基于双重感知注意力与特征度量协同的小样本遥感图像目标检测方法

周建军,陈少波*
(中南民族大学 电子信息工程学院,湖北 武汉 430074)

摘 要 遥感图像目标检测在众多领域应用前景广阔,小样本背景下却面临特征提取不充分、定位精准度欠佳以及分类易出错等难题.针对上述提出的问题,首先,构建双重感知注意力模块,其背景衰减注意力有效抑制背景干扰,空间感知注意力引导网络关注目标定位关键信息,助力区域建议网络生成更好的区域建议框,降低目标遗漏的概率,提升小样本目标定位的性能;同时,引入内省度量学习,在训练中促使模型深度挖掘样本间相似性与差异性特征,强化特征学习与理解,提高分类准确率.最后,设计了一种基于微调的迁移学习方法的小样本目标检测模型,并在遥感数据集NWPU VHR-10上验证.实验结果表明,相较于基准算法,本文算法平均精度均有大幅提升.

关键词 小样本学习;目标检测;遥感图像;度量学习
中图分类号 TP181 **文献标志码** A **文章编号** 1672-4321(2026)05-0640-08
doi:10.20056/j.cnki.ZNMDZK.20260709

Few-shot remote sensing image object detection method based on dual perceptual attention and feature metric collaboration

Zhou Jianjun, Chen Shaobo*
(College of Electronics and Information Engineering, South-Central Minzu University, Wuhan 430074, Hubei China)

Abstract Remote sensing image object detection has broad application prospects in many fields. However, under the background of few-shot, it faces challenges such as insufficient feature extraction, poor positioning accuracy, and prone to classification errors. To address these issues, firstly, a dualperceptual attention module is constructed. Its background attenuation attention effectively suppresses background interference, and the spatial perceptual attention guides the network to focus on the key information for target positioning, helping the Region Proposal Network(RPN) generate better region proposal boxes, reducing the probability of target omission, and improving the performance of few-shot target positioning. At the same time, introspective metric learning is introduced, it prompts the model to deeply explore the similarities and differences between samples during training, strengthening feature learning and understanding, and improving classification accuracy. Finally, a few-shot object detection model based on fine-tuning transfer learning is designed and verified on the remote sensing dataset NWPU VHR-10. Experimental results show that compared with the benchmark algorithm, the proposed algorithm achieves significant improvements in average precision.

Keywords few-shot learning; object detection; remote sensing image; metric learning

遥感图像目标检测是计算机视觉领域的一个重要研究方向,旨在从遥感图像中准确地识别并定位出感兴趣的目标对象,在资源勘探、自然灾害评估、军事目标检测和识别等领域[1]发挥着越来越重要的作用.然而,在实际应用场景里,受限于高昂的数据采集成本、复杂的标注流程以及特定目标自身的稀缺性,以至于有限的数据量难以构建大规模的遥感图像数据集,导致模型难以学习到足够丰富且

收稿日期 2025-07-04
* **通信作者** 陈少波(1980-),男,副教授,博士,研究方向:小样本学习,图像处理,E-mail:chenshaobo1980@qq.com
基金项目 国家自然科学基金资助项目(61201448);中央高校基本科研业务费专项资金资助项目(CZY22012)

具代表性的目标特征.基于上述有限数据样本问题，小样本遥感图像目标检测具有很高的研究价值与广泛的应用前景.

遥感图像目标检测不仅要求完成目标语义分类，更需精准回归目标空间位置，而小样本学习通过挖掘样本先验知识为该任务提供了创新性解决方案，其核心机制在于通过基础类别数据集与新类别数据集的协同训练，借助迁移学习[2-3]、度量学习[4]或元学习[5-7]等策略，将基础类别中习得的特征表征能力迁移至新类别目标检测中.例如，通过构建跨类别共享的特征提取器，结合度量学习构建样本相似度空间，使模型能够基于少量新类别样本快速建立判别边界.Li 等[8]介绍了一种开创性的基于微调的技术，称为 InfRS，旨在使用有限的支持集促进新类别的增量学习，同时保持在已建立的基础类别上的性能，而无需重新访问以前的数据集.Lin 等[9]提出了泛化增强小样本目标检测模型，该模型引入了用于增强多尺度特征表示的跨层融合金字塔注意力网络、用于更准确区域提议的多阶段细化区域提议网络以及用于改进小样本场景下分类性能的广义分类损失.Zhao 等[10]提出一种用于遥感图像的多尺度小样本目标检测方法，该方法建立在 Faster RCNN[11]架构上，用卷积算子构建检测器的整个主干，增强特征提取器的分类能力.在路径聚合模块中学习多尺度特征，通过自下而上的信息流缩短信息传输路径，并利用低级特征的语义信息进行定位，在训练阶段增加了检测器中的形状偏差，进一步提高了模型的鲁棒性.Wang 等[12]基于 Faster RCNN 将微调检测器的最后一层作为关键策略，提出一种针对小样本目标检测的高效方法 TFA（Two-stage Fine-tuning Approach），该方法在相同数据集上准确率比元学习高出 20%，解决了传统元学习方法架构复杂、内存效率低的问题；然而对于基础类别与新类别语义差异大时特征迁移能力不足以及目标分类中噪声存在的影响等问题，仍需进一步改进.

