移学习方法不同^[6],域适应方法因域监督信息的加入,从而拥有更好的分类效果,易于作到跨度更大的域间适应。无监督域适应方法不仅解决了深度学习网络在不同域中适应性的问题,同时也解决了监督学习所需图像标签较多、标签标注成本较大的问题。Tzeng 等^[7]利用最大均值差异(MMD)衡量域间距离,通过端到端训练使得 MMD 最小,从而完成单核 MMD 域适应图像分类。Long 等^[8]将 MMD 引入网络后有更多特征的层级,使用多核 MMD 获得了更好的域适应图像分类效果。Sun 等^[9]利用深度学习网络提取特征的二阶统计量距离,设计了新的损失函数与正则项,用于减小源域与目标域间的特征距离。Ganin 等^[10]利用梯度反转层(GRL)并引入对抗思想,使得网络在不同域的数据中提取的特征无法得到区分,同时又能很好地完成图像分类任务,从而完成域适应。Zhang 等^[11]从理论层面分析了无监督域适应图像分类问题,提出了一种基于边界差异差距的无监督域适应方法,利用梯度反转层进行域适应图像分类。滕文秀等^[12]提出了一种基于生成对抗网络的高分辨率遥感影像跨域分类方法,利用对抗学习最小化域间的分布差异,后进行无监督域适应图像分类。

相较于域适应图像分类,域适应目标检测因需要进行定位而更具有挑战性。Chen 等^[13]使用梯度翻转层对齐图像级和实例级特征实现了基于 Faster-RCNN(Faster Region Convolution Neural Networks)^[14]的端到端域适应目标检测。Inoue 等^[15]提出了一种基于伪标签与 CycleGAN(Cycle Generative Adversarial Networks)^[16]的域适应弱监督目标检测网络,在多种常用的目标检测网络中证明其泛化能力。Tzeng 等^[17]基于 CYCADA 域适应对抗网络^[18]提出一种无监督学习方法,对 CYCADA 的结构进行了精简,只使用一个生成对抗损失和语义一致损失就可取得很好的效果,训练速度有所提升。Rodriguez 等^[19]提出了一种基于自适应实例归一化(AdaIN)^[20]图像风格迁移技术的低层级差别适应与高层级特征适应的域适应目标检测算法,相比于基于 CycleGAN 的域适应方法取得了更好的效果。Kim 等^[21]提出了一种解决源域偏移及域间对译不完善的无监督域适应目标检测方法,方法中设计了域多样化模块和多域不变表示学习模块,而且该方法在多个数据集中均有检测精度的提升。Wang 等^[22]将通道注意力机制拓展到对不同域的注意力机制,从而为不同域的数据集赋予不同的权重,实现对通用目标的检测。蒋宁等^[23]在 SSD 目标检测网络^[24]的全局和局部层级分别配置相应的域适配器以减少域间差异,通过一致性正则化手段来稳定模型训练过程。吴泽远等^[25]基于 CycleGAN 设计了一种域适应目标检测方法,该方法能够进行图像翻译与目标检测的端到端训练,简化了域适应流程。

无监督域适应技术多用在可见光图像中,少有研究将其用于不同红外探测器之间提升模型在不同红外探测器应用中的泛化能力。本文提出了一种基于单阶段轻量级目标检测网络 YOLOv5s 的无监督域适应目标检测网络,该网络有助于降低低空海面背景下不同红外探测器的非均匀性对轻量级检测网络的影响,同时提升网络在不同红外探测器应用中的泛化能力,减少多套红外探测器面对相同检测任务的图像标注成本。

2 无监督域适应目标检测原理

定义源域数据集 $D_s = \{x_s^{(i)}, y_s^{(i)}\}, i = 1, 2, \dots, n_s$, 其中 $x_s^{(i)}$ 为服从源域数据分布 X_s 的样本, $y_s^{(i)}$ 为对应标签, n_s 为源域样本个数, 则无监督域适应目标域数据集定义为 $D_t = \{x_t^{(j)}\}, j = 1, 2, \dots, n_t$, 其

