

基于跳跃特征金字塔的域适应目标检测模型

张永福, 宋海林

(陕西师范大学 计算机科学学院, 陕西 西安 710119)

摘要:针对训练数据和测试数据来源不同,特征分布差异较大,以及目标大小不一时,深度目标检测模型性能下降明显的问题,提出基于跳跃特征金字塔的域适应遥感图像目标检测模型。首先,在目标检测网络中增加域适应部分,将训练数据作为源域,测试数据作为目标域,通过对抗训练的方式,对具有不同特征分布的两个域,关注其差异,同时提取源域和目标域数据的特征,减小遥感图像由光照、角度等不同造成的源域和目标域图像在图像级的域偏移对目标检测的影响,提升检测模型的推广性能。其次,提出跳跃特征金字塔结构,通过特征上采样,以及同层连接、相隔层间的跳跃连接与特征融合,来增强特征图的细节信息和语义信息,以提高模型对不同尺度目标的检测精度。最后,使用区域推荐网络在多个不同分辨率的特征图上提取候选区域,检测不同尺度的目标。所提模型在NWPUVHR-10数据集上检测精度达到了98.2%、误检率为5.4%、漏检率为8.3%;在RSOD-DATA的低亮度数据集上检测精度达到了62%,误检率、漏检率为18.2%和18.5%。与其他模型相比,所提模型的性能有明显提升,具有更好的推广性。

关键词:遥感图像;目标检测;特征金字塔;跳跃连接;域适应

中图分类号:TP391

文献标识码:A

文章编号:1673-629X(2022)09-0028-08

doi:10.3969/j.issn.1673-629X.2022.09.005

Skip Feature Pyramid Based Domain Adapted Model for Object Detection

ZHANG Yong-fu, SONG Hai-lin

(School of Computer Science, Shaanxi Normal University, Xi'an 710119, China)

Abstract: In order to solve the problem that the detection performance of the deep object detection model drops significantly when the feature distribution differ greatly between training data and testing data, and when objects vary in size, a skip feature pyramid based domain adapted model for remote sensing image object detection is proposed. Firstly, the domain adaptation part is added into the object detection network. Regarding training data as the source domain and testing data as the target domain and focusing on their differences, the network can extract robust deep convolution features for both source and target domains with different feature distributions, through the adversarial training. This reduces the influence of the deviation between source and target domain image due to different illumination and angle in image acquisition, and improves the generalization of the model. Secondly, a skip feature pyramid structure is proposed to enhance the detail information and semantic information of the feature map by feature upsample, connection and jump connection between the same layer and the nonadjacent layer, and feature fusion, so as to improve the detection accuracy for different scale objects. Finally, the region proposal network is used to extract candidate regions on multiple feature maps with different resolutions to detect targets of different scales. The experimental results show that the detection accuracy on the NWPUVHR-10 dataset is 98.2%, the false detection rate is 5.4%, and the missed detection rate is 8.3%. The accuracy on the RSOD-DATA low-brightness dataset reaches 62%, and the false detection rate and missed detection rate are 18.2% and 18.5% respectively. Compared with the other models, the performance of the proposed model has improved significantly with better generalization.

Key words: remote sensing image; object detection; feature pyramid; skip connection; domain adaptation

0 引言

遥感图像目标检测技术是一种实现遥感目标自动分类和定位的智能化数据分析技术,也是遥感图像解译的重要研究方向之一^[1-2]。在该任务中,既要考虑

诸如光照、遮挡和几何形变等客观因素,又要考虑目标实例的多尺度性、密集分布、形状差异等问题^[3],因此遥感图像目标检测具有很大的挑战性。

早期方法一般将目标检测转化成滑动窗口分类问

收稿日期:2021-09-17

修回日期:2022-01-18

基金项目:国家自然科学基金项目(61701290)

作者简介:张永福(1973-),男,实验师,研究方向为智能信息处理;宋海林(1994-),男,硕士,研究方向为机器学习与图像处理。

题^[4],卷积神经网络(Convolutional Neural Network, CNN)^[5]的迅速发展为目标检测研究提供了新思路。2014年,Girshick等^[6]提出了区域卷积神经网络R-CNN,之后提出了一种改进的Fast R-CNN^[7]算法,实现了在所有ROI(Region Of Interest)中共享卷积特征。Fast R-CNN在提取区域候选框时使用选择性搜索算法,耗时大,Ren等^[8]提出了Faster R-CNN算法,用区域建议网络来代替选择性搜索算法,减少了提取区域候选框的时间。针对目标检测中待检测目标尺度变化的问题,Hong等^[9]提出了深度轻量级神经网络,通过融合多层特征预测多尺度目标。Liu等^[10]提出的SSD模型在不同分辨率的特征图上检测多尺度目标,对小目标的检测性能较差。Lin等^[11]提出特征金字塔(Feature Pyramid Network, FPN),融合浅层特征图和深层特征图,在融合后的特征图上检测多尺度目标,但FPN的自顶向下路径使深层特征图的语义信息难以有效传播至浅层特征图。Li等^[12]提出了一种新的跨层特征金字塔网络(CFPN),其中允许直接的跨层通信来改进特征的渐进式融合。该模型主要包括两个模块,跨层特征聚合模块和跨层特征分布模块,聚合模块将不同层的多尺度特征进行融合增强,而分布模块则利用多尺度池将聚集的要素分配回其对应的层。特征融合时,由于每个特征图的区分程度不同,直接将它们结合起来可能会将一些有价值的信息转化为噪声,从而降低性能。Xu等提出了一种新的通道关系特征金字塔网络(CR-FPN),该网络通过相似性度量函数从不同层次的特征映射中捕捉通道的长期关系,进而放大最相关的通道,抑制不相关的通道。单义等^[13]通过跳跃连接的方式融合了不相邻特征图之间的信息,目标检测实验取得了较好的效果。上述代表性方法各有其优点,但就遥感图像的多尺度目标和小目标检测而言,仍需在特征提取时寻求细节和语义特征的有效获取和融合机制。

