

基于径向基激活函数的高光谱小目标检测

王勃凡, 赵海涛*

华东理工大学信息科学与工程学院, 上海 200237

摘要 基于深度学习的目标检测方法是当前计算机视觉领域的研究热点,但在小目标的检测问题上,基于深度学习的检测器存在较多的漏检。高光谱图像的每个像元包含了物质的光谱信息,能够提升小目标的检测率。然而,高光谱图像的相邻波段相关性高,需要从中选取具备代表性的波段以降低计算冗余。为此,提出了一种高光谱小目标检测模型,使用径向基激活函数(RBAF)进行光谱筛选与目标检测。具体而言,针对高光谱图像波段冗余的特点,利用 RBAF 设计注意力机制进行光谱维的特征筛选;针对小目标纹理模糊,相对于背景不显著的特点,先对输入图像进行分辨率重建,随后利用 RBAF 构建径向基目标输出子网络(RBOON),以加强目标分类。为了简化模型,将光谱筛选与分辨率重建整合为注意力分辨率重建子网络(ABRRN),配合径 RBOON,检测模型能够筛选特定光谱,抑制虚警,从而提高小目标检测的正确率。高光谱小目标检测实验表明,本研究方法可以使两种检测精度评价指标 AP50 和 AP50:95 分别提升 5.4% 和 0.2%,相较其他方法更具备优势。

关键词 机器视觉; 图像检测系统; 高光谱图像; 目标检测; 径向基函数; 注意力机制

中图分类号 TP391

文献标志码 A

doi: 10.3788/AOS202141.2311001

Small Object Detection in Hyperspectral Images Based on Radial Basis Activation Function

Wang Bofan, Zhao Haitao*

School of Information Science and Engineering, East China University of Science and Technology, Shanghai 200237, China

Abstract Object detection methods based on deep learning are the current research focus of computer vision. However, when detecting small objects, existing detectors often suffer from missing detection. Every pixel of hyperspectral images contain the spectral information of small object materials. Therefore, they can provide additional support for improving the detection performance on small objects. However, the adjacent bands of hyperspectral images are highly correlated. It is thus necessary to select representative bands to reduce the computational redundancy. In response, this paper proposed a hyperspectral small object detection model, which used the radial basis activation function (RBAF) to carry out spectral screening and object detection. Specifically, in view of the band redundancy of hyperspectral images, an attention mechanism based on the RBAF was designed for spectral screening. As for the high texture fuzziness and low distinguishability against the background of small objects, the resolution of input images was reconstructed first. Then, a radial basis object output network (RBOON) based on the RBAF was constructed to enhance object classification. For model simplification, spectrum screening and resolution reconstruction were integrated into an attention-based resolution reconstruction network (ABRRN). With the combination of the ABRRN and RBOON, the detection model can screen the specific spectrum and suppress false alarms and thus improve the accuracy of small object detection. Hyperspectral small object detection experiments show that the proposed method improves the two detection accuracy criteria, namely AP50 and AP50:95, by 5.4% and 0.2%, respectively, which means it is better than other methods.

收稿日期: 2021-04-15; 修回日期: 2021-05-13; 录用日期: 2021-06-10

基金项目: 国家自然科学基金(61973122)

通信作者: *Haitaozhao@ecust.edu.cn

Key words machine vision; image detection system; hyperspectral image; object detection; radial basis function; attention mechanism

OCIS codes 110.2970; 110.4234; 100.6640

1 引言

随着深度学习的发展,基于卷积神经网络的目标检测算法取得了巨大的进展^[1-3]。然而小目标检测一直是目标检测领域的热点和难点问题之一。针对可见光图像,为了提高小目标的检测精度,有学者从多尺度检测^[4-5]、专注小目标样本训练^[6-7]、提高分辨率^[8-9]等方面展开研究。但与中、大目标的检测性能相比,现有方法得到的检测器针对小目标检测精度仍然存在较大的差距。

小目标检测的主要难点是在空间维度上小目标所包含的像素点少,分辨率低,卷积神经网络难以提取有效的纹理特征,许多学者从特征尺度与分辨率的角度进行研究。Lin 等^[5]提出特征金字塔网络(FPN),通过将不同尺度的网络中间特征进行聚合,形成特征金字塔以提高针对小目标的检测性能。Singh 等^[6]提出特征金字塔尺度归一化(SNIP)算法,在训练中只针对尺度与预训练数据集相似的小目标进行训练,并采用特征金字塔进行多尺度检测。王伟锋等^[9]将多个不同大小的卷积核拼接使用,提出感受野模块,以适应不同大小的目标。这些尺度与分辨率策略能够一定程度上改善检测器对于小目标的检测性能,但由于可见光图像的信息有限,仍然会出现大量的虚警或漏检测情况。

通常而言,各类物质对电磁辐射的吸收和反射特性与电磁波所在的波段显著相关。而传统的图像只能依靠可见光波段的纹理细节检测目标,在目标较小或图像分辨率不足的情况下难以获取足够的检测信息。高光谱成像传感器能够采集所观测场景的三维光谱图像,这些光谱波段被划分成几十到上百个通道,光谱分辨率高且能够捕捉到不同物质在不同波段通道上的差异。因此,高光谱图像被广泛地应用于遥感影像分类^[10-11]、环境与气体监测^[12-13]、矿物勘探^[14-15]、农业与食品分析^[16-17]等对物质成分信息敏感的领域。然而,高光谱图像相邻波段的相关性高,冗余信息一方面带来高计算复杂度,另一方面会使检测器倾向于选择相关性高而非鉴别力强的波段进行目标检测。因此需要进行波段筛选,以使网络关注具有目标鉴别能力的波段。