中 $x_i^{(j)}$ 为服从目标域数据分布 X_i 的样本, n_i 为目标域样本个数。目标域泛化误差的上界^[26]可表示为

$$\epsilon_t(h) \leq \epsilon_s(h) + d + \epsilon_\lambda(h), \quad (1)$$

$$d = \sup_{h, h' \in \mathcal{H}} |E_{x \sim P(X_s)}[h(x) \neq h'(x)] - E_{x \sim P(X_t)}[h(x) \neq h'(x)]|, \quad (2)$$

$$\epsilon_\lambda(h) = \min_{h \in \mathcal{H}} [\epsilon_s(h) + \epsilon_t(h)], \quad (3)$$

式中: \mathcal{H} 为假设空间; h 和 h' 为假设空间内的不同假设; $\epsilon_s(h)$ 为源域泛化误差; $\epsilon_t(h)$ 为目标域泛化误差; d 为源域 X_s 与目标域 X_t 间的差异; $\epsilon_\lambda(h)$ 为最小联合泛化误差; $E[\cdot]$ 为期望值; $\sup(\cdot)$ 为上确界。

深度学习的单阶段目标检测任务需要同时对目标的位置及类别进行预测, 可将问题表示为学习后验概率 $P(B, C | I)$, 其中 I 为待检测图像, B 和 C 分别为目标边界框的位置与类别。令 $P_s(B, C, I)$ 与 $P_t(B, C, I)$ 分别为源域与目标域的联合概率分布, 且存在领域漂移, 即 $P_s(B, C, I) \neq P_t(B, C, I)$ 。由贝叶斯公式可得

$$P_s(B, C | I)P_s(I) \neq P_t(B, C | I)P_t(I). \quad (4)$$

根据协变量偏移假设^[27]可知 $P_s(B, C | I) =$

$P_t(B, C | I)$, 则完成域适应需要对齐边缘概率分布, 即 $P_s(I) = P_t(I)$ 。本文通过减少领域间差异 d 与获取最小联合泛化误差 $\epsilon_\lambda(h)$ 两部分进行领域适应。对于减少领域间差异 d , 从像素级与特征级两部分入手, 对源域边缘概率分布 $P_s(I)$ 与目标域边缘概率分布 $P_t(I)$ 进行对齐。首先, 利用图像翻译网络将源域图像翻译为目标域图像, 使翻译后的图像符合目标域的像素级分布, 然后进行像素级域间的对齐。其次, 共享源域图像与翻译后图像的标签, 使检测网络能够进行无监督训练, 减少目标域的标签标注成本。最后, 利用所设计的梯度反转层与最大均值差异损失, 对检测网络在不同深度提取的源域图像特征与目标域图像特征进行优化。梯度反转层通过混淆域分类器来促使检测网络能够在域监督信息的指引下寻找域间的共同特征, 而最大均值差异损失能够在统计意义上进一步对齐域间特征分布, 从而进行特征级域间的对齐。对于减少 $\epsilon_\lambda(h)$, 利用源域数据与图像翻译网络所生成的目标域数据, 通过共享标签进行联合训练可以获得最小联合泛化误差 $\epsilon_\lambda(h)$ 。本文设计的无监督域适应目标检测网络结构如图 1 所示。