因此，针对上述问题，本文提出一种基于双重感知注意力模块和内省度量学习的小样本遥感图像目标检测算法，改进目标检测对于小样本类别目标的分类和定位能力，通过双重感知注意力模块的指导以及内省度量学习对类别的准确区分，提升模型整体的检测性能.

1　研究方法

本文采用两阶段微调的训练策略，模型的训练过程包括基础训练阶段和小样本微调阶段.在基础训练阶段，用充足的基础类别训练样本来训练双重感知注意力模块，提升模型对目标的辨识度；在小样本微调阶段，制作的小样本数据集中只选取少量有标注的图像，借助从大规模基础类别中学习的感知注意力，指导特征图的生成，同时引入内省度量学习，促进模型深度挖掘样本间相似性与差异性特征，增大类间距离，减少类内距离，提升模型分类的准确性.网络整体结构如图 1 所示.

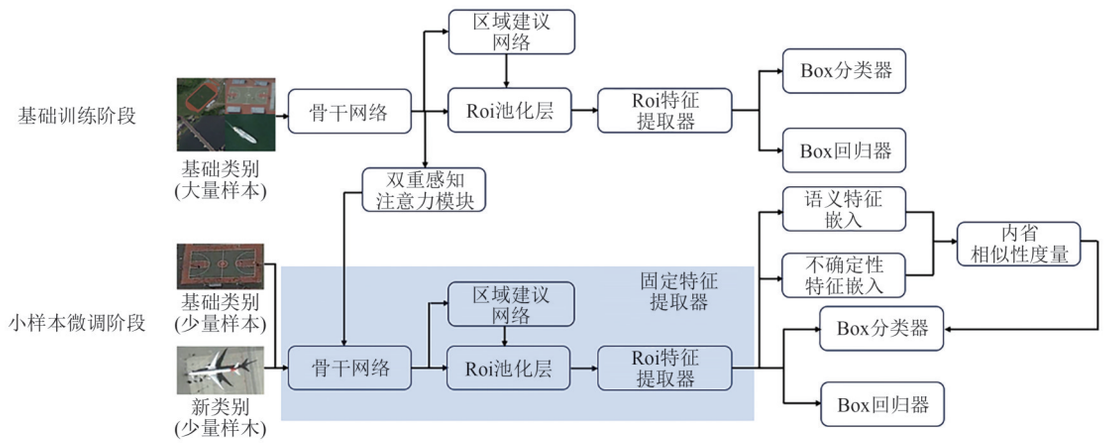


图 1　网络整体结构

Fig. 1　The overall structure of the network

1.1　双重感知注意力模块

在小样本遥感图像目标检测中，由于遥感图像含有丰富的地学纹理区域和高频信息，运用常规的深度学习网络模型在小样本数据集上特征提取能力比较弱，不能获取充足的语义信息，使得对小样本的类别判断和目标定位比较差.本文提出一种双重感知注意力模块，包括背景衰减注意力和图像空间感知注意力，旨在从大量的基础类别数据集中提

取更多与目标定位相关的注意力,运用到微调阶段中,赋予特征点在多个维度上的不同权重,帮助模型更好地检测小样本遥感图像中的目标.背景衰减注意力结构如图2所示.

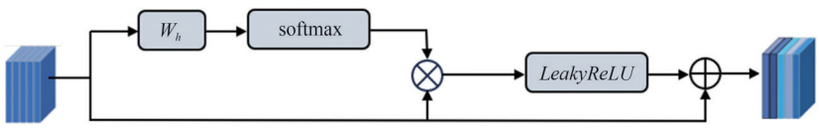


图2 背景衰减注意力结构

Fig. 2 Background attenuation attention structure

该过程可以表示如下:

$$\sigma\left(W_h y_i\right)=\frac{\exp(W_h y_i)}{\sum_{j\in\Omega}\exp(W_h y_i)},\tag{1}$$

其中,y_i表示原有特征图Y上第i点的特征向量,Ω表示空间点的集合,W_h表示线性可学习矩阵.$\sigma\left(W_h y_i\right)$是学习到的空间维度的权重分布,为该模块提供前景背景的判断依据,将计算出的结果与原有特征图进行加权,得到特征增强后的特征图.该过程可以表示如下:

$$Z=Y+\alpha\cdot\mathrm{LeakyReLU}\left(\sigma\left(W_h y_i\right)\cdot Y\right),\tag{2}$$

其中,α是一个超参数常数.图像经过背景衰减模块,能够让特征图的原始信号沿着通道维度与重要性得分对齐,与目标相关的区域可以增强或保持特征;相反,无关区域信号经加权后变得模糊,相当于信号的一种叠加,从而去除背景噪声,实现突出有效目标特征、抑制无效特征的效果.

