

基于不变特征的多源遥感图像舰船目标检测算法

杨 曦¹, 张 鑫¹, 郭浩远¹, 王楠楠¹, 高新波^{1,2}

(1. 西安电子科技大学综合业务网理论与关键技术国家重点实验室, 陕西西安 710071;

2. 重庆邮电大学图像认知重庆市重点实验室, 重庆 400065)

摘 要: 由于域偏移的存在,多源图像舰船目标检测任务面临着不同源传感器带来的图像风格差异难题. 另外,为特定数据源训练特定的检测模型会消耗大量的计算资源,严重限制了其在军民领域的应用. 因此,设计一个通用网络以有效检测来自不同源遥感数据的舰船目标成了当下的研究热点. 针对该需求,本文提出了一种基于不变特征的通用舰船目标检测方法,通过充分利用多源数据之间的共享知识实现通用遥感目标的网络检测. 本方法由2部分组成:图像级的风格转换网络和特征级的域自适应网络. 具体地,前者采用风格转换网络生成接近真实分布的伪多源图像,拉近多源数据之间的分布,在图像层面上学习多源数据的不变特征;为学习特征层面上多源数据的不变特征,后者通过适应网络对多源特征进行信息解耦,通过域注意力网络的自适应权重分配实现特征重组. 本文在NWPU VHR-10, SSDD, HRSC和SAR-Ship-Dataset数据集上进行实验验证,结果表明:所提方法通过不变特征之间的信息互补,缓解了域偏移问题,可有效检测多源遥感数据. 本文方法在上述多源数据集上的平均mAP为90.8%,相比现有主流舰船目标检测方法可以提高1.4%~10.6%.

关键词: 舰船检测; 遥感图像; 深度学习; 风格转换; 域自适应

中图分类号: TN911.73 **文献标识码:** A **文章编号:** 0372-2112(2022)04-0887-13

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20210842

Invariant Features Based Ship Detection Model for Multi-source Remote Sensing Images

YANG Xi¹, ZHANG Xin¹, GUO Hao-yuan¹, WANG Nan-nan¹, GAO Xin-bo^{1,2}

(1. State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an, Shaanxi 710071, China;

2. Chongqing Key Laboratory of Image Cognition, Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

Abstract: Due to the domain shift, ship detection in multi-source data suffers from image variations caused by different source sensors. In addition, training a specific model for a particular data source consumes high computational cost, which severely limits its practical application in military and civilian fields. Therefore, designing a universal network to effectively detect ship objects from multi-source remote sensing images has become a research hotspot. To this end, the paper proposes a universal ship detection algorithm based on invariant features, which realizes a universal remote sensing object detection network by fully utilizing the shared knowledge among multi-source data. Our method mainly consists of two parts, i.e., an image-level style transfer network and a feature-level domain adaptive network. Specifically, the former employs style transfer network to generate pseudo-multi-source images that are close to the real distribution, narrow the distribution between multi-source data, and extract the invariant features of multi-source data at the image level; To extract invariant features at the feature level, the latter decouples the multi-source features through adaptive network, and realizes feature reorganization through adaptive weight allocation of domain attention network. We evaluate the proposed method using multiple datasets including NWPU VHR-10, SSDD, HRSC and SAR-Ship-Dataset. Experimental results show that the proposed method alleviates the problem of domain shift by complementing the information between invariant features, and can effectively detect multi-source remote sensing data. The average mAP of our method on the above-mentioned multi-source datasets is 90.8%, which exceeds 1.4%-10.6% compared with the existing mainstream ship object detection methods.

Key words: ship detection; remote sensing images; deep learning; style transfer; domain adaptation

收稿日期: 2021-07-05; 修回日期: 2021-09-26; 责任编辑: 王天慧

基金项目: 国家自然科学基金(No.61976166, No.62036007, No.61772402, No.62050175, No.61922066, No.61876142); 陕西省重点研发计划(No.2021GY-030); 陕西省创新人才推进计划(No.2020KJXX-027); 中央高校基本科研业务费(No.JB210115)

1 引言

遥感图像中舰船目标检测是海洋监管和国防建设中的重要基础工作^[1],旨在分类和定位一幅遥感图像中的所有舰船.近年来,随着遥感卫星技术的快速发展,多源遥感数据不断扩增.光学遥感图像具有判读直观性强及数据翻译便捷等优点,引起了广大学者的重视和研究.但是光学图像容易受到云雾遮挡、强烈光照等气候的影响.相对地,合成孔径雷达(Synthetic Aperture Radar, SAR)凭借其不依赖光照条件且不受天气影响的优势,能够实现全天时、全天候大面积海域侦查工作.由于上述差异性,针对不同成像机理的遥感图像,需要设计专门的算法进行舰船目标检测.

恒虚警率算法^[2-5]是传统舰船目标检测的经典方法,该算法通过对背景杂波统计建模来检测舰船目标,从而保持恒定的虚警率.然而,这类算法利用浅层弱语义信息进行预测,对复杂环境下的舰船检测效果不佳.近年来,人工智能领域迅速发展,卷积神经网络(Convolutional Neural Network, CNN)凭借其强大的特征提取能力,成为舰船目标检测的有力工具.目前基于CNN的检测器主要分为2类:(1)单阶段检测器,如EFGRNet^[6],CenterNet^[7]; (2)双阶段检测器,如Fast-

er R-CNN^[8], Libra R-CNN^[9].因此,众多学者开始研究基于深度学习的舰船目标检测系统.

尽管这些舰船检测算法已经取得重要进展,但通常是针对特定域的.图1是单源和多源遥感图像舰船目标检测器的对比图,其中紫罗兰色框表示预测结果.如图1(a)所示,由于不同源数据之间存在域偏移,现有舰船目标检测算法大多是有针对性地为各个数据源分别训练一个检测器用于对应数据的检测.若需要检测多源遥感图像,模型的总体尺寸会随着数据源的数量线性增加,造成资源浪费.此外,在实际应用中,若舰船数据来源模糊,无法为其选择特定的检测器.由于数据源的多样性以及它们之间存在着一定域偏移,错误地选择检测器会带来精度的急剧下降.而多源遥感图像检测器只需要训练一个通用的模型,在处理多源图像时只需要执行一次推理过程,如图1(b)所示.这使得多源遥感图像检测器能够节省计算资源和提升计算速度,更符合实际工程的实时检测需求,在解决模糊数据源检测问题的同时,也充分利用了多源数据之间的共享知识,实现了不同数据源之间的信息互补,进一步提高了检测性能.因此,如何使用一种方法同时检测来自不同源的舰船目标,即多源数据舰船目标检测,对于我国国防智能化建设具有重要研究意义.

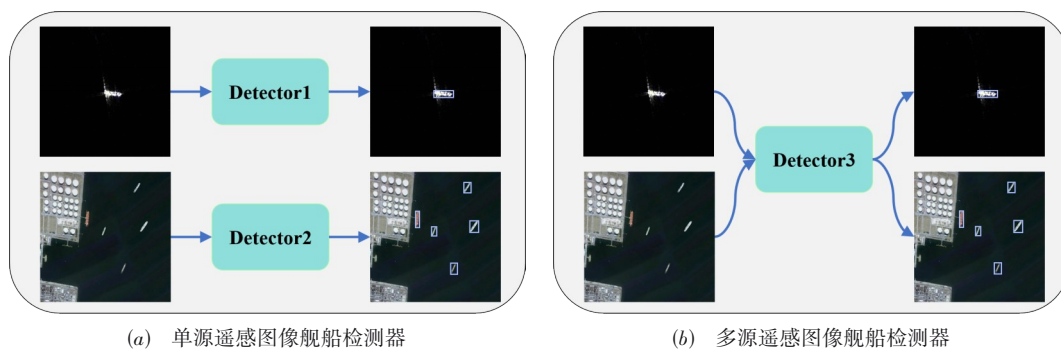


图1 单源和多源遥感图像舰船目标检测器对比图

本文提出了一种新颖的多源遥感图像舰船目标检测算法,基于多源舰船目标不变特征(不随域改变而影响舰船目标判别的特征,例如舰船形态、海陆中的分布位置等)之间的信息互补,使用一个通用网络检测不同源舰船目标并提升检测准确率,总体框架如图2所示.该方法由基于图像级的风格转换网络和特征级的域自适应网络两个模块组成.首先,为学习多源数据图像层面上的不变特征,利用循环生成对抗网络^[10](Cycle Generative Adversarial Network, Cycle GAN)将光学风格图像转换成伪SAR图像,并将SAR风格图像转换成伪光学图像.为了促进生成对抗网络的转换效果,引入循环一致性损失隐式相互约束.接着将生成的伪多源遥