目标检测模型在训练集图像和测试集图像特征分布相似的数据集中的检测性能较高,但是当训练集图像和测试集图像差异较大时模型的检测性能较差。遥感图像进行目标检测时,不同的遥感数据集由于采集图像设备的不同造成图像遥感的尺寸、风格等不同;另外,同一数据集可能因为采集的场景、时间、角度不同,导致数据集中遥感图像的特征分布也不同。因此,因为训练集和测试集的差异使检测性能下降。为了解决此问题,目前主要有两种方式:(1)通过收集更多的训练数据减少域变化带来的影响,一种是标注数据来增加数据,但是标注数据费时费力,另一种是扩充现有数据,但是它们依然来源于同样的数据集,不是跨域的;(2)利用域适应提高目标检测模型的鲁棒性,提高对

不同场景中目标的检测性能。Hong等^[14]将GAN集成到目标分割模型,通过对抗训练实现域适应,实现了语义分割。Chen等^[15]将两个自适应组件集成到Faster R-CNN模型中,分别在图像和实例两个级别进行域适应,得到一种一致性正则化的域适应目标检测模型。为了解决跨域目标检测中的负迁移问题,Chen等^[16]提出了一种基于强化学习的SIR模型,该模型中包含的两个智能体S-Agent和T-Agent能够学习删除低相关度的源实例和离群目标实例。传感器不同的相同地面区域的遥感影像获取条件的差异性使其特征有差异,且往往场景复杂、目标可能大小不一且密集,这使得域适应思想有望应对这些问题,提升模型的推广性能,但是这方面的研究才开始,针对文中采用的遥感图像数据集的相关研究尚未检索到。

基于以上分析,提出了基于跳跃特征金字塔的域适应遥感图像目标检测模型(Skip Feature Pyramid Domain Adapted Model for Remote Sensing Image Object Detection, SFP-DAM)。SFP-DAM是端到端的两阶段目标检测模型,主要贡献有两点:首先引入跳跃特征金字塔结构,设计不同的横向、纵向层间连接方式和融合方式,促进了语义信息和细节信息的融合和语义信息回传,有益于获取不同尺度目标的有效特征。其次,通过增加域适应部分,使模型能适应不同来源的图像,同时获取训练和测试图像的特征,提高模型在不同遥感图像数据上的泛化能力,从而使检测模型获得更好的鲁棒性。

1 基于跳跃特征金字塔的域适应遥感图像目标检测模型

基于跳跃特征金字塔的域适应遥感图像目标检测模型以Faster R-CNN模型为基础,模型结构如图1所示。该模型由目标检测网络和域适应网络两个分支组成。其中目标检测网络分为特征提取网络、候选区域建议网络以及目标分类回归网络三部分。候选区域建议网络和域适应网络共享特征提取网络提取的特征。

1.1 目标检测网络

如图1所示,目标检测网络包括三个部分,其处理过程为:首先,将遥感图像输入特征提取网络中,生成不同层次的特征图。其次,将不同层次的特征图送入跳跃特征金字塔结构,通过跳跃连接和融合,更好地提取和保留深层和浅层特征图的信息。特征提取网络包括13个卷积层、5个池化层。卷积过程分成5个阶段conv1-conv5,各阶段生成的特征图记为 $\{C_1, C_2, C_3, C_4, C_5\}$ 。其后接跳跃特征金字塔结构融合不同层次的特征图,增强浅层特征图的语义信息,提高小目标的检测性能。然后,在不同分辨率的特征图上使用区域