图像空间感知注意力结构如图3所示.

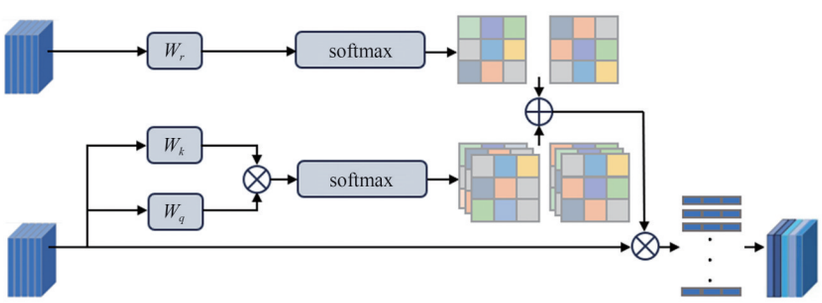


图3 图像空间感知注意力结构

Fig. 3 Image spatially perceptual attention structure

通过两个学习的权重矩阵$W_q, W_k\in\mathbb{R}^{C\times C'}$将输入的特征图处理为嵌入向量,$C'$是原始特征通道数$C$的1/4,计算过程如式(3)所示:

$$Q=W_q X, K=W_k Z.\tag{3}$$

特征图上各部分之间的相似度得分可以如式(4)计算得来:

$$\delta=\sigma\left(\left(Q-\mu_Q\right)^{\mathrm{T}}\left(K-\mu_K\right)\right),\tag{4}$$

其中,μ_Q和μ_K是所有像素上的平均嵌入值,softmax函数σ在空间维度上执行.除了考虑特征图与自身的相关性,还应该基于图像本身.因此,图像空间感知注意力函数表示如下:

$$A=\delta+\beta\cdot W_r Z,\tag{5}$$

其中,β为常系数,$W_r\in\mathbb{R}^{C\times 1}$.此时,$A$的输出具有$H_Q W_Q\times H_S W_S$的形状,代表基于背景衰减注意力处理后特征图($H_Q W_Q$)的每个空间位置的多个注意力图$H_S W_S$.因此,位置感知向量可以由式(6)得到:

$$p_i=\sum_{j\in\Omega}A_{ij}\cdot z_j,\tag{6}$$

其中,Ω表示特征图的所有像素.由式(6)可知,Z的所有向量根据每个位置i自适应加权并聚合为向量$p_i\in\mathbb{R}^{1\times C}$.

1.2 内省度量学习

传统的深度度量学习方法侧重于学习一种具有判别性的嵌入向量来描述图像的语义特征,忽略了每幅图像中因噪声或语义歧义而存在的不确定性.在没有意识到这些不确定性的情况下进行训练,会导致模型在训练过程中过拟合已标注的标签,并在推理过程中产生不令人满意的判断.一个好的相似度模型应该在考虑语义差异的同时意识到不确定性,以更好地处理有歧义的图像,以获得更鲁棒的训练.为了实现这一目标,不仅使用语义嵌入来表示图像,还用伴随的不确定性嵌入来表示图像,分别描述图像的语义特征和模糊性,进一步提出一种内省的相似性度量策略.

1.2.1 内省相似性度量(ISM)

为了促进不确定性感知的相似性度量,首先需要对图像中的不确定性进行建模.与现有的将图像建

模为分布的概率嵌入学习方法不同，使用语义嵌入 s 和不确定性嵌入 u 表示图像即 $Y_{IN}=\{s,u\}$. 语义嵌入 s 描述图像的语义特征，不确定性嵌入 u 描述歧义.

为了比较两个图像 x_1 和 x_2，将语义距离定义为 $\alpha(x_1,x_2)=\|s_1-s_2\|_2$，类似于传统的度量学习方法，但进一步计算相似度的不确定性为 $\beta(x_1,x_2)=\|u_1+u_2\|_2$. 在计算范数之前将嵌入不确定性的向量相加，而不是直接将它们的范数相加.