感图像与原始遥感图像同时输入到主干网络以提取多源特征.最后,为学习多源数据特征层面上的不变特征,利用多个适应网络对多源特征进行解耦表示,同时由域注意力网络进行权重分配实现重组.本文提出的通用多源检测模型与单源检测模型相比,不需要感兴趣域的先验知识,所有的参数和计算都是跨域共享的.该模型通过提取图像级和特征级2个层面上具有判别性的不变特征,减轻不同域之间存在的域偏移,同时实现多源数据不变特征之间的信息互补,这些都是单源检测器所不具备的优势.实验结果表明,所提方法在多源舰船目标检测任务中的平均准确率高于其他前沿方法.

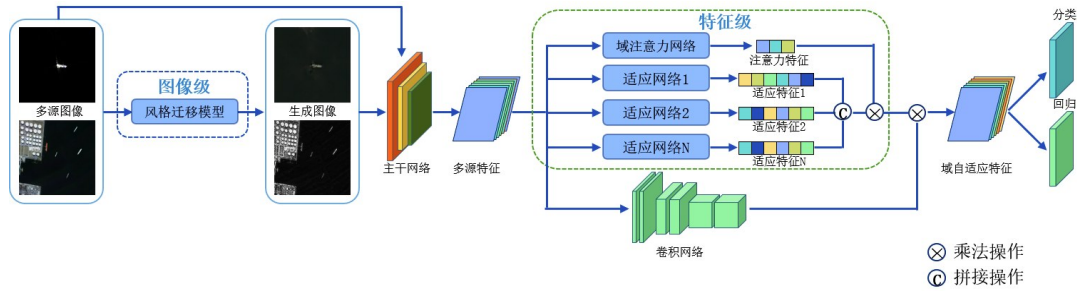


图2 多源遥感图像舰船目标检测框架

2 相关工作

本节从单源检测和多元检测两方面对相关工作进行了总结和回顾。单源检测部分首先介绍通用目标检测的代表性工作,然后聚焦舰船检测问题,介绍已有的舰船检测算法。多元检测部分首先总结了常用的多元检测算法,其次对图像风格转换相关工作进行了回顾。

2.1 单源检测

2.1.1 目标检测

作为计算机视觉的基础任务,目标检测旨在检测一幅输入图像中的所有特定目标。目前基于CNN的检测器大致分为2类:双阶段检测器和单阶段检测器。

(1)双阶段检测器首先利用区域建议框生成网络(Region Proposal Network, RPN)提取感兴趣区域,然后基于感兴趣区域进行分类和回归特定目标。Faster R-CNN^[8]通过增加RPN以扩展Fast R-CNN^[11],实现了首个基于深度学习的端到端双阶段检测器。为了提升检测器的精度,FPN^[12]采用一种自上而下的路径,同时通过横向连接进一步丰富了各层特征的语义信息。为了解决分类和回归任务之间的不平衡问题,IoU-Net^[13]采用精准的RoI池化层取代了原始的RoI池化层,并提出了IoU指导下的非极大值抑制算法。

(2)单阶段检测器直接在大量密集分布的锚框中检测特定的目标,而不需要依赖RPN。RetinaNet^[14]通过引入焦点损失解决了单阶段算法中正负样本不平衡的问题。为了满足实时检测需求,YOLO系列算法^[15,18]陆续被提出,以提升目标检测速度。近年来,无锚框检测器兴起,CornerNet^[19]将目标检测任务转换为关键点检测任务,通过预测检测框的左上角和右下角2对关键点进行定位,同时不再设置锚框的超参数。CenterNet^[7]直接预测每个目标的中心点坐标并且无需非极大值抑制等后处理,大幅度提升了检测速度。

2.1.2 舰船检测

近十年来,基于遥感图像的舰船目标检测越来越受到人们的关注。传统的舰船目标检测方法是基于恒虚警率算法的。为了引入更多的空间信息,Wang等

人^[2]提出了一种基于空间域的恒虚警率算法,该方法将像素的灰度和像素之间的相关性融合为一个特征,利用恒虚警率算法提取对应像素的融合特征以增强舰船检测效果。随后,Pappas等人^[3]提出利用超像素代替矩形滑动窗口来增强恒虚警率算法,旨在实现更好的背景波段目标过滤,降低舰船目标的误检率。为了有效检测复杂背景中的舰船目标,Huo等人^[4]首先通过最大稳定极值区域方法对输入图像进行预筛选以获得低计算复杂度的候选区域,然后利用加权信息熵评估候选区域与邻域之间的相似性,最后进行舰船目标检测。但是这些传统方法不能提取深层次语义信息的特征,因此无法有效检测多源遥感图像舰船目标。

由于卷积神经网络具有较强的判别特征提取能力,研究者们开始更加关注基于CNN的舰船目标检测算法。Cui等人^[20]提出了一种基于密集连接的特征金字塔网络,采用金字塔结构将卷积注意力模块与自上而下的每层特征密切连接,提升了SAR图像检测精度。为了快速定位多角度舰船目标并对其准确分类,Yang等人^[21]提出了一种端到端的精修单阶段检测器,通过引入特征细化模块以实现特征的重构和对齐。Fu等人^[22]提出了一种特征平衡和细化网络,采用注意力机制引导的平衡特征金字塔以平衡多层次特征,同时通过特征精修模块实现特征对齐。上述方法虽然针对单源遥感图像具有较好的检测准确率,但是不善于同时处理多源图像中的舰船目标。

不同源数据之间存在着信息鸿沟,上述方法由于缺少域自适应思想无法直接处理多源数据。为此,基于上述工作,学者们广泛研究了多源目标检测的相关算法。

2.2 多源检测

2.2.1 多源检测算法

多源目标检测首要解决的是不同源数据之间的域偏移问题。DA Faster R-CNN^[23]通过图像和实例层面的域迁移,构建了一个域自适应网络用于检测多源图像。但该方法没有关注特征层面域自适应表征对检测的影响。文献^[24]提出了全局特征弱对齐和局部特征强对

齐策略,通过不同域数据之间的特征对齐,拉近不同域的数据分布. Universal DA^[25]采用域注意力机制构建了一个通用检测器,能够使用单一网络处理所有域.但是上述方法未能考虑图像级别域自适应对检测的影响.由于风格转换领域的迅猛发展,可借助风格转换思想,从图像级层面上减小域偏移难题.

2.2.2 风格转换策略

风格转换一直是重要的研究领域,其目的是通过纹理特征的学习将图像从源域转换至目标风格域. Gatys 等人^[26]提出一种基于 CNN 的风格转换网络,将自然图像的内容和风格进行分离和重组,从而获得高质量的新图像. Berger 等人^[27]在图像生成中添加长期一致性约束,通过引入局部特征和全局特征的空间信息以生成具有全局对称性的纹理,并将纹理用于风格转换中. Gu 等人^[28]提出了一种渐进特征优化方法,通过排列特征图的空间位置实现任意图像风格的转换,避免局部风格转换失真并保持原图像的全局相似性.为缓解纹理尺度不匹配问题, Wang 等人^[29]提出了一种多模态卷积神经网络,该网络利用分层训练的策略分别训练颜色和亮度通道. Zhu 等人^[10]提出了一种基于循环一致损失风格转换网络,利用 2 个生成对抗网络实现非成对图像的风格转换. 与以往的方法主要考虑生成样本质量不同,本文旨在利用风格转换样本提升多源遥感图像舰船目标检测的性能.