高斯径向基函数是一类取值仅依赖样本距离中

心点距离的函数,可以描述样本点与模型中心点的相似程度。在高光谱小目标检测中,目标点像元在每个光谱通道上的分布通常与目标的物质成分相关,并以该物质的漫反射光谱为中心成径向分布。因此径向基函数能够筛选特定的光谱波段,压缩通道数。此外,在目标判别输出阶段,由于小目标与背景点的差异不显著,使用传统 Sigmoid 激活范围较大,易使与目标相似的背景点落入正输入饱和区,而径向基函数由于其只激活一个径向范围的特性,可以加强对目标的筛选,抑制复杂纹理的背景点产生的虚警。基于上述特点,本文提出一种基于径向基激活函数(RBAF)的高光谱小目标检测模型,主要包含两个子网络:

1)注意力分辨率重建子网络(ABRRN),该子网络起到两个主要作用:分辨率重建与光谱筛选。分辨率重建能够改善目标过小,与预训练数据尺度不匹配的问题;光谱筛选用于压缩高光谱通道数,降低冗余。具体而言,ABRRN 先采用反卷积层对高光谱图像进行分辨率重建,形成高分辨率特征图,使小目标在高分辨率特征图中的尺度分布与骨干网络预训练数据中常见的目标尺度分布相似;随后 ABRRN 利用 RBAF 执行注意力机制,自适应地调整筛选的光谱波段与激活范围,从而完成光谱筛选。

2)径向基目标输出子网络(RBOON),该子网络使用 RBAF 对目标判别输出分类进行激活,只对特征空间内一个径向范围进行激活,避免背景点落入 Sigmoid 函数的正输入饱和区内,从而抑制小目标虚警。

2 研究方法

早期的研究认为,激活函数应具备单调性,以利于模型进行凸优化。近年来越来越多的研究表明单调性并非激活函数的必要特性^[18-19],且 Ramachandran 等^[20-21]提出的 Swish、Mish 激活函数更表明非单调的激活函数具备更好的性能。在可见光目标检测中,大多数方法采用类似 Sigmoid 的单调函数作为注意力机制^[22-23]以及目标判别输出分类^[24-25]的激活函数,然而在高光谱小目标检测中,Sigmoid 在正输入区域的激活范围过大,难以有效筛选光谱,同时易将具备复杂纹理的背景点误判为

目标。因此,本文构造光谱筛选径向基激活函数(RBAF-SS)和小目标判别径向基激活函数(RBAF-SOD),分别用于光谱筛选与小目标判别。

2.1 光谱筛选径向基激活函数

在光谱筛选中,RBAF 应针对径向范围内的光谱波段特征进行激活,从而使模型更关注鉴别能力较高的波段。然而,高光谱图像在光谱维度上只包含固定通道的离散波段,无法在光谱维度上对参数进行求导与反向传播。为了将光谱维度视为一个连续整体,充分利用光谱特性,同时保证其反向传播,本研究设计了 RBAF-SS。先对通过拉格朗日插值获取像元内指定波段的值,再对该值进行径向基计算,其表达式为

$$y_{b,w,h,d_{out}} = \exp \left[-\frac{(x_{b,w,h,s_{d_{out}}} - c_{d_{out}})^2}{\eta + 2\lambda_{d_{out}}^2} \right],$$

$$\{b \in \{1,2,\dots,B\}, \omega \in \{1,2,\dots,W\},$$

$$h \in \{1,2,\dots,H\}, d_{out} \in \{1,2,\dots,D_{out}\},$$

$$s_{d_{out}} \in (0, D_{in})\}, \quad (1)$$

式中, $y_{b,w,h,d_{out}}$ 为激活函数在第 b 个训练样本,输出特征图中坐标 w,h 的特征向量第 d_{out} 通道的元素,输出特征图的形状为 $B \times W \times H \times D_{out}$; $x_{b,w,h,s_{d_{out}}}$ 表示输入高光谱特征图在第 b 个训练样本中,坐标 w,h 的像元,波段 $s_{d_{out}}$ 的值,其由输入的高光谱向量 $x_{b,w,h}$ 在 $s_{d_{out}}$ 波段经过拉格朗日插值产生; $s_{d_{out}}, c_{d_{out}}$ 和 $\lambda_{d_{out}}$ 为激活函数的参数,其中 $s_{d_{out}}$ 表示筛选的光谱波段, $c_{d_{out}}$ 表示该波段值的激活中心, $\lambda_{d_{out}}$ 表示激活范围; η 为提高运算稳定性的正小参数。整个隶属度激活函数层将通道数从 D_{in} 变为 D_{out} ,如图 1 所示。

对于特征图中所有像元的 D_{in} 维光谱向量,首先将其展开,随后挑选 D_{out} 个光谱波段,对这些光谱波段的值进行径向基计算,从而得到 D_{out} 维的输出向量。所挑选的波段以及波段的激活范围均由模型自适应训练得到,从而实现三维空间与光谱结合的 RBAF。

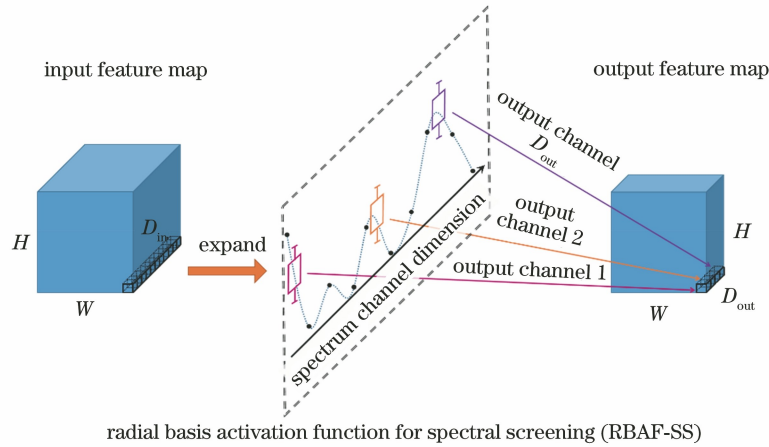


图 1 用于光谱筛选的径向基激活函数示意图

Fig. 1 Schematic diagram of the radial basis activation function for spectral screening (RBAF-SS)