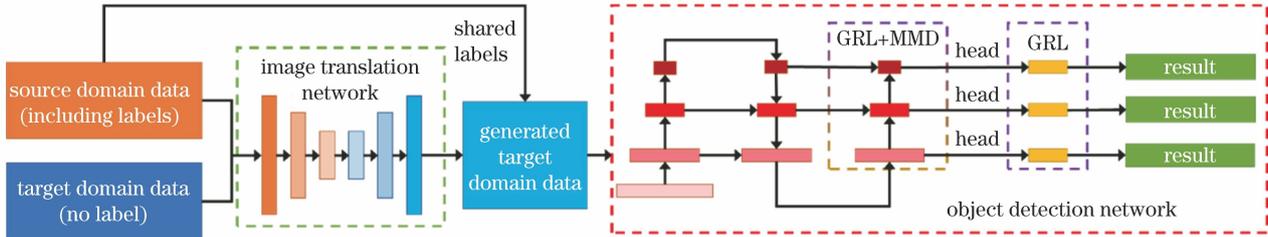


图 1 无监督域适应目标检测网络结构

Fig. 1 Structure of unsupervised domain adaption for object detection network

3 本文方法

3.1 CycleGAN 图像翻译网络

CycleGAN 图像翻译网络利用两个生成对抗网络来构成对偶结构, 使源域红外图像与目标域红外图像循环对译, 并通过循环一致性损失来确保图像翻译的准确性。CycleGAN 网络为无监督训练网络, 不需要任何标签辅助, 且数据集不受配对图像需求的限制。利用图像翻译网络将源域图像翻译为符合目标域像素级分布的图像, 然后进行像素级域间的对齐。同时, 源域图像与翻译后的目标域图像共享标签, 可以减少目标域标签标注成本。图像翻译网络的总损失函数为

$$L_{\text{CycleGAN}} = \lambda_{\text{GAN}} L_{\text{GAN}, S \rightarrow T} + \lambda_{\text{GAN}} L_{\text{GAN}, T \rightarrow S} + \lambda_{\text{cyc}} L_{\text{cyc}} + \lambda_{\text{ident}} L_{\text{ident}}, \quad (5)$$

式中: $L_{\text{GAN}, S \rightarrow T}$ 和 $L_{\text{GAN}, T \rightarrow S}$ 分别为源域与目标域间的生成对抗损失, 用于衡量生成图像与原图像的相似程度, λ_{GAN} 为对应权重; L_{cyc} 为循环一致性损失, 用于限制两个翻译网络的重构误差后进行像素级对齐, λ_{cyc} 为对应权重; L_{ident} 为正则化损失, 用于保证翻译图像色彩的一致性, λ_{ident} 为对应权重。

3.2 基于 YOLO 的无监督域适应目标检测

基于 YOLO 的域适应目标检测网络结构如图 2 所示, 其中 GRL 与 MMD 的损失仅在训练过程中使用, 故不会对网络推理时间造成任何影响。

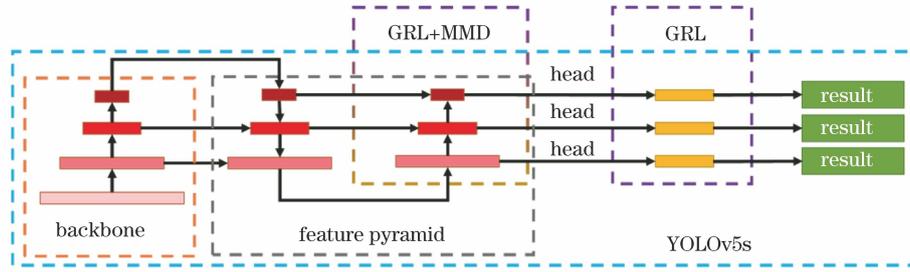


图 2 基于 YOLO 的域适应目标检测网络的结构

Fig. 2 Structure of YOLO-based domain adaption for object detection network

3.2.1 YOLO 目标检测网络

YOLOv5 在 YOLOv4^[28] 的基础上结合更加多样的训练手段与 EfficientDet 目标检测网络^[29] 中的复合缩放方法,使得网络在检测精度、推理速度与模型参数量之间动态权衡。为了保证网络在实际应用过程中的推理速度,选用轻量级 YOLOv5s 网络,其网络结构如图 2 所示,网络主要包括骨干网络、特征金字塔以及预测头。采用多任务损失函数对轻量级 YOLOv5s 网络进行训练,包括坐标损失、置信度损失以及类别预测损失。目标检测网络的总损失函数为

$$L_{\text{Det}} = \lambda_{\text{coord}} L_{\text{coord}} + \lambda_{\text{conf}} L_{\text{conf}} + \lambda_{\text{cls}} L_{\text{cls}}, \quad (6)$$

式中: L_{coord} 为坐标预测损失, L_{conf} 为置信度预测损失, L_{cls} 为分类预测损失,分别用于衡量预测框与真值框的定位偏差、置信度偏差与分类偏差; λ_{coord} 、 λ_{conf} 和 λ_{cls} 分别为对应部分损失的权重。