建议网络生成候选区域,对候选框进行筛选。最后,将筛选后的多尺度候选区域对应的特征图送入 ROI

Pooling 层,进行尺度归一化,并进行第二次检测过程。

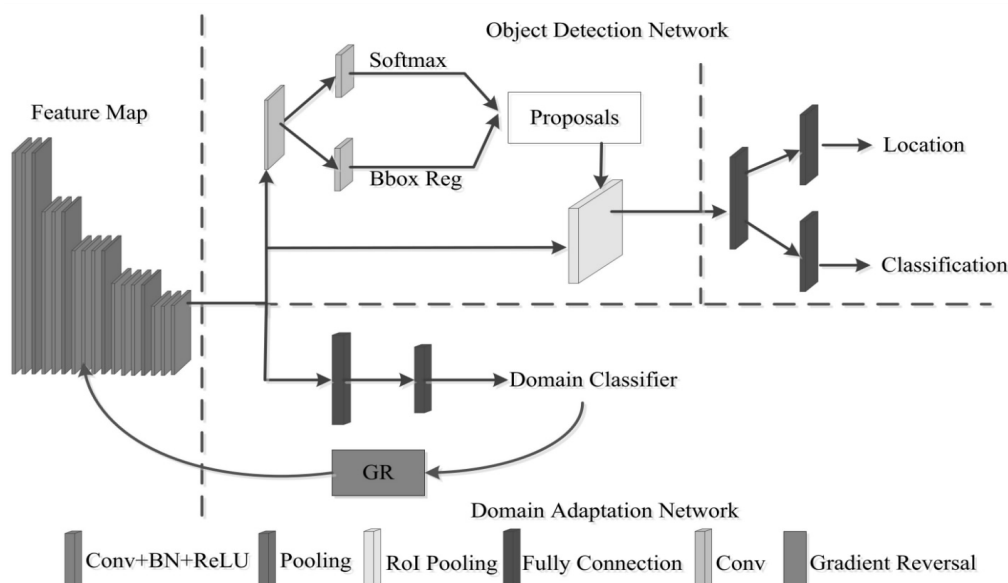


图 1 基于跳跃特征金字塔的域适应遥感图像目标检测模型

1.1.1 跳跃特征金字塔

在特征提取网络中,不同层次的特征图含有不同程度的语义信息和细节信息,浅层特征具有丰富的细节信息但缺乏语义信息,深层特征具有丰富的语义信息但缺乏细节信息^[17]。就遥感图像目标检测而言,小目标、密集目标以及多尺度目标在检测时,更需要不同层次的特征进行融合,以增强特征图中包含的语义和

细节特征信息。为了在不同层次上都更好地综合细节和语义信息,该文设计了跳跃特征金字塔结构(Skip Feature Pyramid Network, SFPN),一方面将深层特征图的语义信息传递至浅层特征图,另一方面使用跳跃连接减少传递语义信息过程中的信息丢失情况。

跳跃特征金字塔结构如图 2 所示。

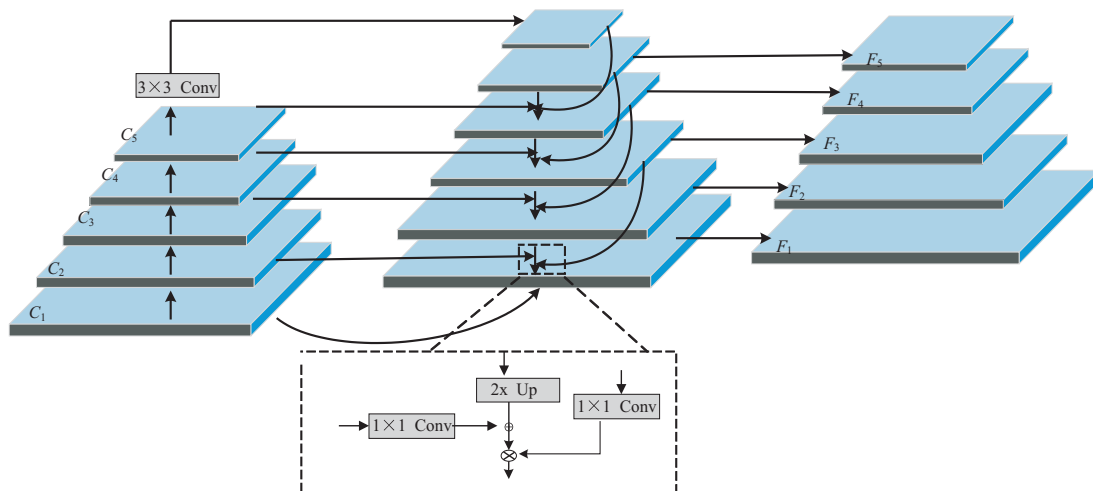


图 2 跳跃特征金字塔结构

输入图像经过特征提取网络形成一组不同层次的特征图 $\{C_1, C_2, C_3, C_4, C_5\}$,即为自底向上路径上的特征图。采用双线性插值上采样深层特征图构造自顶向下的路径,并与自底向上路径上对应的相同分辨率的特征图通过像素相加来融合,同时使用跳跃连接结构将相隔一层的深层特征图与融合后的特征图逐像素相乘,最终形成一组做检测的特征图 $\{F_1, F_2, F_3, F_4, F_5\}$ 。将更深层的特征图与浅层特征图通过跳跃连接

方式融合,有助于更好地传递深层特征图的语义信息,缓解语义信息传递过程中信息丢失的问题。SFPN 输出的特征图 F_i 见公式(1)。

$$F_i = \text{conv}(p(C_i) + \text{up}(C_{i+1}) \times p(C_{i+2}))$$

$$i \in \{1, 2, 3, 4\} \quad (1)$$

其中,conv()表示卷积运算,去除上采样过程中产生的特征冗余及混叠情况; $p()$ 表示使用 1×1 卷积降维映射;up()表示双线性插值上采样; $+$ 、 $\times$ 分别表示逐像