在判断两幅图像之间的语义相似度时，一种内省的度量方式需要同时考虑语义距离和相似度的不确定性. 当不够确定时，该度量拒绝区分语义上的差异，图 4 展示了提出的不确定感知图像比较.

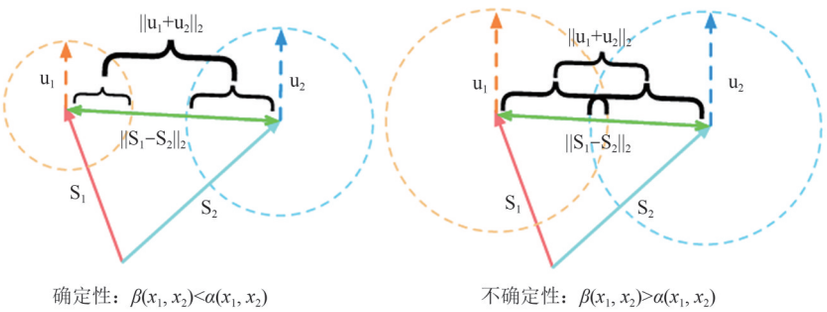


图 4　提出的不确定感知图像比较的示意图

Fig. 4　Illustration of comparison of the proposed uncertain-aware images

在形式上，认为式(7)表示识别 x_1 和 x_2 之间的语义差异是不确定的：

$$\beta(x_1,x_2)+\gamma\geqslant\alpha(x_1,x_2), \tag{7}$$

其中，$\gamma\geqslant 0$ 是内省偏差，表示度量的内省程度. γ 正值表示即使图像表示模型没有提供不确定性，度量仍然是可疑的. 然后定义一个严格的内省相似性度量，如式(8)所示：

$$\tilde{D}_{IN}(x_1,x_2)=\alpha(x_1,x_2)\cdot I\big(\alpha(x_1,x_2)-\beta(x_1,x_2)-\gamma\big), \tag{8}$$

其中，$I(x)$ 是一个指示函数，如果 $x>0$ 输出 1，否则输出 0.

然而，指标函数的使用过于严格，难以在训练过程中进行优化. 相反，比较语义距离和相似度的不确定性来定义两个图像的相对不确定性，如式(9)所示：

$$\tilde{\beta}(x_1,x_2)=\frac{\beta(x_1,x_2)+\gamma}{\alpha(x_1,x_2)}, \tag{9}$$

而且，相对不确定性总是非负的，使用它来软化语义差异，以获得内省相似性度量(Introspective Similarity Metric，ISM)，如式(10)所示：

$$D_{IN}(x_1,x_2)=\alpha(x_1,x_2)\cdot e^{\left(-\frac{1}{\tau}\tilde{\beta}(x_1,x_2)\right)}, \tag{10}$$

其中，$\tau>0$ 是控制弱化程度的超参数.

内省相似性度量同时考虑了两幅图像之间的语义距离和相似度的不确定性，以进行最终的语义差异判断. 由于意识到不确定性，它通常会对两幅图像产生较小的距离. 给定两对具有相同语义距离的图像，内省度量对于相似度不确定性较小的一对图像区分度更好. 此外，当两幅图像的不确定性在很大程度上超过语义距离时，内省度量仅输出一个接近为零的语义距离，避免对网络产生不必要的影响.

1.2.2　内省式深度度量学习(IDML)

将提出的内省相似性度量应用于现有方法，其中 IDML(Introspective Deep Metric Learning)的整体框架如图 5 所示.

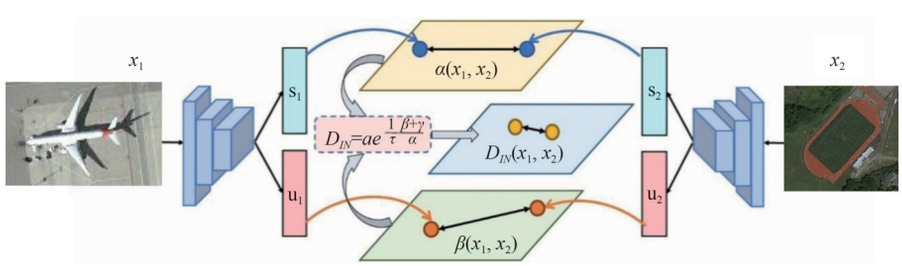


图 5　IDML 框架的一个实例

Fig. 5　An instance of the IDML framework

虽然原始图像中普遍存在数据的不确定性，但很难准确地量化和比较每幅图像的不确定性. 另一方面，数据增强扩展了训练数据，以提高学习到的模型的泛化能力，但许多增强后的图像，如：混合、模

糊和部分遮挡的图像与原始图像相比显示出更大的语义不确定性.生成的图像既具有原始图像的特征,又具有很大程度的语义不确定性.因此,将IDML与Mosaic[13]一起使用,以展示IDML在处理具有不确定性的数据方面的优势.