3.2.2 梯度反转层

本文在检测网络中特征金字塔与检测头的输出特征两部分设计了梯度反转层,通过混淆域分类器来实现域分类误差的最大化,使得模型在不同红外探测器的差异图像间寻找共同特征,从而完成领域适应,前向传播的表达式为

$$R(x') = x', \quad (7)$$

式中: x' 为输入特征图,输出与输入在前向传播中保持一致。反向传播的表达式为

$$dR(x')/dx' = -\lambda I, \quad (8)$$

式中: λ 为常参数,取值为 0.1; I 为经反向传播输入梯度反转层的域分类器损失梯度大小。

梯度反转层通过取反反向传播的梯度,可以在训练过程中使域间特征的分类损失最大化,使目标检测网络的检测损失最小化,最终实现域分类损失与检测损失间的平衡,从而促使检测网络提取出能够混淆分类器的域间最优特征,实现特征级域间的对齐。

3.2.3 最大均值差异

本文在检测网络的特征金字塔部分设计了最大

均值差异损失,并将其作为网络提取出的不同红外探测器图像特征分布差异的衡量指标,使网络提取出的源域特征分布与目标域相匹配,进一步提升域适应性能。设随机变量 x^* 的分布为 X , 随机变量 y^* 的分布为 Y , 若随机变量间的任意阶矩相同,则其分布一致。最大均值差异是衡量随机变量间任意阶矩的最大差异,将其作为衡量两种随机变量间的分布差异。最大均值差异的表达式为

$$M_{\text{MMD}}[F, X, Y] = \left[\frac{1}{m^2} \sum_{i=1}^m \sum_{j=1}^m f(x_i^*, x_j^*) - \frac{2}{mn} \sum_{i=1}^m \sum_{j=1}^n f(x_i^*, y_j^*) - \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n f(y_i^*, y_j^*) \right]^{\frac{1}{2}}, \quad (9)$$

式中: F 为高维空间映射函数集; $f(\cdot)$ 为映射函数; m 和 n 分别为两种分布的样本数量。

3.2.4 稳定训练

本文注意到梯度反转层的引入使得检测网络在训练过程中需要取得检测损失最小与分类损失最大的平衡,这种类似于对抗的思想,该思想会使网络训练容易变得不稳定,甚至出现梯度爆炸的情况。为此,在降低梯度反转层部分学习率与损失权重的同时,设计了一种异步更新优化算法。在训练域适应目标检测网络的过程中,首先更新梯度反转层部分的网络参数,优先使得检测网络提取出能够同时适应不同域的特征,然后更新检测网络参数,促使提取出的特征能够同时完成不同域的检测任务。异步更新优化算法使得检测模型在具有更好泛化能力的同时,训练更加稳定。此外,使用 AdamW^[30] 代替传统 SGD 的优化算法,使网络在训练过程中更快收敛。

3.2.5 损失函数

本文使用交叉熵损失对域分类器所输出的特征图进行逐像素点回归,表达式为

$$L_{\text{WE}} = \sum_{w=1}^W \sum_{v=1}^V \sum_{u=1}^U [-q_u \ln(p_u)]_{vw}, \quad (10)$$

式中: V 与 W 分别为特征图的宽高; U 为域类别总

数; p 为域分类器关于域真值标签 q 的预测值。

域分类损失可表示为

$$L_{\text{Dom}} = \frac{1}{M_{\text{FPN}}} \sum_{m'=1}^{M_{\text{FPN}}} L_{\text{CE},m'} + \frac{1}{M_{\text{HEAD}}} \sum_{n'=1}^{M_{\text{HEAD}}} L_{\text{CE},n'}, \quad (11)$$

式中: M_{FPN} 为特征金字塔的层数; M_{HEAD} 为用于检测的特征层数量。

使用(9)式中的最大均值差异距离,作为评估网络提取出的不同域图像特征的分布差异损失,并结合(6)式与(11)式,总损失函数可表示为

$$L_{\text{Total}} = \lambda_{\text{Det}} L_{\text{Det}} + \lambda_{\text{Dom}} L_{\text{Dom}} + \lambda_{\text{MMD}} L_{\text{MMD}}, \quad (12)$$