素相加和逐像素相乘。

1.1.2 区域建议网络

将跳跃特征金字塔结构生成的特征图送入区域建议网络以生成候选区域。卷积神经网络中深层特征图的分辨率较小、语义信息丰富、感受野较大,适合检测大尺度的目标;浅层特征图的分辨率较大、细节信息丰富、感受野较小,融合了深层特征图传递来的语义信息后,可以更好地检测小尺度的目标。因此,将在不同层次的特征图上分别使用区域建议网络来生成候选区域,以更好地检测多尺度目标。在区域建议网络中针对不同分辨率的特征图设置不同尺度的锚框,对于各层次特征图依次设定锚框大小为 $\{162, 202, 322, 642, 1282\}$,每一尺度锚框设定三个比例 $\{1:1, 1:2, 2:1\}$ 。区域建议网络生成候选区域后将其送入 ROI Pooling 层进行尺度归一化,并进行第二次分类回归的检测过程。

1.2 域适应

域适应是一种特殊的迁移学习方式,对于图像任

务,源域、目标域分别对应于来源不同的训练集和测试集。域适应通过学习一个复合映射,使源域和目标域数据映射到一个特征空间中,复合映射同时学习源域和目标域上的特征,源域和目标域数据分布不同时,复合映射可以缩小源域和目标域分布差异,达到同时提取源域和目标域特征的目的^[18-20]。数据扩充也可以在一定程度上弥补数据差异,但是扩充前后的数据来源相同,扩充前后数据的差异也取决于数据扩充的操作类型,而域适应可以应对各种目标域与源域数据的不同。

如图3所示,SFP-DAM模型中域适应部分由特征提取网络、标签分类器以及域分类器、梯度反传(Gradient Reversal, GR)等部分组成。特征提取网络提取的卷积特征将送入两个分支,一个分支是标签分类器,训练标签分类器以判定目标的类别;另一个分支是域分类器,训练域分类器以判定图像是源域图像还是目标域图像。

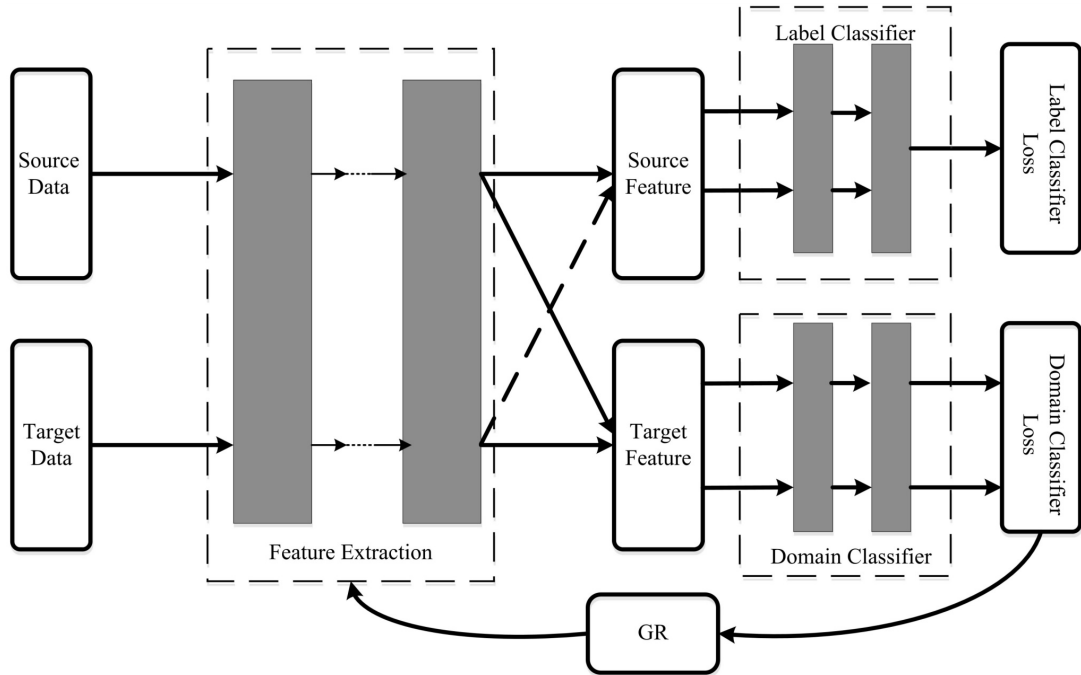


图3 域适应处理流程

定义 θ_f 代表特征提取网络参数, θ_y 是标签分类器参数, θ_d 是域分类器参数。 L_y 是标签分类器损失, L_d 是域分类器损失。 G_y 是标签分类器; G_d 是域分类器; G_f 是特征提取网络。 x_i 是第 i 个特征图, y_i 为其标签。网络参数通过式(2)、式(3)来更新,式(4)为网络的总损失函数。

$$\hat{\theta}_d = \max_{\theta_d} L_{\theta_d}(\hat{\theta}_f, \hat{\theta}_y, \theta_d) \quad (2)$$

$$(\hat{\theta}_f, \hat{\theta}_y) = \min_{(\theta_f, \theta_y)} L_{(\theta_f, \theta_y)}(\theta_f, \theta_y, \hat{\theta}_d) \quad (3)$$

$$L(\theta_f, \theta_y, \theta_d) = \sum_{i=1, d_i=0}^N L_y(G_y, (G_f(x_i; \theta_f); \theta_y) y_i) -$$

$$\lambda \sum_{i=1}^N L_d(G_d(G_f(x_i; \theta_f); \theta_y); \theta_d) y_i) \quad (4)$$