采用深度神经网络提取原始图像和混合图像的特征嵌入$y_{\mathrm{IN}}=f(x)=\{s,u\}$,其中$s$和$u$分别表示图像的语义嵌入和不确定性嵌入.然后,一对样本的距离或相似性计算遵循提出的内省相似性度量.IDML可以应用于具有不同损失函数和采样策略的深度度量学习方法中.对于训练目标$J(y,L;\ D)$,其中D是欧氏距离,将嵌入y和指标D分别替换为提出的内省嵌入y_{IN}和D_{IN},这样所提出的IDML框架的目标为$J(y_{\mathrm{IN}},L;\ D_{\mathrm{IN}})$.

大多数基于成对的损失计算样本之间的距离,可以直接使用提出的ISM.然而,许多基于代理的损失,如softmax损失,通常计算余弦相似度$C(x_i,p_j)$,而不是图像x_i和类级代表p_j之间的距离.为了适应这一点,将基于相似性版本的ISM表示如下:

$$C_{\mathrm{IN}}\left(x_i,p_j\right)=1-\left(1-C\left(x_i,p_j\right)\right)\cdot \mathrm{e}^{\left(-\frac{1}{\tau}\tilde{\beta}\left(x_i,p_j\right)\right)},\tag{11}$$

其中,$\tilde{\beta}(x_i,p_j)$表示图像与类级代表之间的相对不确定性,它模糊了相似性差异以获得更大的不确定性.

对比损失[14]让模型在嵌入空间中减小同类样本的距离,同时增大异类样本的距离,如式(12)所示:

$$\begin{aligned}J_{\mathrm{con}}(y_{\mathrm{IN}},L;\ D_{\mathrm{IN}})=&\sum_{l_i=l_j}D_{\mathrm{IN}}\left(x_i,x_j\right)+\\&\sum_{l_i\neq l_j}[\delta-D_{\mathrm{IN}}(x_i,x_j)]_+,\end{aligned}\tag{12}$$

其中,δ是一个边际超参数常数.

softmax损失函数是一个非常常用的损失函数,主要用于分类任务.它可以看作是基于代理的度量学习损失,最大化样本与正代理的相似性,最小化样本与其他代理的额外相似性,如式(13)所示:

$$J_s\left(y_{\mathrm{IN}},L;\ C_{\mathrm{IN}}\right)=\frac{1}{N}\sum_{i=1}^{N}\left(-\log\frac{\sum_{l_{p_j}=l_i}\mathrm{e}^{C_{\mathrm{IN}}\left(x_i,p_j\right)}}{\sum_{l_{p_j}\neq l_i}\mathrm{e}^{C_{\mathrm{IN}}\left(x_i,p_j\right)}}\right),\tag{13}$$

其中,l_{p_j}是类级代表p_j的类别,p_j可以通过线性分类器的第j行向量来实现.

因此,通过在相应的softmax损失中引入内省相似性,所提出的IDML框架可以自然地应用于图像分类任务.

1.3 模型训练

在基础训练阶段,利用大量基础类别训练样本对双重感知注意力模块展开训练,该模块涵盖背景衰减注意力与空间感知注意力,此过程旨在让模块充分学习基础类别数据中的特征信息.进入小样本微调阶段,先构建小样本数据集用于模型微调.针对基础类别与新类别的各个类别,仅挑选少量带标注的图像,借助基础训练阶段从大规模基础类别样本中习得的感知注意力,引导微调阶段特征图的生成.同时,引入内省式度量学习方法,驱动模型在训练时深入挖掘样本间的相似特征与差异特征,强化从不同类别样本中学习高效分类能力的过程,从而提升模型在小样本场景下的性能.

2 实验结果与分析

2.1 实验环境

本文实验使用的GPU为NVIDIA 1080Ti,实验在Linux系统下运行,操作系统的型号及版本为Ubuntu 16.04.1,Python版本为3.11.

2.2 实验数据集

本文采用NWPU VHR-10[15]作为实验数据集,它包含650张有目标的图像和150张无目标的背景图像,一共有3651个实例.数据集包含10个用于检测的地理空间对象类:飞机、车辆、船舶、港口、棒球场、篮球场、网球场、地面田径场、储罐和桥梁.各目标尺度变化较为多样,地形分布的情况各不相同.在实验中将10个类别分布在3个新类别和7个基础类别中,并将数据集划分为两个子集,70%的图像被分配用于训练,其余30%被指定用于测试.