式中: λ_{Det} 为检测任务的损失权重; λ_{Dom} 为域分类任务的损失权重; λ_{MMD} 特征分布差异的损失权重。

4 实验分析

4.1 实验设置

实验使用 Pytorch 1.6 环境,在 CPU 为 Intel core i9-10900K、GPU 为 Quadro RTX 8000 (48 GB)、CUDA 11.0, CUDNN 8.0 和系统版本为 Ubuntu 20.04 LTS 的计算机上进行。图像翻译网络使用 Adam 优化方法,初始学习率为 10^{-4} ,总共进行 500000 次参数迭代,当训练至一半时学习率线性衰减至训练结束,训练批大小为 1。域适应目标检测网络使用 AdamW 优化方法和余弦退火学习率调整

策略进行训练,检测网络部分的初始学习率为 10^{-2} ,终止学习率为 10^{-3} ,梯度反转层部分的学习率为检测网络部分的 1/10。训练共计 300 个回合,训练批大小为 16。图像翻译网络部分损失权重的设置: $\lambda_{\text{GAN}} = 1, \lambda_{\text{cyc}} = 10, \lambda_{\text{ident}} = 10$ 。目标检测网络部分损失权重的设置: $\lambda_{\text{coord}} = 0.05, \lambda_{\text{conf}} = 1.0, \lambda_{\text{cls}} = 0.5$ 。总损失函数部分权重的设置: $\lambda_{\text{Det}} = 1, \lambda_{\text{Dom}} = 0.1, \lambda_{\text{MMD}} = 1$ 。

4.2 数据集与评估准则

数据集包含两个非均匀性差异较大的红外探测器所采集的低空海面船只与无人机,每个采集 5000 张图像,图像分辨率为 $640 \text{ pixel} \times 512 \text{ pixel}$ 。数据的标注格式采用 TXT 文本标注格式,内容依次为目标类型、归一化目标中心点坐标 (x, y) 以及归一化目标宽高 (w, h) 。训练集为含有标签的源域图像与不含标签的目标域图像,在目标域图像中进行结果测试,将在交并比阈值为 0.5 的情况下计算得到的 mAP(mean Average Precision)作为评价指标,用来验证所提方法的有效性。

4.3 实验结果

4.3.1 图像翻译结果

通过图像翻译网络,利用源域图像生成目标域图像,生成的目标域图像与源域共享标签如图 3 所示。

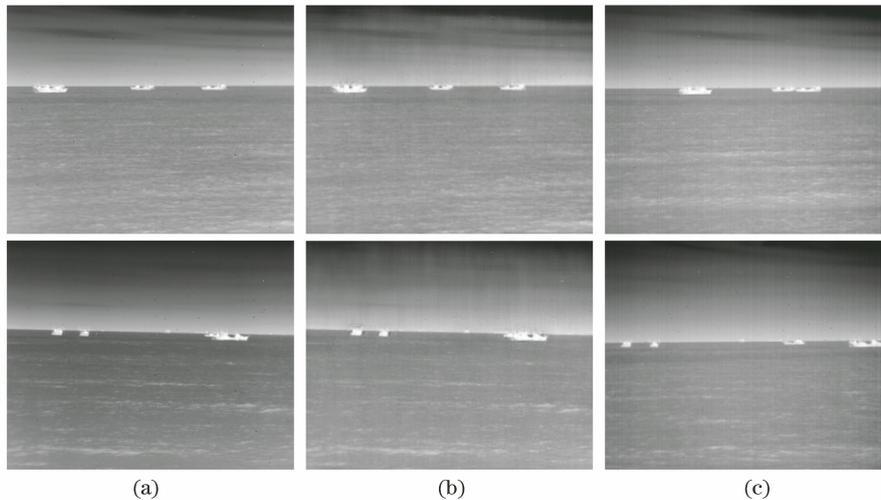


图 3 图像翻译结果。(a)源域图像;(b)生成的目标域图像;(c)目标域图像

Fig. 3 Image translation results. (a) Source domain image; (b) generated target domain image; (c) target domain image