3 图像级风格转换网络

来自不同源传感器的遥感图像具有背景复杂多变的特点,在分辨率、纹理和细节等方面各不相同.为了学习不同源传感器下数据的鲁棒性特征,需要采集大量多源遥感数据.但是在实际应用中,多源遥感数据获取成本高.为了降低成本并实现多源舰船目标检测,本文从图像风格适应角度出发利用 Cycle GAN 进行数据生成.风格转换网络具有以下优点:(1)实现数据自动扩增,解决数据匮乏问题;(2)生成数据同时具有不同源数据的特点,有助于网络学习数据之间的不变特征;(3)生成数据与原始数据内容信息保持一致,只是图像风格发生变化,因此无需人工标注,进一步降低了专家标注的成本.

3.1 网络结构

如图 3 所示,图像风格转换网络由左右两个镜像对称的子网络组成,每个子网络由 2 个生成器(G_1 和 G_2) 和 1 个判别器(D_1 或 D_2) 组成.为使图像风格转换网络学习到不同源传感器的鲁棒性特征,2 个生成器共享参数且以端到端学习的方式连接.其中一个生成器用来生成伪风格图像,而另一个生成器将伪风格图像重构为原始风格图像.因此,本网络的目标是学习 2 个映射函数 $G_1: A \rightarrow B$ 和 $G_2: B \rightarrow A$,从而实现 G_1 和 G_2 生成的图像和真实图像在数学分布上无法区分.其中,域 A 表示光学图像域,域 B 表示 SAR 图像域.特别地,在训练和测试过程中,均不需要 SAR 图像和光学图像是一一对应的,本网络做的仅仅是风格转换.

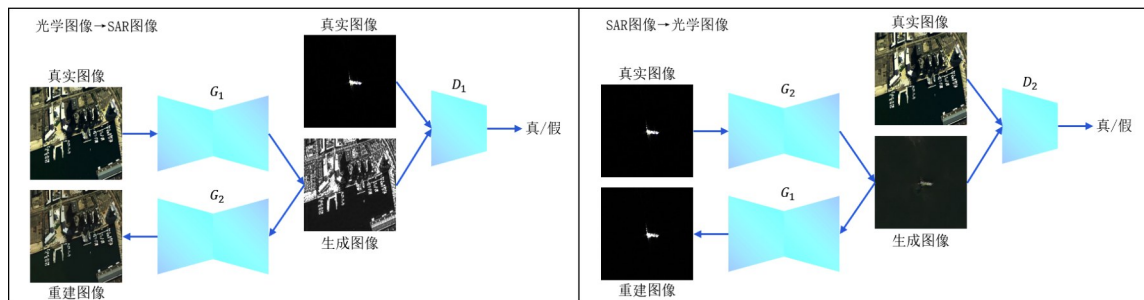


图3 风格转化网络

3.2 目标函数

2 个子网络损失函数形式相同,下面仅介绍图 3 左边子网络的损失函数.对于图像级风格转换网络,为保证生成图像能够保留更多的舰船信息,引入图像对抗损失来引导生成图像的分布接近目标域分布.对于映射函数 $G_1: A \rightarrow B$ 以及对应的判别器 D_1 ,图像对抗损失函数的定义为

$$\mathcal{L}_{\text{GAN}}(G_1, D_1, A, B) = \mathbb{E}_{b \sim P_{\text{data}}(b)} [\log D_1(b)] + \mathbb{E}_{a \sim P_{\text{data}}(a)} [\log (1 - D_1(G_1(a)))] \quad (1)$$

其中 \mathbb{E} 表示数学期望.为了使伪 SAR 风格图像转换为重建光学图像时舰船目标像素准确对齐,利用循环一致性损失函数进行约束,函数的定义为

$$\mathcal{L}_{\text{CYC}}(G_1, G_2, A, B) = \mathbb{E}_{a \sim P_{\text{data}}(a)} [\|G_2(G_1(a)) - a\|_1] + \mathbb{E}_{b \sim P_{\text{data}}(b)} [\|G_1(G_2(b)) - b\|_1] \quad (2)$$

其中 $\|\cdot\|_1$ 表示 L1 距离.因此,图像级风格转换网络的目标函数为

$$\mathcal{L}(G_1, G_2, D_1, A, B) = \mathcal{L}_{GAN}(G_1, D_1, A, B) + \lambda \mathcal{L}_{CYC}(G_1, G_2, A, B) \quad (3)$$

其中, λ 表示平衡因子, 用于平衡对抗损失任务和循环一致损失任务, 经验上, 取 $\lambda = 10$.

通过使用图像级风格转换网络, 训练集被扩充

为原始图像和风格转换图像的组合. 如图 4 所示, 经过循环一致性损失的约束, 每个生成图像保留了原始内容, 新样本和原始图像具有相同的标注. 因此, 可利用生成图像以及对应的相关标注来训练多源遥感图像舰船目标检测网络.

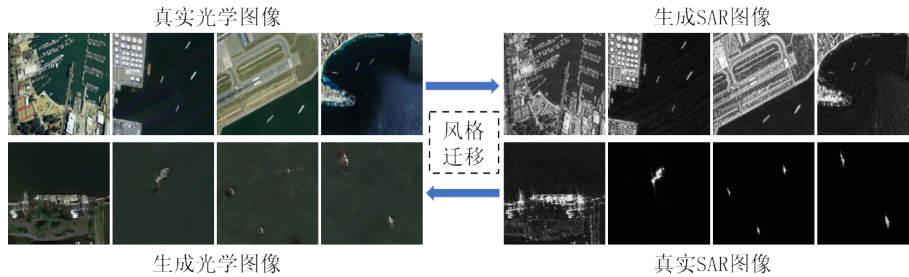


图 4 光学图像和 SAR 图像之间的风格转换示例

4 特征级域自适应网络

经过图像级风格转换网络后, 将原始图像和生成的风格转换图像相结合, 输入到后续的检测网络中. 利用特征级域自适应网络提取得到判别性特征, 包括利用适应网络来提取不同域的一致性表征以及利用域注意力网络进行权重分配, 从而实现多源信息的解耦以及重组. 同时, 引入焦点损失函数和 $L1$ 损失函数来约束检测网络的分类和回归任务.

4.1 网络结构

为了提取含有丰富语义信息的一般性特征, 本文提出一种域自适应网络(图 5). 域自适应网络分为适应网络和域注意力网络 2 个模块. 首先利用多个适应网络实现多源特征的解耦, 分别提取输入数据的适应性特征, 不同的适应特征反映了其在不同域空间特征表示; 然后使用域注意力网络计算所有域空间对应的权重向量, 实现对适应特征新的线性组合; 最后利用矩阵运算实现权重的分配和多源信息的一致性空间表征.