2.2 小目标判别径向基激活函数

在小目标判别中,RBAF 应针对特征图的分布对特定的空间区域的候选点进行激活,加强模型对小目标从背景中的筛选能力。本文设计了 RBAF-SOD,其表达式为

$$y_{b,w,h,d_{out}} = \prod_{d_{in}=0}^{D_{in}} \exp \left[-\frac{(x_{b,w,h,d_{in}} - c_{d_{out},d_{in}})^2}{\eta + 2\lambda_{d_{out},d_{in}}^2} \right],$$

$$\{b \in \{1,2,\dots,B\}, \omega \in \{1,2,\dots,W\},$$

$$h \in \{1,2,\dots,H\}, d_{out} \in \{1,2,\dots,D_{out}\},$$

$$d_{in} \in \{1,2,\dots,D_{in}\}\}, \quad (2)$$

式中, $x_{b,w,h,d_{in}}$ 为激活函数直接选取的第 b 个训练样本,输入特征图中坐标 w,h 的特征向量第 d_{in} 通

道的元素,无需进行插值;整个激活函数为 D_{out} 个无归一化的 D_{in} 维高斯函数,将特征图的通道数从 D_{in} 变为 D_{out} 。RBAF-SOD 相对于传统 Sigmoid 函数,激活范围更小,对目标的筛选更严格,从而能够在目标输出子网络中抑制虚警。

2.3 总体网络结构

本文所提出的网络结构可分为四阶段:第一阶段(Stage 1)为分辨率重建与光谱筛选阶段,高光谱图像经过 2 个 ABRRN,完成光谱筛选;第二阶段(Stage 2)和第三阶段(Stage 3)采用 CenterNet^[24]的框架,其中第二阶段(Stage 2)采用经过预训练的残差网络(ResNet-18)^[26]作为骨干网络,提取图像

特征; 第三阶段(Stage 3)由可变形卷积网络(DCN)^[27]与反卷积层(Deconv)组成,其中 DCN 用于增强特征图空间变换能力,Deconv 用于重建空间信息;第四阶段(Stage 4)使用 RBOON 输出无锚框的目标检测结果,包括目标存在热点图、目标位置偏

置图和目标边框大小图,经过解码即可得到目标输出,除 ABRRN 和 RBOON 将在 2.4 与 2.5 节详细介绍外,其余所有的卷积层、DCN^[28]与反卷积层后均添加批归一化层(BN)和 ReLU 激活函数,完整网络结构如表 1 所示。

表 1 总体网络结构

Table 1 Overall structure of the detection network

Stage	Sub-network	Block	Layer details			Output shape / (pixel×pixel×pixel)
Input	—	—	—	—	384×192×25	
1	ABRRN	—	Deconv	Conv+ReLU	768×384×6	
				Conv RBAF-SS GAP		
	ABRRN	—	Element-wise multiplication Same as the above ABRRN	Conv+BN+ReLU	1536×768×3	
				Max pooling		
2	ResNet-18 backbone	Down sampling	Conv+BN+ReLU	Identity	384×192×64	
		ResBlock (no down sampling)	Conv+BN	Element-wise Addition	192×96×128	
		ResBlock (down sampling)	Conv+BN	Conv+BN		
		ResBlock (no down sampling)	Element-wise addition	192×96×128		
		ResBlock (down sampling)	Same as the above ResBlock (no down sampling)			
		ResBlock (no down sampling)	Same as the above ResBlock (down sampling)	96×48×256		
		ResBlock (no down sampling)	Same as the above ResBlock (no down sampling)	96×48×256		
		ResBlock (down sampling)	Same as the above ResBlock (down sampling)	48×24×512		
3	Up sampling network	Up sampling block	DCN+BN+ReLU	96×48×256		
		Up sampling block	Deconv+BN+ReLU			
		Up sampling block	Same as the above upsampling block	96×48×128		
4	RBOON	—	Conv+ReLU	Conv+ReLU	96×48×1/2/2	
			RBAF-SOD	Conv		Conv

2.4 注意力分辨率重建子网络

高光谱图像相邻波段存在一定的相关性,需要运用注意力机制对光谱进行筛选;另一方面,针对小目标检测问题,Singh 等^[6]指出,一般目标检测所采用的预训练骨干网络,其预训练数据集中的目标尺度分布与实际存在的小目标尺度分布存在较大差异,因此需要进行分辨率重建以适配预训练数据集的尺度。设计了 ABRRN,将注意力光谱筛选和分辨率重建集成在单个子网络中。在提高图像分辨率的同时,进行光谱筛选。其结构如图 2 所示。

在 ABRRN 中,对于大小 $W \times H \times C$ 为输入特征图,将其分为两条数据流分支。其中一条分支负责重建图像的分辨率,采用反卷积层将特征图分辨率提升 2 倍,同时将通道数压缩到原始通道数的 1/3。另一条分支负责生成高光谱通道注意力向量,首先经过 2 层卷积层和 RBAF-SS,筛选出 1/3 的光谱波段,随后经过全局均值池化层(GAP),得到 $1 \times 1 \times (C/3)$ 的全局光谱注意力向量。最终,RBAF 生成的光谱注意力向量与分辨率重建后的特征图相乘,形成经过光谱筛选的高分辨率特征图。