4.3.2 梯度反转层位置

表 1 展示了在不同位置处加入梯度反转层后对域适应效果的影响,其中 w/o GRL(YOLOv5s)为不含有梯度反转层。从表 1 可以看出,梯度反转层提升了模型在目标域红外探测器中的泛化能力,

mAP 值提升了 5.50 个百分点;此外,在改进网络特征金字塔提取的全局目标特征的基础上,将检测头输出后的特征图引入梯度反转层中,能够进一步改进网络提取的局部目标特征,mAP 值提升 1.21 个百分点。

表 1 不同位置的梯度反转层对检测精度影响

Table 1 Influence of GRL in different positions on detection accuracy

GRL position	mAP / %
w/o GRL(YOLOv5s)	84.84
Feature pyramid	89.13
Feature pyramid+head	90.34

4.3.3 最大均值差异损失

表 2 展示了不同特征分布差异损失权重取值下的对比结果,可以看出 λ_{MMD} 取为 1.0 拥有最好的检测精度。对比表 1 与表 2 的结果可以看出,最大均值差异损失的引入使得目标域检测精度得到进一步提升,mAP 值提升了 1.84 个百分点。

表 2 不同 λ_{MMD} 取值对检测精度的影响

Table 2 Influence of different λ_{MMD} values on detection accuracy

λ_{MMD}	mAP / %
0.5	89.56
1.0	92.18
2.0	91.98

4.3.4 稳定训练

为了证明本文提出的稳定训练方法的有效性,图 4 展示了训练过程中损失值的变动情况。从图 4 可以看出,SGD 异步更新方式有效改善了 SGD 同步更新过程中训练不稳定的情况,损失曲线波动更小,而且训练结束后拥有更小的损失值,mAP 值提升了 2.73 个百分点。为了进一步加快训练过程中的收敛速度,使用 AdamW 优化方法对网络进行训练。AdamW 同步更新方式较 SGD 异步更新方式拥有更快的训练收敛速度,而 AdamW 异步更新方式拥有最快的训练收敛速度与最小的训练损失曲线波动,同时训练结束后拥有最小的损失值。表 3 展示了不同训练方法得到的域适应检测结果,可以看

表 4 源域与目标域的检测精度对比

Table 4 Comparison of detection accuracy between source domain and target domain

Serial number	Training dataset	Test dataset	Detection accuracy / %	
			Source domain	Target domain
1	Source domain	Source domain	94.32	90.31
2	Source domain	Target domain	81.94	84.84
3	Target domain	Target domain	87.33	86.64
4	Source domain+target domain	Source domain+target domain	89.26	89.56
5	Supervised domain adaptation	Target domain	95.85	92.44
6	Unsupervised domain adaptation	Target domain	95.82	92.18

到本文改进后的 AdamW 异步更新方式的 mAP 值较传统 SGD 同步更新方式提升了 4.07 个百分点,较 SGD 异步更新方式提升了 1.34 个百分点。

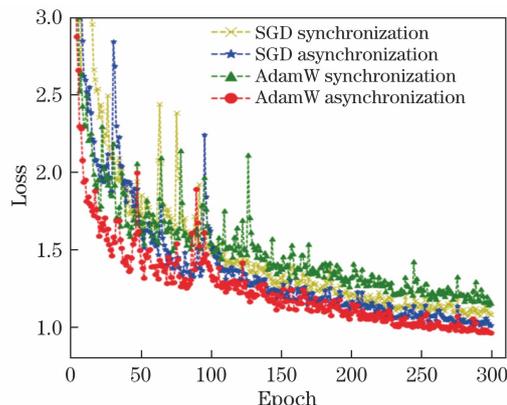


图 4 不同优化算法在训练过程中的损失值变动曲线

Fig. 4 Variation curves of loss values of different optimization algorithms during training

表 3 不同训练优化算法对检测精度的影响

Table 3 Influence of different training optimization algorithms on detection accuracy

Optimization algorithm	mAP / %
SGD synchronization	88.11
SGD asynchronization	90.84
AdamW synchronization	85.18
AdamW asynchronization	92.18