域分类器损失是二分类损失函数,取值为 $\{0, 1\}$,如果图像来自于源域则取值为0,否则取值为1。使用梯度反传 GR 将域分类器的误差信息反馈到特征提取网络,更新特征提取网络参数,促使特征提取网络生成对源域图像和目标域图像鲁棒的卷积特征(即域不变特征),减小源域图像和目标域图像特征分布不同时对目标检测网络性能的影响。

域适应网络训练过程为:

步骤 1: 固定特征提取网络参数 $\hat{\theta}_f$ 和标签分类器参数 $\hat{\theta}_y$, 训练学习域分类器参数 θ_d , 最大化域分类器参数 θ_d , 见公式(2)。

步骤 2: 将步骤 1 得到的参数 θ_d 代入公式(4)第二项, 最小化域分类器损失 L_d , 使域适应网络的域分类器能准确地预测卷积特征来自于源域图像还是目标域图像。

步骤 3: 步骤 1 和步骤 2 完成一次域训练过程, 然后固定域分类器参数 $\hat{\theta}_d$, 训练学习特征提取网络参数 θ_f 和标签分类器参数 θ_y 。即最小化特征提取网络参数 θ_f 和标签分类器参数 θ_y , 见公式(3)。

步骤 4: 将步骤 3 得到的参数 θ_f 和参数 θ_y 代入公式(4)中, 最小化标签分类器损失 L_y , 使标签分类器准确预测目标的类别。公式(4)第二项由于负号作用, 最大化域分类器损失 L_d , 使域分类器难以准确预测卷积特征来自于源域图像还是目标域图像。

步骤 5: 重复步骤 1 至步骤 4, 通过最大最小化对抗训练的方式使特征提取网络提取到对源域图像和目标域图像鲁棒的域不变特征。

2 实验结果与分析

2.1 数据集与参数设置

(1) 数据集介绍。

实验中使用的数据集为 NWPUVHR-10^[8] 遥感图像数据集。该数据集包括十类地物目标, 共有 800 幅超高分辨率遥感图像, 本实验使用其中的飞机目标遥感图像。实验中使用 NWPUVHR-10 数据集中所有的飞机图像作为源域, 对其标注图像级标签 0, 同时对数据集中的飞机目标遥感图像进行裁剪、旋转等操作, 扩充数据集, 最终得到 500 张带飞机目标的遥感图像作为源域数据。另外从 Google Earth 上收集 64 张包含飞机的遥感图像作为目标域数据, 主要包括晴天、有雾、低亮度等情况的飞机图像。遥感图像采集设备和条件、场景不同使数据集中的图像场景、光照等不同, 这里源域和目标域图像的来源不同, 具有较大的差异。由于遥感图像尺寸较大, 使用滑动窗口对目标域图像进行裁剪, 得到大小为 600×800 的图像, 并对这些图像进行旋转扩充数据集, 最终得到 500 张遥感图像, 为其标注图像级标签 1。最终目标检测网络的训练验证样本是 880 张, 测试集样本为 120 张。

RSOD-DATA 遥感数据集中飞机目标的遥感图像有 446 张, 每张遥感图像的场景不同。RSOD-DATA 中包括正常亮度的图像和亮度较低或过亮的图像。在公开数据集中正常光照的图占大多数, 因此实验中收集 RSOD-DATA 遥感数据集的飞机图像制作

了亮度较低的飞机目标数据集。故源域和目标域图像的来源相同, 但是亮度差异大。将数据集中的部分飞机图像进行低亮度处理得到 200 张低亮度图像, 同时将图像旋转 180°。将 RSOD-DATA 数据集中的原图像作为源域数据, 将处理后获得的低亮度图像作为目标域数据, 其中源域数据标签为 0, 目标域数据标签为 1。将目标域图像中的 30% 作为测试集数据, 另外 70% 的图像和源域数据一起作为训练集对模型进行训练。因此实验中训练验证集有 586 张图像, 测试集有 60 张图像。

(2) 实验环境及配置。

实验在配置了 Intel(R) Xeon(R) CPU E5-2690 v32.6 GHz 处理器、256 GB 内存、4 TB 硬盘, 安装 Ubuntu 系统的工作站上完成。训练和测试在 Caffe 深度学习平台完成, 使用一块 NVIDIA Tesla K40c 12 GB GPU 加速。使用 Image Net 数据集预训练 VGG16 模型初始化参数, 之后使用遥感数据集对模型进行训练, 学习率设置为 0.000 1, batch_size 为 2, gamma 为 0.1, 权重衰减为 0.000 5, 最大迭代次数为 80 000, 用随机梯度下降法更新网络权值。