4.1.1 适应网络

本文利用迭代深层聚合 (Deep Layer Aggregation, DLA) 网络^[30] 作为特征提取的主干网络. 基于 SE (Squeeze-and-Excitation) 模块^[31] 的适应网络能够表征多源信息. 每个适应网络具有相同的结构, 包含 2 个全连接 (Full Connection, FC) 层以及 2 个激活层 (ReLU 函数和 Sigmoid 函数). 适应网络的输入是 DLA 网络提取的多源特征 x . 其计算过程可以表示为

$$\delta_i = f_{SE}(f_{GAP}(x)) \quad (4)$$

$$\Gamma = [\delta_1, \delta_2, \dots, \delta_N] \in \mathbb{R}^{C \times N} \quad (5)$$

其中 $f_{GAP}(\cdot)$ 表示全局平均池化 (Global Average Pooling, GAP) 操作, 该操作能够提取关键信息, 避免过拟合

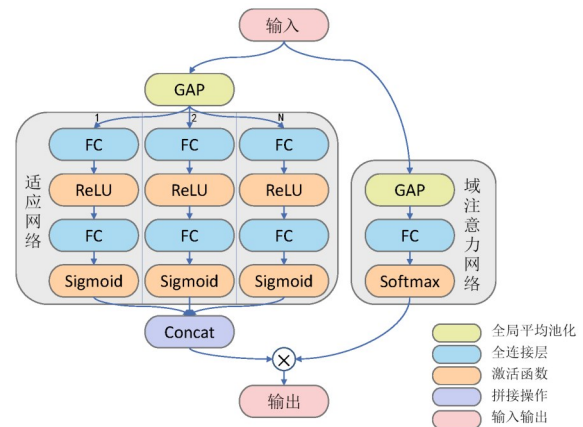


图 5 域自适应网络结构

并且不会引入待学习参数, 另外, 全局平均池化通过对特征求平均, 可以整合输入的空间信息, 对舰船目标的空间平移有更好的鲁棒性; $f_{SE}(\cdot)$ 表示全连接层、ReLU 函数、全连接层和 Sigmoid 函数的组合; δ_i 表示第 i 个适应网络输出的适应特征, $i = 1, 2, \dots, N$, 由实验结果可得此处 $N = 3$; Γ 表示 N 个适应特征经过拼接操作得到的 $N \times C$ 大小的拼接适应特征. 为了减少适应网络参数量, 全连接层的维度下采样率为 16. 相比使用 2 个检测器分别检测光学遥感图像和 SAR 图像, 域自适应网络整体的参数量大约减少了一半, 这对于特定需求的场景具有重要意义.

4.1.2 域注意力网络

基于特征的注意力机制借鉴了人类视觉机理, 使感知行为适应不同的任务和环境. 因此, 使用基于特征的注意力机制来辅助多源数据舰船目标检测任务是合理的. 域注意力网络结构简单, 包含 1 个 GAP 层、1 个 FC 层和 1 个激活层 (Softmax 函数). 此处的 GAP 层可以

视为一种结构正则化器,它可以将特征强制映射为概率置信度的权重.其计算过程可以表示为

$$\omega = \sigma \left(f_{\text{FC}} \left(f_{\text{GAP}}(x) \right) \right) \quad (6)$$

其中, ω 表示权重向量; σ 表示激活函数 Softmax.

拼接适应性特征 Γ 本质上是一个尺寸为 $N \times C$ 大小的特征向量,不同的适应网络可以看作将输入特征转换为不同域上的一个空间投影.注意力特征实际上是一个尺寸为 $C \times 1$ 大小的域敏感权重向量,以数据驱动的方式组合这些投影.如图6所示,对域注意力网络输出的权重向量进行统计分析.其中横坐标为权重,纵坐标为响应的样本数量,3种不同颜色的统计图分别对应着适应网络的3个分支.可以发现,各个分支中不同权重下所响应的样本数目接近正态分布,因此本文对其进行了高斯拟合.对于NWPU VHR-10数据,适应网络第一个分支的权重均值为0.16,第二个分支的权重均值为0.65,第三个分支的权重均值为0.18;而对于SSDD数据,网络3个分支的权重均值分别为0.16,0.66,0.18.2种数据集的权重分布十分相近,这表明该网络在应对多源数据时并非是只响应其中一个分支屏蔽另一个分支,而是针对不同的数据同时响应所有分支,同时侧面说明网络各个分支能够提取表征多源数据中舰船目标的一致性特征.整体上看,域自适应网络就是借鉴了类似正交坐标系的概念,对解耦后的多源特征进行线性组合.

4.2 目标函数

该多源遥感图像检测网络将进行分类预测(分类

置信度 conf)和回归预测(中心点偏移 point、预测框偏移 wh).为了避免简单样本主导模型的训练过程,引入焦点损失函数来挖掘难样本.分类损失函数 L_{cls} 由焦点损失函数实现,可以表示为

$$L_{\text{cls}} = (1 - \text{conf})^\alpha \log(\text{conf}) \quad (7)$$

其中, α 表示焦点损失函数的超参数,用于平衡难易样本,经验上取 $\alpha = 2.0$.另外,回归损失函数由 L_{wh} 和 L_{point} 两部分组成,使用 $L1$ 损失函数来微调预测框使其逼近真实标注框.真实标注框由左上角 (x_1, y_1) 和右下角 (x_2, y_2) 表示,中心点 p 对应的计算公式可以表示为

$$p = \left(\frac{x_1 + x_2}{2}, \frac{y_1 + y_2}{2} \right) \quad (8)$$

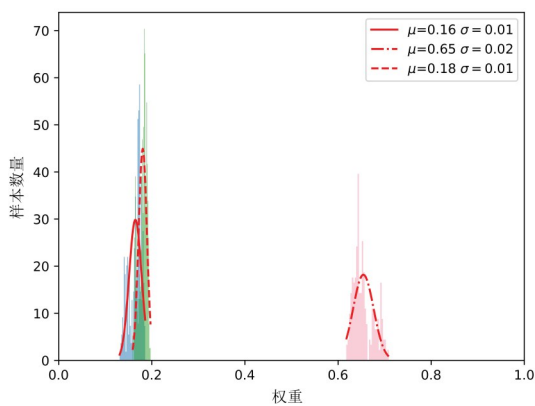
损失函数 L_{wh} 的表达式为

$$L_{\text{wh}} = \frac{1}{n} \sum_{i=1}^N \sqrt{(\hat{w}_i - w_i)^2 + (\hat{h}_i - h_i)^2} \quad (9)$$

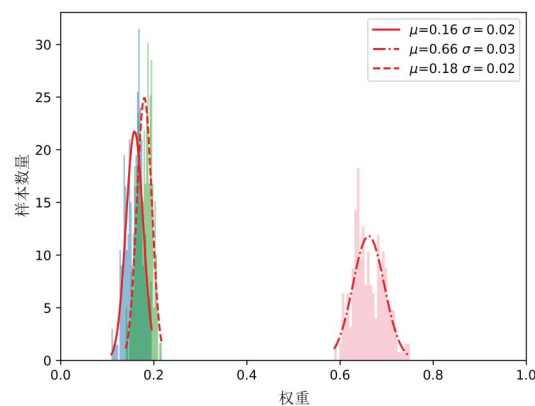
其中, n 表示输入特征中心点的数目; $w_i(h_i)$ 和 $\hat{w}_i(\hat{h}_i)$ 分别表示真实标注框宽(高)和预测框宽(高)偏移量.此外,由于下采样操作的影响,网络在前向传播过程中会出现特征像素未对齐的问题.为了解决该问题,采用中心点偏移 point 的预测值 Δp 实现特征像素对齐.中心点偏移对应的损失函数 L_{point} 可以表示为

$$L_{\text{point}} = \frac{1}{n} \sum_{i=1}^N \left| \Delta p - \left(\frac{p}{R} - \left\lfloor \frac{p}{R} \right\rfloor \right) \right| \quad (10)$$

其中, R 表示下采样因子,经验上设置 $R = 4$; $\lfloor \cdot \rfloor$ 表示下取整操作; n 表示分类置信度特征中的像素点数目.



(a) NWPU VHR-10数据权重向量以及高斯拟合结果图



(b) SSDD数据权重向量以及高斯拟合结果图

图6 域注意力网络学习的权重向量统计图

5 实验结果及分析

5.1 数据集和评估指标

本节主要介绍多源遥感图像舰船目标检测数据

集,其中包括光学遥感数据NWPU VHR-10,HRSC,SSDD,SAR-Ship-Dataset.本文在NWPU VHR-10和SSDD这2种不同数据源上进行消融实验,验证所提方法的有

效性. 接着在这 2 个数据集上进行对比实验, 将所提算法与现有前沿算法进行比较. 为了进一步验证本文方法对任意多源数据具有普适性, 本节还在上述 4 种不同源数据上进行了对比实验.