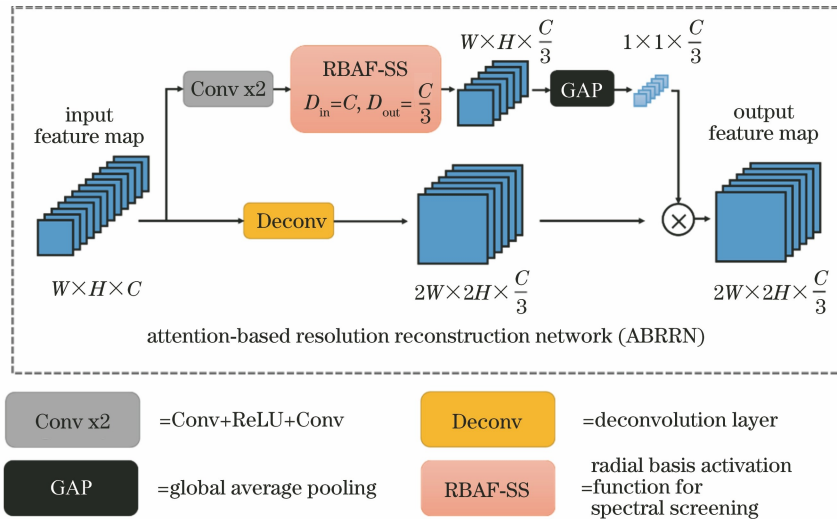


图 2 注意力分辨率重建子网络

Fig. 2 Attention-based resolution reconstruction network (ABRRN)

2.5 径向基目标输出子网络

目标输出子网络可分为锚框目标输出子网络和无锚框目标输出子网络。早期目标检测器大多使用锚框目标输出子网络^[1,3]。近几年,一些学者提出直接基于特征点进行目标判别的无锚框目标输出子网络,如 CornerNet^[28]、CenterNet 等^[24]。文献[29-30]指出,基于锚框的输出子网络受锚框形状与大小的约束,无法适应目标的大幅畸变与尺度变换。因此对于小目标的应用场景,本文采用无锚框目标输出子网络。同时,使用 RBAF-SOD 对进行目标输出,形成 RBOON。该目标输出子网络只激活特征空间内落入的特定范围的像素点,减少对具备复杂纹理的背景点的误判,从而抑制小目标虚警。

如图 3 所示,RBOON 包含三支信息:目标存在热点图、目标位置偏置图和目标边框大小图。目标存在热点图用于描述在 $W \times H$ 大小的空间中各位置存在第 N 类目标的概率,目标位置偏置图用于修正目标位置从原始输入图像分辨率大小 $W_{in} \text{ pixel} \times H_{in} \text{ pixel}$ 到特征图分辨率大小 $W \text{ pixel} \times H \text{ pixel}$ 的下采样偏差,目标边框大小图用于描述在各位置存在的目标边框高度与宽度。通过设置阈值从目标存在热点图中筛选位置坐标,并从目标边框大小图和目标位置偏置图中获取边框大小和位置偏置,即可解码得出目标检测的输出。

在目标存在热点图的网络分支中,使用 RBAF-SOD 对目标分类输出进行激活。针对每类目标,其

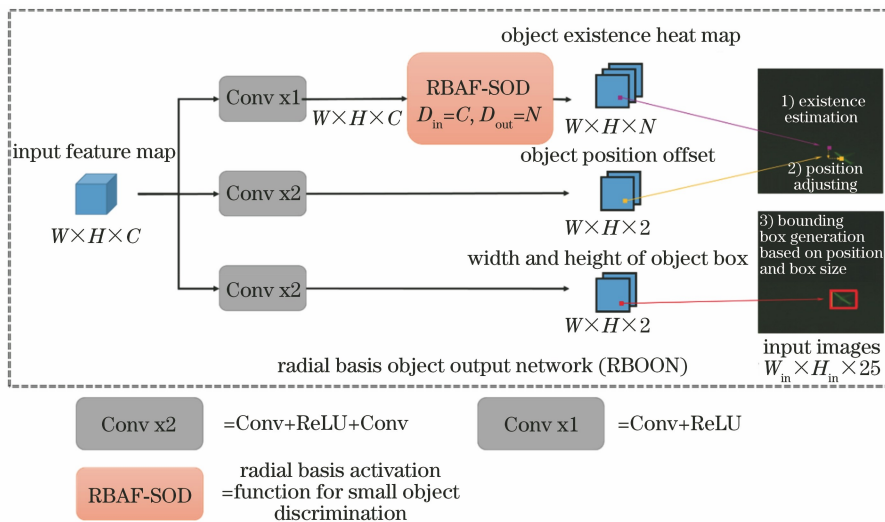


图 3 径向基目标输出子网络(RBOON)

Fig. 3 Radial basis object output network (RBOON)

在每个特征维度上具备可学习的激活区域,且只对落入该区域范围内的特征图像素进行激活,从而加强网络对小目标从背景点中的筛选,抑制虚警。

3 实验环境

3.1 实验平台与数据集

本文所提出的模型均使用 Pytorch 深度学习框架实现,其中 RBAF 采用 C++ 语言编写并作为 Pytorch 的 cuda 拓展整合进模型,从而能够对速度和内存进行合理优化。模型训练与测试平台为单块 NVIDIA RTX 3080。

为了验证模型性能,本文采用自制高光谱数据集进行实验,由手持的 XIMEAMQ022HG-IM 光谱仪于地面对空中的飞机进行拍摄,模拟地对空目标检测的应用场景。随后针对每张图像进行人工标注,标注只包含单个图像中目标框的位置与大小,不包含跨图像间的目标关联,用于验证非时序静态检测模型。其中包含 14555 张高光谱图像和 9735 个