(3) 评价指标。

评价指标包括平均精度 AP(Average Precision)、误检率 FAR(False Alarm Rate)和漏检率 MR(Missing Ratio), 其定义见式(5)~式(9)。AP 是查准率的平均值, 用来衡量检测效果; FAR 计算的是误检为飞机的数量占飞机总数的比例; MR 计算的是没有被检测出的飞机占飞机总数的比例。

$$r = \frac{TP}{TP + FN} \quad (5)$$

$$p = \frac{TP}{TP + FP} \quad (6)$$

$$AP = \frac{1}{n} \sum_{i=1}^n \max_{r \in [i-1/n, i/n]} p(r) \quad (7)$$

$$MR = \frac{FN}{TP + FN} \quad (8)$$

$$FAR = \frac{TN}{TP + TN} \quad (9)$$

其中, r 是查全率, p 是查准率, TP 是真正例, TN 是真正例, FP 是假正例, FN 是假反例。

2.2 实验结果分析

(1) NWPUVHR-10 数据集。

在 NWPUVHR-10 数据集上, 训练时先对图像级域分类网络进行训练, 之后再训练目标检测网络。表 1 给出了 FRCN VGG-16^[7]、MSCNN^[21]、Faster R-CNN^[8]、Faster R-CNN+FPN^[11]、SFP-DAM 模型在 NWPUVHR-10 测试集上的实验结果。表 1 中的对比模型均为两阶段目标检测模型, 这些方法采用相同的

训练集和测试集进行实验。比较方法的主干网络都是VGG-16, FRCN VGG-16的特征提取网络最终产生感受野大小固定的特征图进行检测框的预测及分类。MSCNN模型通过加入膨胀卷积结构,控制采样的步长,来增强网络对多尺度目标的表征能力,并将softmax分类损失函数替换成焦点分类损失。与Faster R-CNN相比, Faster R-CNN+FPN使用了金字塔结构,而 SFP-DAM不仅加入了金字塔结构,还增加了跳跃连接和域适应部分,从三者的结果可以看出增加金字塔结构、跳跃连接和域适应后的性能增加。

表1 不同模型在NWPUVHR-10测试集上的平均精度、误检率、漏检率和每幅图像测试时间

模型	平均精度/%	误检率/%	漏检率/%	每幅图像测试时间/s
FRCN VGG-16	85.4	16.9	18.6	0.21
MSCNN	97.6	10.2	10.5	0.16
Faster R-CNN	95.6	14.2	13.8	0.18
Faster R-CNN+FPN	96.2	12.5	12.3	0.18
SFP-DAM	98.2	5.4	8.3	0.19

部分检测结果如图4所示,从左到右依次为Faster R-CNN、Faster R-CNN+FPN和SFP-DAM的检测结果。从结果可见, FRCN VGG-16产生感受野大小固定的特征图进行检测框的预测以及分类,特征图的分辨率较低,不利于多尺度目标的检测,检测精度最低,漏检率和误检率较高。Faster R-CNN+FPN的特征金字塔结构有效利用了不同层次的特征,为浅层特征图增加了语义信息,对不同尺度目标的检测性能比Faster R-CNN有提高。SFP-DAM的跳跃连接金字塔和域适应使其更好地将与背景相近的小目标检测出来,克服图像光照与亮度不同等造成的检测困难,漏检率和误检率更低。

将表1及图4的结果综合来看, SFP-DAM模型包含了特征融合、域适应等更多的处理,但测试时间和其他方法基本相当,在同等测试时间下取得了性能的提升,并且在误检率、漏检率上提升更为显著。这些结果表明由于有效的金字塔结构和域适应模块的增加,使所提模型可以更好地应对不同场景及不同尺度的目标检测,模型具有更好的推广性能。