4.3.5 双域对比

表 4 展示了有监督训练方法与本文无监督域适应方法的训练对比结果。从表 4 可以看出,直接将不同红外探测器采集的数据进行混合训练可以使网络同时适应源域与目标域的数据(序号 4),但相较于域内单独训练(序号 1、3)检测精度无法得到保障性的提升,直接进行跨域训练的检测精度最低(序号 2);域适应手段的引入有效提升了网络对源域与目标域的泛化能力,且相较于有监督训练方法(序号

1~4)检测精度更高;相较于有监督方法(序号 4),本文方法(序号 6)的源域精度提高了 6.56 个百分点,目标域精度提高了 2.62 个百分点,相较于有监督域适应方法(序号 5)拥有相近的检测精度,同时省去了目标域红外图像数据的标注成本。有监督检测与无监督域适应检测的结果对比如图 5 所示。



图 5 有监督检测与无监督域适应检测的结果对比。

(a)有监督检测;(b)无监督域适应检测

Fig. 5 Comparison of results between supervised detection and unsupervised domain adaptation detection.

(a) Supervised detection; (b) unsupervised domain adaptation detection

5 结 论

本文提出了一种基于无监督域适应的低空海面红外目标检测方法。在 YOLOv5s 目标检测网络的基础上,引入梯度反转层与最大均值差异损失,这可以提升模型在不同红外探测器间的适应性。通过图像翻译网络,利用源域图像生成目标域图像同时共享标签,再进行无监督域适应目标检测。通过 AdamW 异步更新优化算法,进一步提高模型在训练过程中的稳定性与检测精度。结果表明,相较于传统有监督训练方法,本文方法的源域精度提高了 6.56 个百分点,目标域精度提高了 2.62 个百分点,有效提高目标检测网络在不同红外探测器间的泛化能力,同时减少了不同探测器

图像的标签标注成本。

参 考 文 献

- [1] Szegedy C, Zaremba W, Sutskever I, et al. Intriguing properties of neural networks[EB/OL]. (2013-12-21) [2021-06-25]. <https://arxiv.org/abs/1312.6199>.
- [2] Su J W, Vargas D V, Sakurai K. One pixel attack for fooling deep neural networks[J]. IEEE Transactions on Evolutionary Computation, 2019, 23(5): 828-841.
- [3] Zhu T Y, Huang L F, Dong F, et al. Infrared-remote-sensing ship detection based on lightweight residual network[J]. Acta Optica Sinica, 2020, 40(1): 0111018.
朱天佑, 黄凌峰, 董峰, 等. 基于轻量级残差网络的红外遥感船只检测[J]. 光学学报, 2020, 40(1): 0111018.
- [4] Wang W X, Fu Y T, Dong F, et al. Infrared ship target detection method based on deep convolution neural network[J]. Acta Optica Sinica, 2018, 38(7): 0712006.
王文秀, 傅雨田, 董峰, 等. 基于深度卷积神经网络的红外船只目标检测方法[J]. 光学学报, 2018, 38(7): 0712006.
- [5] Fan C N, Liu P, Xiao T, et al. A review of deep domain adaptation: general situation and complex situation[J]. Acta Automatica Sinica, 2021, 47(3): 515-548.
范苍宁, 刘鹏, 肖婷, 等. 深度域适应综述: 一般情况与复杂情况[J]. 自动化学报, 2021, 47(3): 515-548.
- [6] Yosinski J, Clune J, Bengio Y, et al. How transferable are features in deep neural networks? [EB/OL]. (2014-11-06) [2021-06-25]. <https://arxiv.org/abs/1411.1792>.
- [7] Tzeng E, Hoffman J, Zhang N, et al. Deep domain confusion: Maximizing for domain invariance [EB/OL]. (2014-12-10) [2021-06-25]. <https://arxiv.org/abs/1412.3474>.
- [8] Long M, Cao Y, Wang J, et al. Learning transferable features with deep adaptation networks[C]//International Conference on Machine Learning, July 6-11, 2015, Lille, France. Cambridge: PMLR, 2015: 97-105.
- [9] Sun B C, Saenko K. Deep CORAL: correlation alignment for deep domain adaptation[M]//Hua G, Jégou H. Computer vision-ECCV 2016 workshops. Lecture notes in computer science. Cham: Springer, 2016, 9915: 443-450.
- [10] Ganin Y, Lempitsky V. Unsupervised domain adaptation by backpropagation[C]//International Conference on Machine Learning, July 6-11, 2015, Lille, France. Cambridge: PMLR, 2015: 1180-1189.
- [11] Zhang Y, Liu T, Long M, et al. Bridging theory and