NWPU VHR-10^[32]是由西北工业大学自动化学院发布的光学高分遥感数据集. NWPU VHR-10 遥感数据集一共有 800 幅图像, 总计 10 个类别, 图像尺寸在 958×556 到 1003×808 不等. 其中含有舰船目标的图像 57 张, 这些图像包含民用港口以及远海等丰富的场景. 由于原始光学图像较少, 因此对原始数据进行旋转、颜色抖动和随机裁剪等操作进行数据扩增, 最终得到图像共 798 张. 训练集、验证集和测试集以 6:2:2 的比例进行随机划分.

SSDD^[33]是由中国人民解放军海军航空大学发布的 SAR 数据集, 用于舰船目标检测任务. 图像主要采集自 RadarSat-2, TerraSAR-X, Sentinel-1, 采用 VH, VV, HV, HH 四种极化方式, 分辨率大致在 1 m 到 15 m 不等. 图像采样场景丰富, 在远海和近海区域均具有舰船目标. 该数据集一共有 1160 幅图像, 总计 2456 艘舰船. 为了训练多源遥感图像舰船目标检测网络, 将 SSDD 数据随机划分成训练集、验证集和测试集 3 部分, 比例为 6:2:2.

HRSC^[34]是由中国科学院大学刘子坤等人发布的包含海面场景图像和近岸舰船图像的数据集, 图像的尺寸范围在 300×300 到 1500×900 之间. 本文将所有类型的船只都视为舰船这一大类, 按照原论文的数据划分, 即测试集为 444 张图片, 另外本文将训练集按照 8:2 的比例随机划分出一部分数据作为验证集.

SAR-Ship-Dataset^[35]是由中国科学院空天信息研究院王原原等人发布的数据集, 该数据集来源于 Gaofen-3 和 Sentinel-1 成像, 共有 43819 张图片, 大小为 256×256. 本文以 7:2:1 的比例将数据集随机划分为训练集、验证集和测试集.

为了评估不同舰船目标检测算法的性能, 评估指标采用平均精度均值 (Mean Average Precision, mAP). mAP 能够反映舰船目标检测算法的全局性能, 对应的计算公式为

$$mAP = \int_0^1 P(R)dR \quad (11)$$

其中, P 表示准确率; R 表示召回率.

5.2 实验设置

为了保证算法的正常运行, 本文在 Ubuntu v16.04 系统上搭建 PyTorch 深度学习环境, 即 CUDA v8.0 + cuDNN v6.0 + PyTorch. 所有试验均在有 24 GB 显存的 TITAN RTX 上运行. 所提算法将采用 CenterNet 作为基准网络, DLA-34 作为判别特征提取网络. 为了公平比

较, 对比算法所使用的参数都是其论文中公布的参数, 所有数据集的图像尺寸被缩放至 512×512 大小. 此外, 模型在训练阶段采用 Adam 算法进行优化, 其中动量参数 momentum 设置为 0.9, 权重参数设置为 0.000 1. 整个训练过程共 140 个 epochs, batch size 设置为 32. 模型的初始学习率为 1.25e-4, 并在第 90 个 epoch 衰减为 1.25e-5, 在第 120 个 epoch 衰减为 1.25e-6. 本文的实验首先在多源数据集上训练一个模型, 然后分别在各个数据集对应的测试集上计算 mAP.

5.3 消融实验

为了验证所提方法中图像级风格转换网络和特征级域自适应网络的有效性, 在多源数据集上进行了消融实验, 如表 1 所示. 为了公平比较, 在消融实验过程中依次添加图像级风格转换网络和特征级域自适应网络进行舰船目标检测. 相比基准模型 CenterNet, 加入图像级风格转换网络后, 模型在 2 种数据集上 mAP 均有提升: 在 NWPU VHR-10 上提升了 0.5%, 在 SSDD 上提升了 0.7%, 平均检测精度提升了约 0.5%. 这说明图像级风格转换网络在 SAR 数据和光学遥感数据的互相生成过程中起到重要的作用, 能够在图像级别实现域迁移并拉近 2 种图像之间的联系, 保证生成图像的质量. 由于 SAR 数据和光学遥感数据存在域偏移, 因此仅依赖图像级别的风格转换网络难以进一步提取多源图像的判别性特征. 引入特征级域自适应网络后, 模型的检测精度进一步获得提升: 在 NWPU VHR-10 上提升了 0.9%, 在 SSDD 上提升了 1.3%, 平均检测精度提升了约 1.2%. 这说明在判别特征的提取阶段, 特征级域自适应网络能够在多维空间上表征多源信息并有效检测多源数据中的舰船目标.

表 1 多源数据集上消融实验 mAP

图像级	特征级	NWPU VHR-10/%	SSDD/%	Avg/%
No	No	72.7	93.6	83.2
Yes	No	73.2	94.3	83.7
Yes	Yes	74.1	95.6	84.9

5.4 与现有方法对比实验

为了验证本文方法的有效性和优越性, 本节将其与现有前沿算法进行了比较, 包括双阶段算法 (Faster R-CNN^[8], Cascade R-CNN^[36], Libra R-CNN^[9])、单阶段算法 (RetinaNet^[14], EFGRNet^[6], CenterNet^[7], CenterNet++^[37]) 以及域自适应算法 (Universal DA^[25], DA Faster R-CNN^[23]).

各种算法在数据集 NWPU VRH-10 和 SSDD 上的结果如表 2 所示. 整体上看, 本文方法性能稳定且能够实时检测舰船目标. Cascade R-CNN 由于采用级联策略优化预测模型, 因此平均检测精度达到 82.8%, 且优于

表2 各种算法在SSDD和NWPU VRH-10数据集上的mAP和Runtime

Methods	NWPU VHR-10/%	SSDD/%	Avg/%	Runtime/s
Faster R-CNN	71.8	91.3	81.6	0.068
Cascade R-CNN	71.9	93.4	82.8	0.087
Libra R-CNN	72.3	93.9	83.1	0.071
RetinaNet	73.1	90.5	81.8	0.059
EFGRNet	72.2	92.4	82.3	0.033
CenterNet	72.7	93.6	83.2	0.031
CenterNet++	73.4	94.3	83.9	0.033
Universal DA	73.1	88.8	81.0	0.122
DA Faster R-CNN	74.1	88.1	81.1	0.069
Ours ($N=1$)	73.7	94.8	84.2	0.037
Ours ($N=2$)	73.9	95.0	84.5	0.037
Ours ($N=3$)	74.1	95.6	84.9	0.037
Ours ($N=4$)	73.8	94.6	84.2	0.038

Faster R-CNN的性能。RetinaNet只解决了网络在正负样本采样阶段的不均衡问题,而Libra R-CNN从特征、采样以及损失3个层面解决了舰船目标检测存在不均衡的问题。因此,在2种数据集上的平均mAP为83.1%,优于RetinaNet(81.8%)。相比于RetinaNet,EFGRNet平均检测精度提升了0.5%,这是因为EFGRNet引入上下文信息增强网络,缓解了多尺度舰船目标检测难题。CenterNet直接将每个舰船目标建模为目标边界框的中心点,避免对复杂背景产生冗余计算并提升检测准确率。所以,CenterNet相比于EFGRNet方法平均检测精度提升了0.9%。相比于CenterNet,CenterNet++的平均检测精度提升了0.7%,这是因为CenterNet++通过特征精修和再分类操作提取到了更细粒度的特征。相比于Universal DA,DA Faster R-CNN的平均