飞机目标。每张高光谱图像包含 25 个光谱通道,响应谱段范围为 600~975 nm,图像原始分辨率为 409 pixel×217 pixel。数据集按照 9:1 的比例划分为训练集与测试集,其中训练集包含 12955 张高光谱图像以及 9006 个标注目标,测试集包含 1600 张高光谱图像以及 729 个标注目标。依照微软目标检测数据集(MS COCO)所提出的标准^[31],将目标框所占像素小于 1024 pixel(32 pixel×32 pixel)、介于 1024 pixel(32 pixel×32 pixel)与 9216 pixel(64 pixel×64 pixel)之间、以及大于 9216 pixel(64 pixel×64 pixel)的目标分别认定为小目标、中等目标与大目标。则在训练集中,小目标、中等目标的数量分别为 7869、1136;在测试集中,其数量分别为 525、203。本文所制数据集的拍摄场景为地对空拍摄,所拍摄的飞机目标以小目标为主,以一定比例的中等目标辅助训练,以验证不同检测器对小目标的检测性能。图 4 所示为高光谱图像与目标标注的示例。

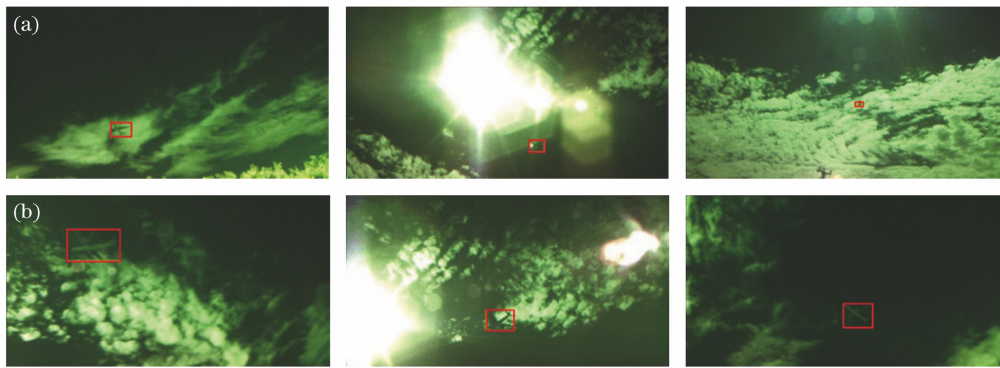


图 4 高光谱数据集示例。(a)小目标;(b)中等目标

Fig. 4 Examples of hyperspectral data sets. (a) Small objects; (b) medium objects

3.2 训练过程与评价指标

在网络输入过程中,所有高光谱图像均调整到 384 pixel×192 pixel 的分辨率,使其分辨率为 32 (ResNet-18 的最大下采样)的整数倍,避免造成失真。在训练过程中,采用随机裁剪和随机水平翻转两种数据增强方法,避免网络过拟合。优化过程采用 Adam 算法,共训练 70 个数据循环(epoch),批次大小(batch size)为 8。其初始学习率为 1.25×10^{-4} ,并分别在第 45、55、60 次循环(epoch)结束后将学习率降低至原来的 0.22,整个训练过程耗时约 25 h。

实验的评价指标采用目标检测常用的平均精度(AP)。列出了面积交并比(IoU)在 0.5 至 0.95 的综合平均精度(AP50:95),以及 IoU 固定在 0.5 的

平均精度(AP50)。此外,本文的实验还给出了不同 IoU 阈值下的虚警率(FAR),以对比验证 RBOON 对虚警的抑制能力。在计算 FAR 以及显示目标检测框时,所有检测算法的置信度阈值均设置为 0.3。

4 实验结果与分析

4.1 光谱筛选的定性分析

在 ABRRN 中,基于 RBAF 的注意力机制是为了在连续光谱维度上进行光谱筛选,因此对其筛选效果进行定性分析。对 ABRRN 的中间层光谱特征像元进行采样,可得到 25 个光谱通道的特征分布,结合像元真实标注,以及光谱筛选 RBAF 所筛选的光谱与激活范围,观察光谱筛选的效果,图 5 所示。

图中横坐标为光谱的 25 个维度特征,纵坐标为各维度光谱特征像元的值,形成像元在光谱维度上的连续特征曲线,每一个目标点的像元采样都以一条曲线表示。为了便于可视化,所有样本曲线均以

半透明形式堆叠在同一个图内,因此图中颜色较深即表示曲线分布密集,目标特征具备较强的可鉴别性。同时以虚线构成的框线标注了 RBAF-SS 的筛选波段。

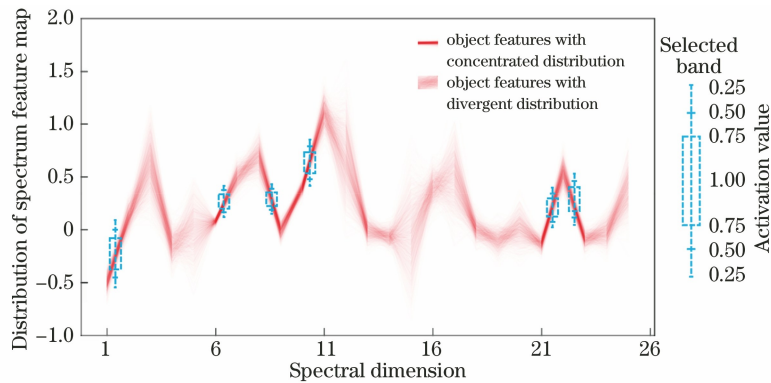


图 5 基于径向基注意力机制的光谱筛选定性分析

Fig. 5 Qualitative analysis of spectral selection based on RBAF

图 5 右侧绘制了每个所选波段 RBAF-SS 的激活区域与激活值的映射参照图,矩形上下边表示激活值为 0.75 的位置,矩形上下的 2 组线由内至外分别代表激活值为 0.5 和 0.25 的位置。图中 6 个虚线所画的框线表示 6 个不同的筛选波段,每个筛选