2.3 评价指标

在目标检测领域中,评估目标检测模型的性能广泛采用均值平均精确度mAP(mean Average Precision)作为一种综合性的评价指标,它是平均精度AP(Average Precision)的平均值.AP是通过对召回率-准确率曲线下的面积进行积分得到的.召回率是指模型正确检测出的正样本占所有真实正样本的比例,反映了模型对正样本的覆盖能力;准确率则是指正确检测出的正样本占所有被预测为正样本的比例,体现了模型预测的精确性.在不同的召回率水平下,准确率会有波动,AP综合考量了这些变化,全面反映了模型在该类别上的检测性能.召回率Recall和准确率Precision的定义分别为:

$$\mathrm{Recall}=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}},\tag{14}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{15}$$

其中，*TP*为正确预测为正样本的数量，*FP*为错误预测为正样本的数量，*FN*为错误预测为负样本的正样本数量.

AP是Precision-Recall曲线下的面积，即为：

$$\text{AP} = \int_0^1 p(r)\,dr. \tag{16}$$

将所有类别目标的*AP*进行平均，得到*mAP*为：

$$\text{mAP} = \frac{\sum_{i=1}^{K} \text{AP}_i}{K}, \tag{17}$$

其中，*K*为类别个数，AP_i表示第*i*类的精度.

2.4　实验设置

实验以ResNet101[16]为主干网络(backbone)，以Faster RCNN为基础，训练的神经网络模型使用SGD(Stochastic Gradient Descent)作为优化器，初始学习率为0.001，在训练时使用Step策略，在训练的第12个和第18个Epoch时将学习率降低10倍，使得梯度下降的速度减慢.在小样本学习阶段，仅从每个新类别中随机选取K-shot实例进行训练，对NWPU VHR-10数据集设置*K*=1，3，5，10.

2.5　实验结果

所有的实验均在NWPU VHR-10基准测试中进行.为了进行公平和直接的比较，对小样本训练数据集和测试数据集采取相同的预处理，采用AP和mAP进行评价.结果在新类别的测试集上进行评估.本文提出的模型与各类小样本目标检测模型进行对比，结果见表1和表2.

表1展示的是将NWPU VHR-10数据集划分为3份分别作为新类别数据集，以及本文方法在3份新类别数据集上的检测结果.表2展示了本文方法与Meta-YOLO[17]、TFA、FSODM[18]、DeFRCN[19]等方法在NWPU VHR-10数据集上的小样本目标检测的实验结果.实验数据表明，本文方法在绝大多数场景下展现出显著优势.在新类别检测的结果中，当训练样本数量为10时，本文方法在3份不同的新类别划分设置下，mAP值分别达到76%、60%、47%，相较于对比方法均提升显著.即使在单样本测试中，本文方法仍实现了性能的大幅跃升，充分验证了其良好的鲁棒性和泛化能力.

表1　NWPU VHR-10数据集在新类别上的性能

Tab. 1　Performance of the NWPU VHR-10 dataset on the new category

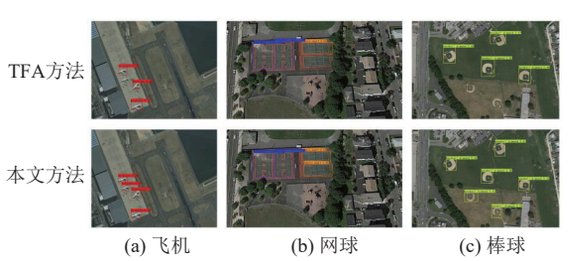
骨干网络	Novel Set 1	1-shot	3-shot	5-shot	10-shot
ResNet101	airplane	0.22	0.43	0.52	0.73
	tennis court	0.46	0.55	0.57	0.69
	baseball diamond	0.20	0.54	0.69	0.86
	mean	0.29	0.51	0.59	0.76
	Novel Set 2	1-shot	3-shot	5-shot	10-shot
	ship	0.12	0.52	0.62	0.73
	storage tank	0.51	0.80	0.79	0.85
	vehicle	0.10	0.14	0.17	0.21
	mean	0.24	0.49	0.53	0.60
	Novel Set 3	1-shot	3-shot	5-shot	10-shot
	basketball court	0.20	0.25	0.37	0.52
	ground track field	0.18	0.53	0.41	0.59
	harbor	0.17	0.02	0.07	0.30
	mean	0.18	0.27	0.28	0.47

表2　NWPU VHR-10数据集各方法的对比实验结果

Tab. 2　Comparative experiment results of various methods on the NWPU VHR-10 dataset

骨干网络	Method	Novel Classes			
		1-shot	3-shot	5-shot	10-shot
ResNet101	Meta-YOLO	0.07	0.13	0.26	0.37
ResNet101	TFA	0.09	0.16	0.26	0.43
DarkNet53	FSODM	0.21	0.32	0.53	0.65
ResNet101	DeFRCN	0.19	0.37	0.57	0.67
ResNet101	本文方法	0.29	0.51	0.59	0.76

从整体性能表现来看，本文方法在不同样本数量下的平均检测精度均优于对比方法，且随着样本量的逐步增加，两者之间的性能差距呈扩大趋势.这一现象表明，本文方法能够更高效地利用新增样本信息，持续优化检测性能.该模型在小样本学习场景下，不仅能够精准检测新类别目标，同时也能有效维持对基础类别目标的检测精度.基线TFA方法和本文方法在新类别上的检测结果见图6.