检测精度提升了0.1%,主要原因是DA Faster R-CNN通过对抗训练方式学习了多源数据的判别特征。

此外,表2给出了不同 N 设置下本文方法的检测准确率。总体上看,特征级域自适应网络对适应网络的个数 N 比较稳健,所测mAP均高于其他方法。但当 N 过小或过大时,准确率会受到一定影响。当 N 特别小时,域自适应网络结构简单,会影响多源图像信息的表征。当 N 特别大时,域自适应网络结构冗余,会影响域自适应特征的判别性。图7展示了不同 N 下,本方法在NWPU VRH-10和SSDD数据集上的检测曲线(PR曲线),可直观看出算法的稳定性。最后,当 N 取3时,本文方法取得了最高的平均检测准确率84.9%,并且在2种数据集上均取得最高检测准确率。这是因为所提算法通过引入风格转换网络和域自适应网络,在图像级和特征级2个层面实现多源数据的域自适应,从而提取到多源数据的一致性特征进行舰船目标检测。

为了进一步验证本文方法对任意多源数据集的普适性,本节在上文提到的4种数据源上进行实验,结果如表3所示。和理论分析一致,所提方法取得了较好性能。相比于2种数据源的平均mAP,4种数据源的检测性能可以提高5.9%。这是因为数据源越丰富,所提算法可以更好地提取不同域的一致性表征,从而更有利于对多源特征进行信息解耦与耦合。图8给出了不同 N 设置下本文方法对多源数据的检测曲线。同样地,本算法对任意多源数据也足够稳定,当适应网络个数为3时取得最高检测准确率。综上所述,本文方法在一定程度上缓解了多源数据存在的域偏移问题,适用于多源遥感图像舰船目标检测。

5.5 可视化实验与分析

图9展示了本方法在4种数据集上的部分检测结果,检测框全部正确。所提方法不仅能检测多尺度舰船

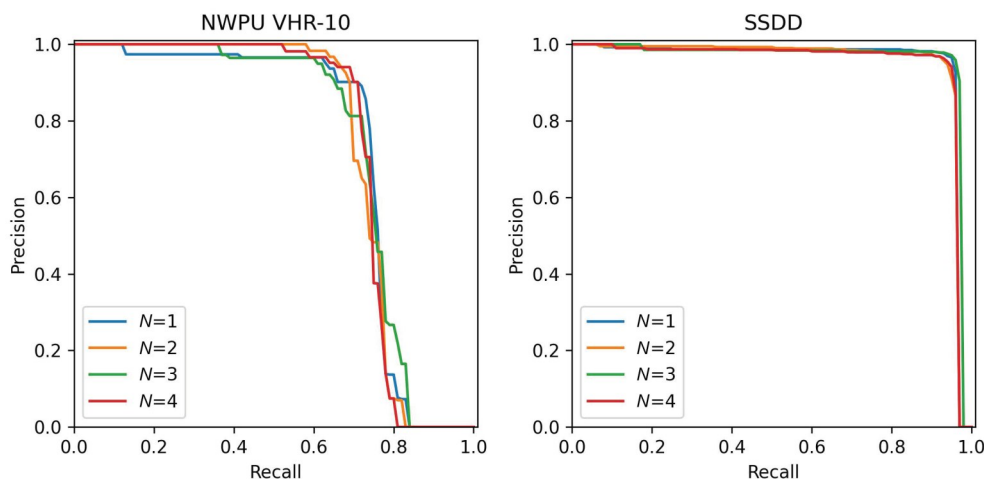


图7 不同 N 设置下NWPU VRH-10和SSDD数据集的检测曲线

表 3 多源数据集上各种算法的 mAP

Methods	NWPU VHR-10	SSDD	HRSC	SAR-Ship-Dataset	Avg
Faster R-CNN	70.2	84.1	85.2	94.9	83.6
Cascade R-CNN	70.0	86.8	85.5	94.8	84.3
Libra R-CNN	75.2	85.6	84.8	95.7	85.3
RetinaNet	69.8	91.3	84.9	96.4	85.6
EFGRNet	79.7	91.6	81.2	94.9	86.9
CenterNet	80.6	91.2	91.8	94.1	89.4
CenterNet++	76.6	92.3	92.7	94.2	89.0
Universal DA	72.8	85.5	74.8	87.7	80.2
DA Faster R-CNN	78.9	88.1	76.3	89.6	83.2
Ours ($N=1$)	78.2	93.0	93.7	94.5	89.8
Ours ($N=2$)	81.1	92.9	93	94.8	90.4
Ours ($N=3$)	82.1	92.7	93.8	94.6	90.8
Ours ($N=4$)	77.3	92.8	92.7	94.6	89.4

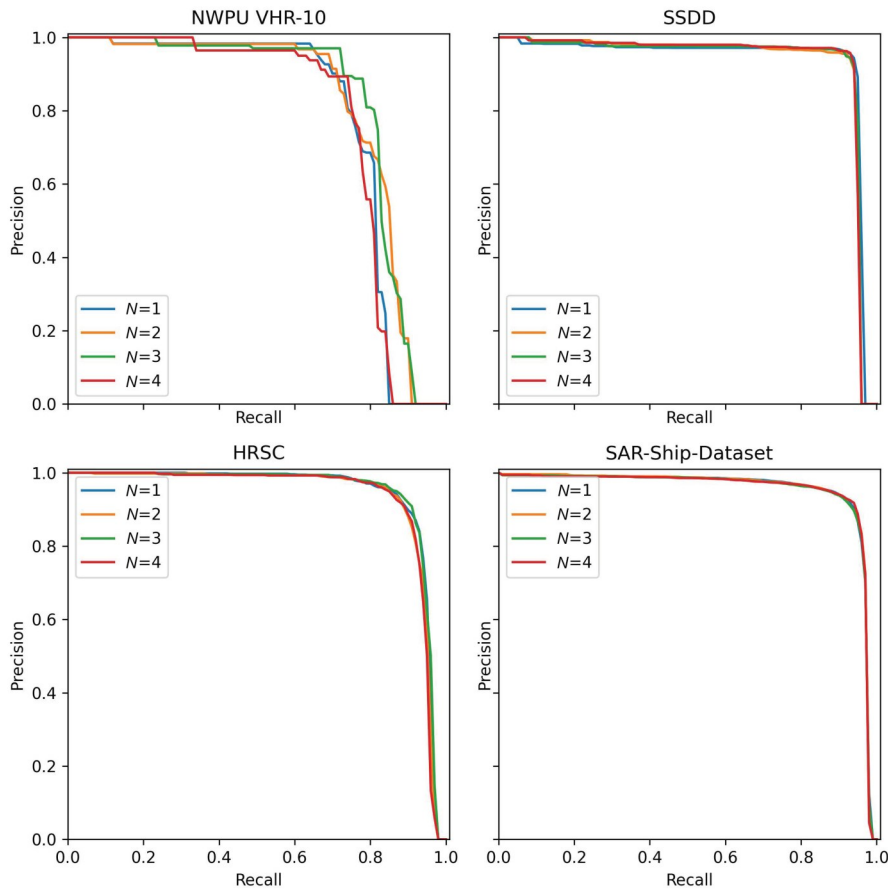


图 8 不同 N 设置下多源数据的检测曲线

目标,尤其是小目标,在复杂背景(SAR 图像中岛屿和陆地,光学图像中云雾)的干扰下也有不错的检测效果. 这表明域自适应网络通过解耦操作避免了复杂背景的干扰同时提取了关键特征,提升了舰船目标检测准确率. 本文方法能够从图像级和特征级拉近不同域,是更为通用的舰船目标检测器. 此外,图 10 对比了不

同算法在多源数据集上的部分检测结果,包括双阶段算法 Libra R-CNN、单阶段算法 RetinaNet、无锚框舰船检测算法 CenterNet++、多源检测算法 Universal DA 以及本文算法. 其中深蓝色框为 ground truths,紫罗兰色框为 true positives,红色框为 false positives,绿色框为 false negatives. 从上到下依次为 NWPU VHR-10, SSDD,