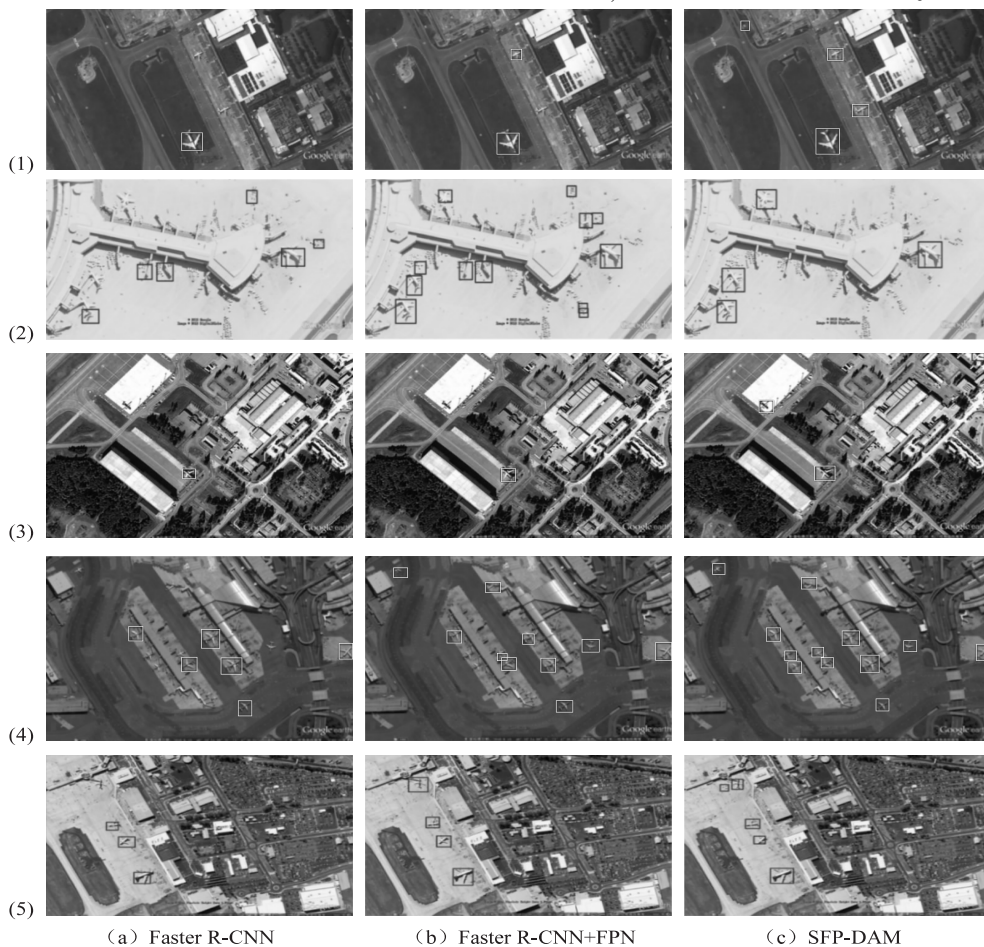


图4 NWPUVHR-10数据集部分实验结果

(2) RSOD-DATA数据集。

表2给出了FRCN VGG-16^[7]、MSCNN^[21]、Faster R-CNN^[8]、Faster R-CNN+FPN^[11]、Faster R-CNN+

SFPN^[11]和SFP-DAM模型在测试集上的部分实验结果。各方法在同样的训练集和测试集上进行实验。由表2可见各方法的定量评价结果都不如前一个数据集

的好,但是 SFP-DAM 依然取得了最好的结果, MSCNN 模型次之。这里 Faster R-CNN+FPN 和 Faster R-CNN+SFPN 相比少了跳跃连接, Faster R-CNN+SFPN 和 SFP-DAM 相比少了域适应部分,这三个模型结合 Faster R-CNN 模型在测试集上的结果可以表明,增加金字塔结构,在金字塔结构上增加跳跃连接,以及增加域适应网络分别带来的性能提升情况,即反映了消融实验的结果。

表 2 不同模型在 RSOD-DATA 测试集上的平均精度、误检率、漏检率结果

模型	平均精度/%	误检率/%	漏检率/%
FRCN VGG-16	50	22.3	34.8
MSCNN	58	20.1	20.2
Faster R-CNN	44	25.3	40.9
Faster R-CNN+FPN	54	21.6	35.5
Faster R-CNN+SFPN	56	20.3	33.6
SFP-DAM	62	18.2	18.5

在数据集 RSOD-DATA 上检测飞机目标的部分实验结果如图 5 所示,图中从左到右依次为原图, Faster R-CNN 模型、MSCNN 模型和 SFP-DAM 模型的检测结果。由图 5 可见,飞机目标较小,所占像素少,大小不一,特别地,由于亮度上的变化,目标和背景的对比度大大降低,增加了目标检测的困难性。Faster

R-CNN 模型对图中很小的飞机漏检多, MSCNN 进行了一定的特征融合操作,能够检测到部分小飞机,但是对和飞机形状相似的围栏等产生了误检的情况,同时对一些处于偏暗背景中的目标检测性能不佳。而 SFP-DAM 模型由于域适应的处理,能更好地检测处于低亮度背景中的飞机。

综上实验结果可得,跳跃特征金字塔结构能够应对不同尺度目标的检测,域适应网络可以提升不同场景下目标检测精度。和其他模型相比, SFP-DAM 在检测精度、误检和漏检方面的性能提升比较明显,表明了所提措施的有效性。

3 结束语

针对遥感图像中检测的目标尺度不一、小目标检测困难,以及训练集和测试集可能差异较大等问题,提出了基于跳跃特征金字塔的域适应遥感图像目标检测模型 SFP-DAM。模型一方面基于特征金字塔,设计特征金字塔结构的连接和融合方式,更好地保留细节特征并融合目标的语义特征,从而有利于不同尺度目标的检测。另一方面,增加域适应部分应对不同来源的训练和测试图像的特征差异,训练时也使用没有标记信息的目标域图像,使特征提取网络同时提取源域和目标域的特征,有效的特征提取使模型对测试图像提升了目标检测的性能,增加了模型的推广性能。实验结果表明,与其他模型相比, SFP-DAM 模型在 NWPUVHR-10 和 RSOD-DATA 数据集上取得了最优的结果,性能提升明显。

参考文献:

- [1] CHENG G, HAN J W. A survey on object detection in optical remote sensing images[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2016, 117: 11-28.
- [2] HAN J, ZHANG D, CHENG G, et al. Object detection in optical remote sensing images based on weakly supervised learning and high-level feature learning[J]. IEEE Transactions on Geoscience and Remote Sensing, 2015, 53(6): 3325-3337.
- [3] 许德刚, 王露, 李凡. 深度学习的典型目标检测算法研究综述[J]. 计算机工程与应用, 2021, 57(8): 10-25.
- [4] 陶晓力, 刘宁钟, 沈家全. 基于深度信息融合的航拍车辆检测[J]. 计算机技术与发展, 2019, 29(9): 117-121.
- [5] KRIZHEVSKY A, SUTSKEVER I, HINTON G, et al. ImageNet classification with deep convolutional neural networks[J]. Advances in Neural Information Processing Systems, 2012, 25(2): 1097-1105.
- [6] GIRSHICK R, DONAHUE J, DARRELL T, et al. Region-based convolutional networks for accurate object detection and segmentation[J]. IEEE Transactions on Pattern Analysis

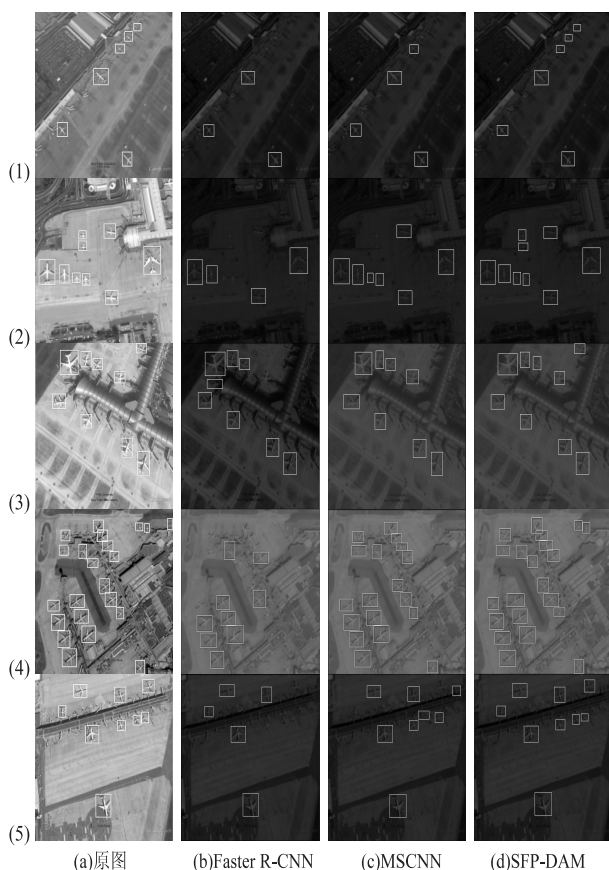


图 5 RSOD-DATA 实验结果

- & Machine Intelligence, 2015, 38(1):142–158.
- [7] GIRSHIK R. Fast R-CNN[C]//The IEEE international conference on computer vision. Santiago; IEEE, 2015:1440–1448.
- [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] HONG S, ROH B, KIM K H, et al. PVANet: lightweight deep neural networks for real-time object detection[J]. arXiv:1611.08588, 2016.
- [10] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]//European conference on computer vision. Amsterdam; Springer, 2016:21–37.
- [11] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Honolulu; IEEE, 2017:2117–2125.
- [12] LI Z, LANG C, LIEW J, et al. Cross-layer feature pyramid network for salient object detection[J]. IEEE Transactions on Image Processing, 2021, 30:4587–4598.
- [13] 单义, 杨金福, 武随烁, 等. 基于跳跃连接金字塔模型的小目标检测[J]. 智能系统学报, 2019, 14(6):1144–1151.
- [14] HONG W X, WANG Z Z, YANG M, et al. Conditional generative adversarial network for structured domain adaptation[C]//The IEEE conference on computer vision and pattern recognition. Salt Lake City; IEEE, 2018:1335–1344.
- [15] CHEN Y H, LI W, SAKARIDIS C, et al. Domain adaptive faster R-CNN for object detection in the wild[C]//The IEEE conference on computer vision and pattern recognition. Salt Lake City; IEEE, 2018:3339–3348.
- [16] CHEN J, WU X, DUAN L, et al. Sequential instance refinement for cross-domain object detection in images[J]. IEEE Transactions on Image Processing, 2021, 30:3970–3984.
- [17] 杨文武, 普园媛, 赵征鹏, 等. 基于多层特征描述及关系学习的智能图像情感识别[J]. 陕西师范大学学报:自然科学版, 2019, 47(5):40–48.
- [18] FERNANDO B, HABRARD A, SEBBAN M, et al. Unsupervised visual domain adaptation using subspace alignment[C]//Proceedings of the IEEE international conference on computer vision. Darling Harbour; IEEE, 2013:2960–2967.
- [19] DUAN L, XU D, TSANG I W, et al. Visual event recognition in videos by learning from web data[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 34(9):1667–1680.
- [20] DUAN L, TSANG I W, XU D. Domain transfer multiple kernel learning[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(3):465–479.
- [21] CAI Z, FAN Q, FERIS R S, et al. A unified multi-scale deep convolutional neural network for fast object detection[C]//European conference on computer vision. Amsterdam; Springer, 2016:354–370.