(a) 飞机　(b) 网球　(c) 棒球

图6　NWPU VHR-10数据集在新类别上的检测结果

Fig. 6　Detection results on the NWPU VHR-10 dataset on the new category

2.6　消融实验

为进一步验证本文提出算法的各个子模块对小样本目标检测结果的影响，针对NWPU VHR-10遥感数据集，选用了ResNet101骨干网络，进行了消融实验，结果如表3所示.

表3充分验证了本文算法各模块的有效性.在

表3 不同模块对实验结果的影响

Tab. 3 The effect of different modules on the experimental results

双重感知注意力模块	内省度量学习方法	mAP			
		1-shot	3-shot	5-shot	10-shot
		0.17	0.40	0.54	0.64
√		0.23	0.45	0.57	0.71
	√	0.22	0.43	0.55	0.70
√	√	0.29	0.51	0.59	0.76

单独引入双重感知注意力模块后,模型在1-shot至10-shot场景下对新类别的mAP指标分别提升至0.23、0.45、0.57、0.71,尤其在低样本量的1-shot场景中,提升幅度高达35.3%.如图7,对比两幅图可见,左图提取的特征多样性较为有限,不同子图之间的差异表现得较小;而右图能够提取出更丰富的语义信息,子图间在明暗、纹理等特征上的区分十分明显,这表明模型对目标区域的特征聚焦能力更为出色.实验前后的对比表明,该模块有效抑制了背景无关特征的干扰,强化了目标区域的特征表达.

当单独加入内省度量学习方法模块时,模型在1-shot至10-shot场景下的新类别mAP分别提升至0.22、0.43、0.55、0.70,并且在图8的可视化结果中,基础模型TFA的新类别与基础类别簇群存在明显重叠,特别是篮球场和网球场之间,而本文模型通过该模块的优化,实现了基础类别与新类别簇群的清晰分离,有效提升了类别间的可区分性.

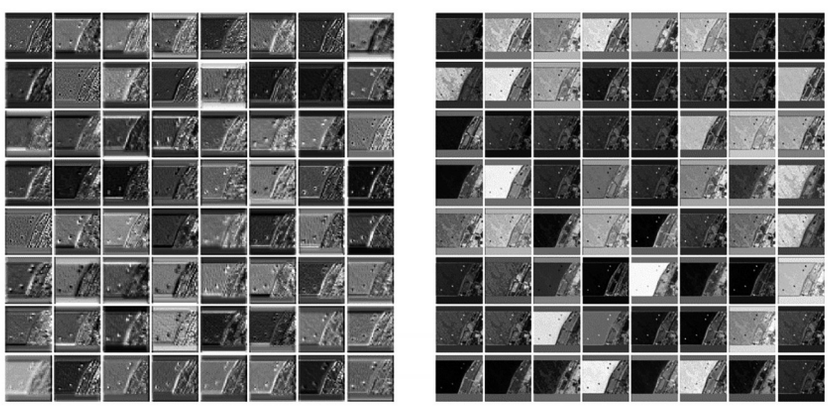


图7 基础模型TFA(左)和本文模型(右)特征提取可视化对比

Fig. 7 Visualization comparison of feature extraction of the basic model TFA (left) and the model proposed in this paper(right)

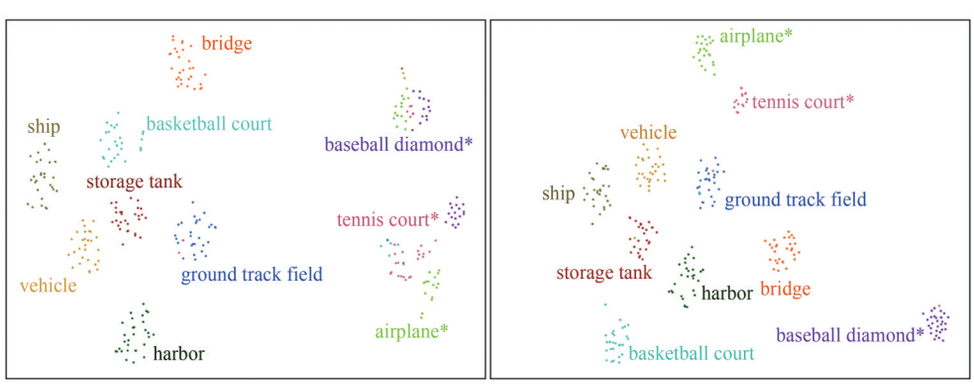


图8 基础模型TFA(左)和本文模型(右)目标类别分类对比

Fig. 8 Comparison of target category classification between the basic model TFA(left) and the model proposed in this paper(right)