的波段位置由参数 s 决定,激活区域由参数 c 与 λ 共同作用决定,参数均由网络训练得到。由图 5 可以看出, RBAF-SS 所选取的波段都位于颜色较深处,即目标光谱特征分布较集中,具备较强的可鉴别性的波段。

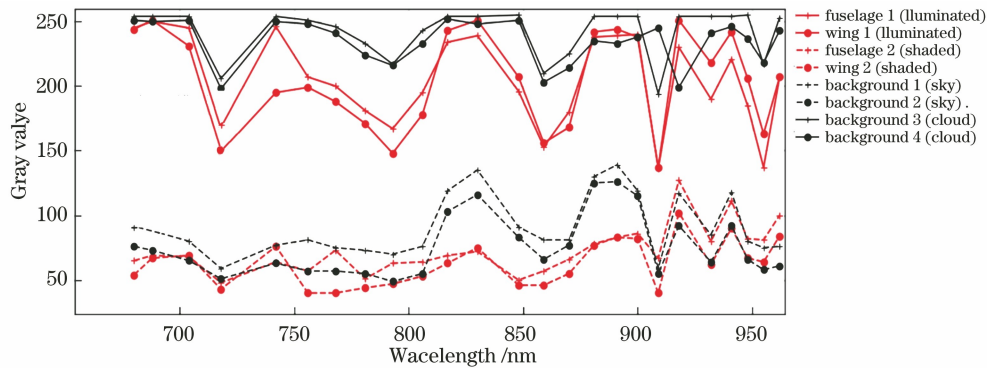


图 6 飞机目标特征的光谱曲线

Fig. 6 Spectral curve of airplane object features

图 6 展示了本研究中数据集中飞机目标特征的光谱曲线,横坐标为光谱波长,纵坐标为目标特征像元的灰度值。其中红色曲线为飞机目标的光谱,黑色曲线为背景光谱。在图左侧所标注的 8 种线形分别代表光照中的机身 1,光照中的机翼 1,背光的机身 2,背光的机翼 2,天空的背景样例 1,天空的背景样例 2,白云的背景样例 3,白云的背景样例 4。

由图 6 可以看出,在上半段的高光照环境下,750~800 nm 以及 825~875 nm 的波段具备较强的可鉴别性,在下半段的低光照环境下,800~900 nm 的波段具备较强的可鉴别性。光照对原始的目标光谱特征的影响较大,但背景与飞机目标仍是可区

分的。

4.2 虚警抑制的定性与定量分析

对于 RBOON,采用 RBAF 代替 Sigmoid 函数以加强对目标点的筛选,以期望其达到抑制虚警的效果,因此对其分别进行定性与定量分析。图 6 显示了本研究方法与 4 种基于 Sigmoid 的目标输出层方法的实验结果对比,其中红色粗线框表示算法检测的目标框,黄色细线表示人工标注的真实目标框。在以下的分析中,将从左到右 3 个场景依次称为场景 1、场景 2 和场景 3。由图 7 可以看出,在相同的置信度阈值下, Faster RCNN^[1]、YOLOv3^[25] 和 FCOS^[32] 在场景 1 或场景 2 中均由于背景纹理或噪

点而出现了虚警,CenterNet^[24]没有出现虚警情况,体现了其无锚框输出层针对小目标的优势。然而,对于存在极端小尺度目标的场景 3, Faster RCNN^[1]、YOLOv3^[25]、FCOS^[32]和 CenterNet^[24]均

产生了漏检测,本文所提出的 RBOON 在场景 1 与场景 2 中有效抑制了虚警,同时在场景 3 中成功检测到了小目标。

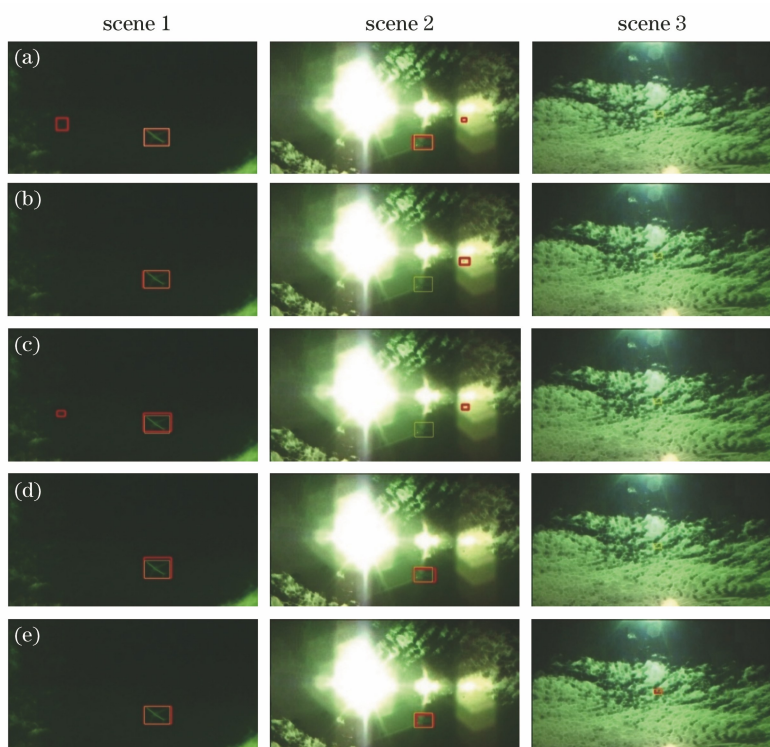


图 7 本文方法与 4 种基于 Sigmoid 的目标输出层方法的实验结果对比。(a) Faster RCNN^[1] (ResNet-50); (b) YOLOv3^[25] (Darknet-53); (c) FCOS^[32] (ResNet-50); (d) CenterNet^[24] (ResNet-18); (e) 本研究方法 (ResNet-18)
Fig. 7 Experimental results of the proposed method are compared with four approaches with sigmoid based object output layers. (a) Faster RCNN^[1] (ResNet-50); (b) YOLOv3^[25] (Darknet-53); (c) FCOS^[32] (ResNet-50); (d) CenterNet^[24] (ResNet-18); (e) proposed method (ResNet-18)