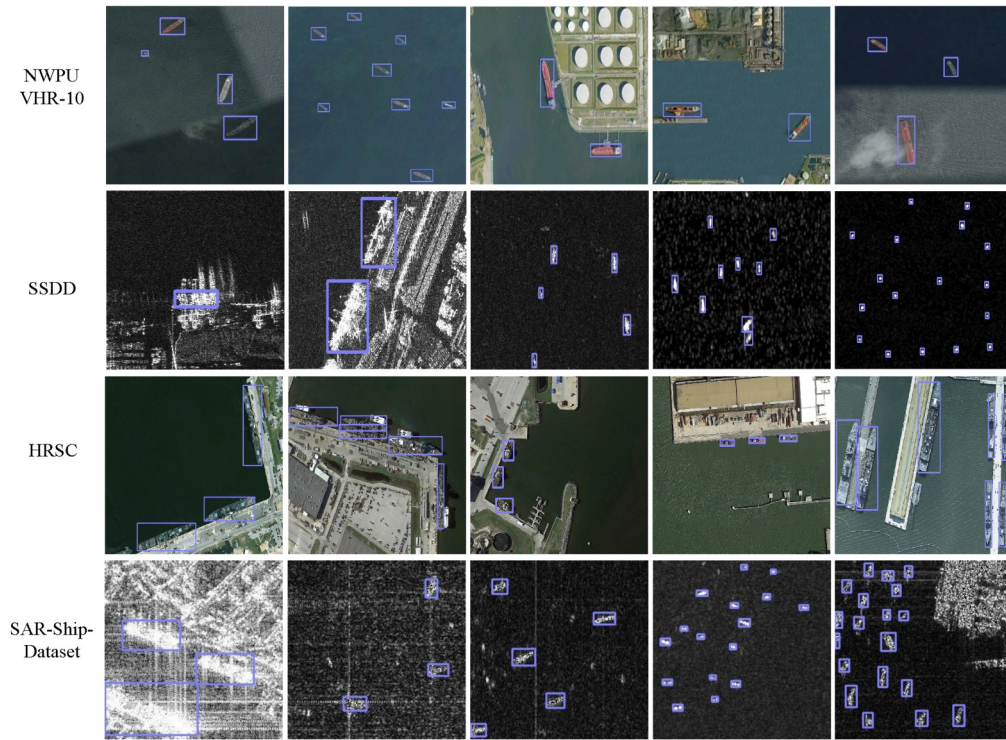


图9 本文方法在多源数据集上的部分检测结果

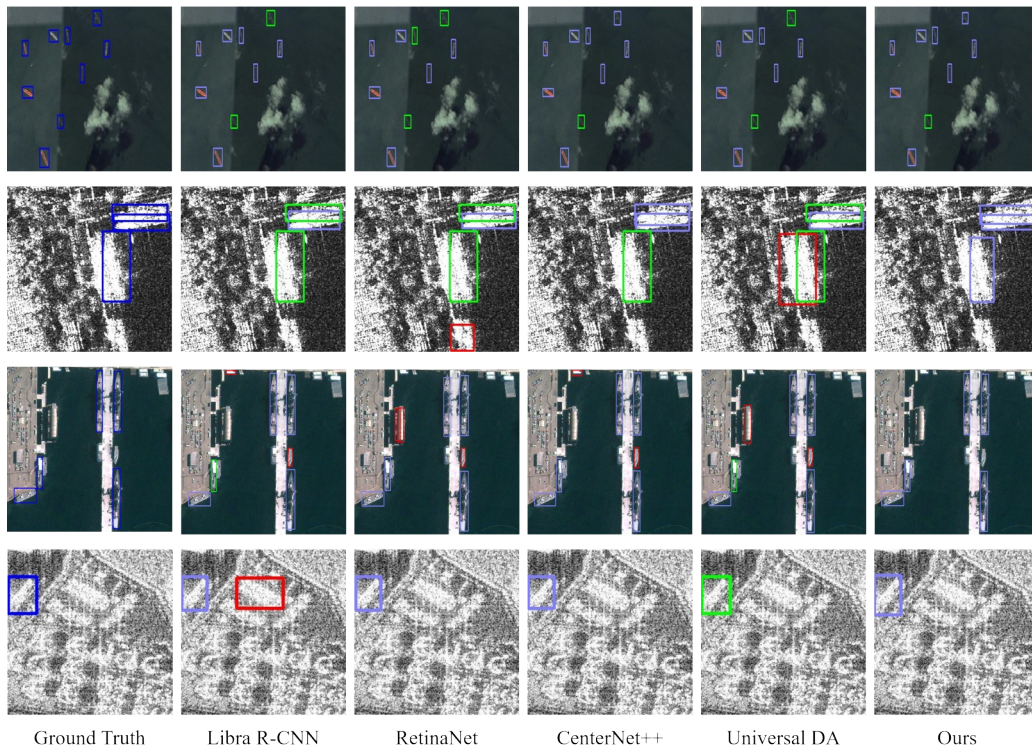


图10 不同方法在多源数据集上的部分检测结果

HRSC, SAR-Ship-Dataset 数据集中的图像. 可以看出, 本算法明显优于其他算法, 减少了漏检和误检, 检测准确率更高.

图11是本文方法在4种数据集上不理想的检测结果. 对于处于更加复杂背景的舰船目标, 本文方法会漏检和误检其中的个别舰船. 这是因为SAR图像自身缺乏细

节纹理信息,背景噪声很容易对其检测造成干扰,而光学遥感图像中存在和舰船外表相似的目标.因此,为了更进一步提升算法的检测能力,可以使用图像超分辨率重建

网络提升舰船目标的分辨率以获得更优秀的判别性特征.具体地,可以采用基于无监督PULSE算法^[38],通过探索生成模型的潜在空间实现遥感图像的超分辨率重建.

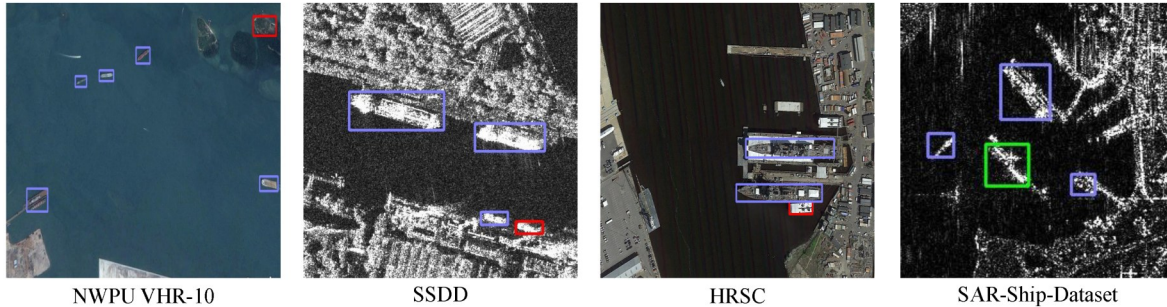


图 11 本文方法不理想的检测结果

6 结束语

为了实现一个通用方法检测多源数据中的舰船,本文提出了一种基于不变特征的多源遥感图像舰船目标检测算法,实现图像级和特征级 2 个层面域自适应.通过引入风格转换网络在图像级拉近光学遥感图像域和 SAR 图像域,这有利于后续网络在图像层面上学习多源数据的不变特征,同时实现数据扩增.另外,通过采用域自适应网络实现对多源特征的解耦以提取到多域的一致性特征,经过注意力机制的自适应权重分配实现特征重组,在特征层面上学习多源数据的不变特征.本文在 NWPU VHR-10, SSDD, HRSC, SAR-Ship-Dataset 数据集上进行了大量实验并分析了其算法性能.实验表明,所提方法实现了最佳检测准确率并且能够满足实时性的需求.

本文方法虽然相比现有方法取得了最佳检测精度,但是在面临更复杂背景时,由于背景噪声对目标特征的干扰,会出现漏检和误检情况.在未来,将使用图像超分辨率重建网络来缓解此问题,以进一步提升检测准确率.