- algorithm for domain adaptation [C] // International Conference on Machine Learning, June 9-15, 2019, Long Beach, USA. Cambridge: PMLR, 2019: 7404-7413.
- [12] Teng W X, Wang N, Chen T S, et al. Deep adversarial domain adaptation method for cross-domain classification in high-resolution remote sensing images [J]. *Laser & Optoelectronics Progress*, 2019, 56(11): 112801.
滕文秀, 王妮, 陈泰生, 等. 基于深度对抗域适应的高分辨率遥感影像跨域分类 [J]. *激光与光电子学进展*, 2019, 56(11): 112801.
- [13] Chen Y H, Li W, Sakaridis C, et al. Domain adaptive Faster R-CNN for object detection in the wild [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: 3339-3348.
- [14] 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.
- [15] 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.
- [16] Zhu J Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks [C] // 2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 2242-2251.
- [17] Tzeng E, Burns K, Saenko K, et al. SPLAT: semantic pixel-level adaptation transforms for detection [EB/OL]. (2018-12-03) [2021-06-25]. <https://arxiv.org/abs/1812.00929>.
- [18] Hoffman J, Tzeng E, Park T, et al. CYCADA: cycle-consistent adversarial domain adaptation [C] // International Conference on Machine Learning, July 10-15, 2018, Stockholm, Sweden. Cambridge: PMLR, 2018: 1989-1998.
- [19] Rodriguez A L, Mikolajczyk K. Domain adaptation for object detection via style consistency [EB/OL]. (2019-11-22) [2021-06-25]. <https://arxiv.org/abs/1911.10033>.
- [20] Huang X, Belongie S. Arbitrary style transfer in real-time with adaptive instance normalization [C] // 2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 1510-1519.
- [21] Kim T, Jeong M, Kim S, et al. Diversify and match: a domain adaptive representation learning paradigm for 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: 12448-12457.
- [22] Wang X D, Cai Z W, Gao D S, et al. Towards universal object detection by domain attention [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: 7281-7290.
- [23] Jiang N, Fang J L, Yang Q. Global-local domain adaptive object detection based on single shot multibox detector [J]. *Journal of Computer Applications*, 2021, 41(2): 517-522.
蒋宁, 方景龙, 杨庆. 基于单点多盒检测器的全局-局部层级的域适应目标检测 [J]. *计算机应用*, 2021, 41(2): 517-522.
- [24] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector [M] // Leibe B, Matas J, Sebe N, et al. *Computer vision-ECCV 2016. Lecture notes in computer science*. Cham: Springer, 2016, 9905: 21-37.
- [25] Wu Z Y, Zhu M. End-to-end cross-domain object detection based on image style transfer [J]. *Computer Systems & Applications*, 2021, 30(1): 194-199.
吴泽远, 朱明. 基于图像风格迁移的端到端跨域目标检测 [J]. *计算机系统应用*, 2021, 30(1): 194-199.
- [26] 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.
- [27] Shimodaira H. Improving predictive inference under covariate shift by weighting the log-likelihood function [J]. *Journal of Statistical Planning and Inference*, 2000, 90(2): 227-244.
- [28] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: optimal speed and accuracy of object detection [EB/OL]. (2020-04-23) [2021-06-25]. <https://arxiv.org/abs/2004.10934>.
- [29] Tan M X, Pang R M, Le Q V. EfficientDet: scalable and efficient object detection [C] // 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 10778-10787.
- [30] Loshchilov I, Hutter F. Decoupled weight decay regularization [EB/OL]. (2017-11-14) [2021-06-25]. <https://arxiv.org/abs/1711.05101>.