为了进一步验证本研究方法对虚警的抑制能力,对其进行定量分析。将 IoU 阈值在 0.4~0.65 范围内(以 0.05 为间隔)的 FAR 进行计算,并以曲线形式可视化,如图 8 所示。其中提出的径向基方法在 0.40、0.45、0.50、0.55、0.60、0.65 的 IoU 阈值下,虚警率分别为 2.1%、2.5%、4.0%、9.2%、16.1%、26.1%,低于其余 4 种基于 Sigmoid 的目标输出方法,该径向基方法有效降低了小目标检测中的虚警率。

4.3 检测精度与消融实验

检测精度是评价小目标检测性能的核心指标。表 2 和表 3 分别为 4 种对照方法的检测精度与本研究方法的消融实验。对于两种精度评价指标 AP50 和 AP50:95,分别列出了针对小目标、中等目标和全部目标的精度。需要注意的是,在消融实验部分(如表 3 所示),由于本研究方法在框架上继承

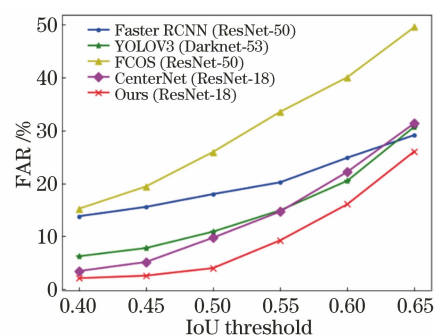


图 8 不同方法在各 IoU 阈值下的虚警率
Fig. 8 False alarm rate of different approaches under different IoU threshold

CenterNet^[24], 因此以其为基准。在 Proposed-ABRRN 实验中,添加了 ABRRN。在 Proposed-RBOON 的实验中,添加了 RBOON。

表 2 检测精度与消融实验

Table 2 Detection accuracy and ablation experiment

Methods	AP50 /%			AP50:95 /%			Time /s
	Small	Medium	All	Small	Medium	All	
Faster RCNN ^[1] (ResNet-50)	59.4	69.5	62.2	21.8	30.4	24.2	0.044
YOLOv3 ^[25] (Darknet-53)	51.8	67.4	56.2	18.5	28.5	21.3	0.015
FCOS ^[32] (ResNet-50)	25.0	46.7	31.1	9.0	9.7	8.6	0.041

表 3 消融实验

Table 3 Ablation experiment

Methods	AP50 /%			AP50:95 /%			Time /s
	Small	Medium	All	Small	Medium	All	
CenterNet ^[24] (ResNet-18)	60.2	70.4	63.1	21.5	28.3	22.7	0.011
Proposed-ABRRN (ResNet-18)	64.9	73.8	67.4	23.4	27.0	24.4	0.020
Ours-RBOON (ResNet-18)	60.7	70.6	63.5	21.6	28.4	22.9	0.011
Proposed (ResNet-18)	65.0	74.2	67.6	23.5	26.7	24.4	0.020

由表 3 可看出,相较于原始 CenterNet^[24],本研究中的 ABRRN 和 RBOON 均能提升网络检测精度。其中 ABRRN 将 AP50 和 AP50:95 分别提升了 4.3% 和 1.7%,在小目标上的提升分别为 4.7% 和 1.9%,同时,其耗时提高了 0.010 s,其时间上的损失是可以接受的。RBOON 也能一定程度上提升检测精度,在 AP50 和 AP50:95 上分别提升 0.4% 和 0.2%,同时耗时与原始 CenterNet^[24] 相同。

在 ABRRN 和 RBOON 的共同作用下,本研究方法 AP50 和 AP50:95 分别达到 67.6% 和 24.4%,在速度和精度上均优于最佳对照方法 Faster RCNN^[1]。在中等目标中,本研究方法的 AP50:95 略低于 Faster RCNN^[1],其可归因与两阶段方法在目标定位准确性方面的优势。总体而言,本研究方法基本实现了抑制虚警以及提升小目标检测精度的设计目标。

5 结 论

针对目标检测对小目标检测精度不足的问题,采用 CenterNet 目标检测框架,引入高光谱图像,并基于 RBAF 设计 ABRRN 和 RBOON,以提高小目标的检测精度。在小目标高光谱数据集上,先验证了径向基方法在光谱筛选与抑制虚警方面的作用,随后验证其检测精度。本研究的方法达到 67.6%

的 AP50 和 24.4% 的 AP50:95,相较 CenterNet 分别提升了 5.4% 和 0.2%。下阶段的工作是尝试将 RBAF 移植到其他高光谱计算机视觉领域中,验证其性能。

参 考 文 献

- [1] 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.
- [2] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 779-788.
- [3] 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.
- [4] Ji Z, Kong Q K, Wang J. Object detection algorithm guided by dual attention models [J]. Laser & Optoelectronics Progress, 2020, 57(6): 061008. 冀中, 孔乾坤, 王建. 一种双注意力模型引导的目标检测算法[J]. 激光与光电子学进展, 2020, 57(6):