参考文献

- [1] 种劲松, 朱敏慧. SAR 图像舰船及其尾迹检测研究综述[J]. 电子学报, 2003, 31(9): 1356-1360.
CHONG J S, ZHU M H. Survey of the study on ship and wake detection in SAR imagery[J]. Acta Electronica Sinica, 2003, 31(9): 1356-1360. (in Chinese)
- [2] WANG C, BI F, ZHANG W, et al. An intensity-space domain CFAR method for ship detection in HR SAR images[J]. IEEE Geoscience and Remote Sensing Letters, 2017, 14(4): 529-533.
- [3] PAPPAS O, ACHIM A, BULL D. Superpixel-level CFAR

detectors for ship detection in SAR images[J]. IEEE Geoscience and Remote Sensing Letters, 2018, 15(9): 1397-1401.

- [4] HUO W, HUANG Y, PEI J, et al. Ship detection from ocean SAR image based on local contrast variance weighted information entropy[J]. Sensors, 2018, 18(4): 1196.
- [5] 王明春, 张嘉峰, 杨子渊, 等. Beta 分布下基于白化滤波的极化 SAR 图像海面舰船目标 CFAR 检测方法[J]. 电子学报, 2019, 47(9): 77-84.
WANG M C, ZHANG J F, YANG Z Y, et al. A CFAR detection method of ship targets in polarimetric SAR imagery based on whitening filter under Beta distributed texture[J]. Acta Electronica Sinica, 2019, 47(9): 77-84. (in Chinese)
- [6] NIE J, ANWER R, CHOLAKKAL H, et al. Enriched feature guided refinement network for object detection[C]// Proceedings of the IEEE International Conference on Computer Vision. Kerkyra: IEEE, 2019: 9537-9546.
- [7] ZHOU X, WANG D, KRAHENBUHL P. Objects as points [EB/OL]. (2019-04-16)[2021-07-05]. <https://arxiv.org/abs/1904.07850v1>.
- [8] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6): 1137-1149.
- [9] PANG J, CHEN K, SHI J, et al. Libra R-CNN: Towards balanced learning for object detection[C]// Proceedings of International Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019: 821-830.
- [10] ZHU J Y, PARK T, ISOLA P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]// Proceedings of the IEEE International Confer-

- ence on Computer Vision. Venice: IEEE, 2017: 2223-2232.
- [11] GIRSHICK R. Fast R-CNN[C]//Proceedings of the IEEE International Conference on Computer Vision. Santiago: IEEE, 2015: 1440-1448.
- [12] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Proceedings of International Conference on Computer Vision and Pattern Recognition. Venice: IEEE, 2017: 2117-2125.
- [13] JIANG B, LUO R, MAO J, et al. Acquisition of localization confidence for accurate object detection[C]//Proceedings of the European Conference on Computer Vision. Munich: Springer, 2018: 784-799.
- [14] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]//Proceedings of the IEEE International Conference on Computer Vision. Venice: IEEE, 2017: 2980-2988.
- [15] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection [C]//Proceedings of International Conference on Computer Vision and Pattern Recognition. Los Vegas: IEEE, 2016: 779-788.
- [16] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger[C]//Proceedings of International Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: IEEE, 2017: 7263-7271.
- [17] REDMON J, FARHADI A. YOLOv3: An incremental improvement[EB/OL]. (2018-04-08)[2021-07-05]. <https://arxiv.org/abs/1804.02767>.
- [18] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal speed and accuracy of object detection [EB/OL]. (2020-04-23)[2021-07-05]. <https://arxiv.org/abs/2004.10934>.
- [19] LAW H, DENG J. Cornernet: Detecting objects as paired keypoints[C]//Proceedings of the European Conference on Computer Vision. Munich: Springer, 2018: 734-750.
- [20] CUI Z, LI Q, CAO Z, et al. Dense attention pyramid networks for multi-scale ship detection in SAR images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(11): 8983-8997.
- [21] YANG X, LIU Q, YAN J, et al. R3Det: Refined single-stage detector with feature refinement for rotating object [EB/OL]. (2019-08-15)[2021-07-05]. <https://arxiv.org/abs/1908.05612>.
- [22] FU J, SUN X, WANG Z, et al. An anchor-free method based on feature balancing and refinement network for multiscale ship detection in SAR images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 59(2): 1331-1344.
- [23] CHEN Y, LI W, SAKARIDIS C, et al. Domain adaptive Faster R-CNN for object detection in the wild[C]//Proceedings of International Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 3339-3348.
- [24] SAITO K, USHIKU Y, HARADA T, et al. Strong-weak distribution alignment for adaptive object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019: 6956-6965.
- [25] WANG X, CAI Z, GAO D, et al. Towards universal object detection by domain attention[C]//Proceedings of International Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019: 7289-7298.
- [26] GATYS L A, ECKER A S, Bethge M. Image style transfer using convolutional neural networks[C]//Proceedings of International Conference on Computer Vision and Pattern Recognition. Bhubaneswar: IEEE, 2016: 2414-2423.
- [27] BERGER G, MEMISEVIC R. Incorporating long-range consistency in CNN-based texture generation[EB/OL]. (2016-01-03)[2021-07-05]. <https://arxiv.org/abs/1606.01286v1>.
- [28] GU S, CHEN C, LIAO J, et al. Arbitrary style transfer with deep feature reshuffle[C]//Proceedings of International Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 8222-8231.
- [29] WANG X, OXHOLM G, ZHANG D, et al. Multimodal transfer: A hierarchical deep convolutional neural network for fast artistic style transfer[C]//Proceedings of International Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 5239-5247.
- [30] YU F, WANG D, SHELHAMER E, et al. Deep layer aggregation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 2403-2412.
- [31] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 7132-7141.
- [32] CHENG G, ZHOU P, HAN J. Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2016, 54(12): 7405-7415.
- [33] LI J, QU C, SHAO J. Ship detection in SAR images

based on an improved faster R-CNN[C]//SAR in Big Data Era: Models, Methods and Applications(BIGSARDA-TA). Beijing: IEEE, 2017: 1-6.

- [34] LIU Z, YUAN L, WENG L, et al. A high resolution optical satellite image dataset for ship recognition and some new baselines[C]//International Conference on Pattern Recognition Applications and Methods. Porto: ICPRAM, 2017: 324-331.
- [35] WANG Y, WANG C, ZHANG H, et al. A SAR dataset of ship detection for deep learning under complex backgrounds[J]. Remote Sensing, 2019, 11(7): 765.
- [36] CAI Z, VASCONCELOS N. Cascade R-CNN: Delving into high quality object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 6154-6162.
- [37] GUO H, YANG X, WANG N, et al. A CenterNet++ model for ship detection in SAR images[J]. Pattern Recognition, 2021, 112: 107787.
- [38] MENON S, DAMIAN A, HU S, et al. PULSE: Self-supervised photo upsampling via latent space exploration of generative models[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 2437-2445.



王楠楠 男, 1986年11月出生, 山东安丘人. 现为西安电子科技大学通信工程学院教授、博士生导师. 研究方向为计算机视觉和机器学习.

E-mail: nnnwang@xidian.edu.cn



高新波(通讯作者) 男, 1972年8月出生, 山东莱芜人. 现为西安电子科技大学电子工程学院教授、博士生导师, 重庆邮电大学计算机学院教授、博士生导师. 研究方向为计算机视觉和机器学习.

E-mail: xbgao@mail.xidian.edu.cn, gaoxb@cqupt.edu.cn

作者简介



杨 曦 女, 1988年10月出生, 陕西咸阳人. 现为西安电子科技大学通信工程学院教授、博士生导师. 研究方向为计算机视觉和机器学习.

E-mail: yangx@xidian.edu.cn



张 鑫 男, 1998年11月出生, 陕西渭南人. 现为西安电子科技大学通信工程学院硕士研究生. 研究方向为计算机视觉和目标检测.

E-mail: xzhang_98@stu.xidian.edu.cn



郭浩远 男, 1996年2月出生, 河南郑州人. 2021年获得西安电子科技大学通信工程学院硕士学位. 研究方向计算机视觉和机器学习.

E-mail: guohy@stu.xidian.edu.cn