消融实验结果充分证明了各模块的协同有效性.此外,图6进一步表明,在小样本条件下,本文算法对遥感图像目标实现了更精准的检测,验证了算法在实际应用中的有效性和鲁棒性.

3 结论

本文从目前小样本遥感图像目标检测方法存在的主要问题出发,提出了一种基于迁移学习的小样本目标检测算法,具体改进如下:

(1)针对特征提取网络对关键特征提取不完全的问题,通过引入双重感知注意力模块,提升网络对目标的敏感度;

(2)针对遥感图像目标准确分类问题,在检测阶段采用内省度量学习来度量图像之间的相似性,提升分类的准确性.

通过实验验证,本文算法可以有效提升小样本遥感图像目标检测的性能,但是,当目标尺度较小

并且分布密集时,算法在这些物体上的检测性能较差.因此,对于算法中存在的问题,下一步将会优化算法,以增强对小尺度目标的捕捉能力.

参 考 文 献

[1] Zhang R, Shao Z, Huang X, et al. Adaptive dense pyramid network for object detection in UAV imagery[J]. Neurocomputing, 2022, 489: 377-389.

[2] Huang Q, Zhang H, Xue M, et al. A survey of deep learning for low-shot object detection[J]. ACM Computing Surveys, 2023, 56(5): 1-37.

[3] Li Z, Wang Y, Zhang N, et al. Deep learning-based object detection techniques for remote sensing images: A survey[J]. Remote Sensing, 2022, 14(10): 2385.

[4] Zhao W, Rao Y, Zhou J, et al. DIML: Deep interpretable metric learning *via* structural matching[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 46(4): 2518-2532.

[5] Li X, Deng J, Fang Y. Few-shot object detection on remote sensing images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 60: 5601614.

[6] Perez-Rua J M, Zhu X, Hospedales T M, et al. Incremental few-shot object detection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE, 2020: 13843-13852.

[7] liu J. Few-shot object detection model based on meta-learning for UAV[C]//Fifth International Conference on Mechatronics and Computer Technology Engineering (MCTE 2022). Chongqing: SPIE, 2022: 138.

[8] Li W, Zhou J, Li X, et al. InfRS: Incremental few-shot object detection in remote sensing images [J]. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62: 1-14.

[9] Lin H, Li N, Yao P, et al. Generalization-enhanced few-shot object detection in remote sensing[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2025, 35(6): 5445-5460.

[10] Zhao Z, Tang P, Zhao L, et al. Few-shot object detection of remote sensing images *via* two-stage fine-tuning [J]. IEEE Geoscience and Remote Sensing Letters, 2021, 19: 8021805.

[11] 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 and Machine Intelligence, 2016, 39(6): 1137-1149.

[12] Wang X, Huang T E, Darrell T, et al. Frustratingly simple few-shot object detection [J]. arXiv preprint arXiv:2020,2003.06957.

[13] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: Optimal speed and accuracy of object detection [J]. arXiv preprint arXiv: 2020,2004.10934.

[14] Hu J, Lu J, Tan Y P. Discriminative deep metric learning for face verification in the wild[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2014: 1875-1882.

[15] Cheng G, Han J, Zhou P, et al. Multi-class geospatial object detection and geographic image classification based on collection of part detectors [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2014, 98: 119-132.

[16] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016: 770-778.

[17] Liu B, Wang Z, Feng Z, et al. Meta-YOLO: Fast few-shot object detection via meta-learning[C]//IEEE. 2021 IEEE International Conference on Multimedia and Expo (ICME). Shenzhen: IEEE, 2021: 1-6.

[18] Chen K, Yang Z, Wang J, et al. FSODM: Few-shot object detection via discriminative mining[C]//IEEE. 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal: IEEE Computer Society, 2021: 9063-9072.

[19] Chen Y, Dai J, Liang X, et al. DeFRCN: Deformable few-shot R-CNN with iterative refinement and attentive training [C]//IEEE. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans: IEEE Computer Society, 2022: 13653-13662.

(责编 曹东,校对 雷建云)