- 061008.
- [5] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection [C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 936-944.
- [6] Singh B, Davis L S. An analysis of scale invariance in object detection-SNIP [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: 3578-3587.
- [7] Singh B, Najibi M, Davis L S. SNIPER efficient multi-scale training[C]//2018 Annual Conference on Neural Information Processing Systems, December 3-8, 2018, Montreal, Canada. MA: Massachusetts Institute of Technology Press, 2018: 9333-9343.
- [8] Zhou P, Ni B B, Geng C, et al. Scale-transferrable object detection[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: 528-537.
- [9] Wang W F, Jin J, Chen J M. Rapid detection algorithm for small objects based on receptive field block[J]. *Laser & Optoelectronics Progress*, 2020, 57(2): 021501.
王伟锋, 金杰, 陈景明. 基于感受野的快速小目标检测算法[J]. *激光与光电子学进展*, 2020, 57(2): 021501.
- [10] Liu J M, Yang S, Huang H. Hyperspectral remote sensing image classification based on local reconstruction Fisher analysis[J]. *Chinese Journal of Lasers*, 2020, 47(7): 0710001.
刘嘉敏, 杨松, 黄鸿. 基于局部重构 Fisher 分析的高光谱遥感影像分类[J]. *中国激光*, 2020, 47(7): 0710001.
- [11] Song L, Cheng Y M, Zhao Y Q. Hyper-spectrum classification based on sparse representation model and auto-regressive model [J]. *Acta Optica Sinica*, 2012, 32(3): 0330003.
宋琳, 程咏梅, 赵永强. 基于稀疏表示模型和自回归模型的高光谱分类[J]. *光学学报*, 2012, 32(3): 0330003.
- [12] Zhang Y S, Wu L, Ren H Z, et al. Mapping water quality parameters in urban rivers from hyperspectral images using a new self-adapting selection of multiple artificial neural networks[J]. *Remote Sensing*, 2020, 12(2): 336.
- [13] Khan F U, Guarnizo G, Martín-Mateos P. Direct hyperspectral dual-comb gas imaging in the mid-infrared[J]. *Optics Letters*, 2020, 45(19): 5335-5338.
- [14] Yokoya N, Chan J C W, Segl K. Potential of resolution-enhanced hyperspectral data for mineral mapping using simulated EnMAP and sentinel-2 images[J]. *Remote Sensing*, 2016, 8(3): 172.
- [15] Malkamäki T, Kaasalainen S, Iinca J. Portable hyperspectral lidar utilizing 5 GHz multichannel full waveform digitization[J]. *Optics Express*, 2019, 27(8): A468-A480.
- [16] Liu L X, He D, Li M Z, et al. Identification of Xinjiang jujube varieties based on hyperspectral technique and machine learning[J]. *Chinese Journal of Lasers*, 2020, 47(11): 1111002.
刘立新, 何迪, 李梦珠, 等. 基于高光谱技术与机器学习的新疆红枣品种鉴别[J]. *中国激光*, 2020, 47(11): 1111002.
- [17] Medus L D, Saban M, Francés-Villora J V, et al. Hyperspectral image classification using CNN: application to industrial food packaging [J]. *Food Control*, 2021, 125: 107962.
- [18] Liu P, Zeng Z G, Wang J. Multistability analysis of a general class of recurrent neural networks with non-monotonic activation functions and time-varying delays[J]. *Neural Networks*, 2016, 79: 117-127.
- [19] Hornik K. Approximation capabilities of multilayer feedforward networks[J]. *Neural Networks*, 1991, 4(2): 251-257.
- [20] Ramachandran P, Zoph B, Le Q V. Searching for activation functions [EB/OL]. (2017-10-16) [2021-04-10]. <https://arxiv.org/abs/1710.05941>.
- [21] Misra D. Mish: a self regularized non-monotonic neural activation function [EB/OL]. (2019-08-23) [2021-04-10]. <https://arxiv.org/abs/1908.08681>.
- [22] Hu J, Shen L, Sun G. Squeeze-and-excitation networks [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: 7132-7141.
- [23] Zhang X D, Wang T J, Zhu S J, et al. Hyperspectral image classification based on dilated convolutional attention neural network [J]. *Acta Optica Sinica*, 2021, 41(3): 0310001.
张祥东, 王腾军, 朱劭俊, 等. 基于扩张卷积注意力神经网络的高光谱图像分类[J]. *光学学报*, 2021, 41(3): 0310001.
- [24] Zhou X Y, Wang D Q, Krähenbühl P. Objects as points[EB/OL]. (2019-05-16) [2021-04-10]. <https://arxiv.org/abs/1904.07850>.
- [25] Redmon J, Farhadi A. Yolov3: an incremental improvement[EB/OL]. (2018-05-08) [2021-04-10]. <https://arxiv.org/abs/1804.02767>.
- [26] 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 (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.
- [27] Dai J F, Qi H Z, Xiong Y W, et al. Deformable convolutional networks[C]//2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 764-773.
- [28] Law H, Deng J. CornerNet: detecting objects as paired keypoints [M] // Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11218: 765-781.
- [29] Chao Y W, Vijayanarasimhan S, Seybold B, et al. Rethinking the faster R-CNN architecture for temporal action localization [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: 1130-1139.
- [30] Yang L, Peng H W, Zhang D W, et al. Revisiting anchor mechanisms for temporal action localization [J]. IEEE Transactions on Image Processing, 2020, 29: 8535-8548.
- [31] Lin T Y, Maire M, Belongie S, et al. Microsoft COCO: common objects in context [M] // Fleet D, Pajdla T, Schiele B, et al. Computer vision-ECCV 2014. Lecture notes in computer science. Cham: Springer, 2014, 8693: 740-755.
- [32] Tian Z, Shen C H, Chen H, et al. FCOS: fully convolutional one-stage object detection [C] // 2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 9626-9635.