[文章编号] 1003－4684(2025)02-0042-07

一种改进的遥感图像多目标识别算法

黎 林，姚 颖，郭鑫轶，叶云龙
（湖北工业大学计算机学院，湖北 武汉 430068）

[摘 要] 由于遥感图像背景复杂、目标尺度差异大，易导致多目标识别效果不佳，针对此问题提出了一种改进的遥感图像多目标识别算法。以 Faster R-CNN 网络为基础，采用 EfficientNet 作为主干网络以提取更完整的图像特征；然后引入改进后的特征金字塔结构，获得更多上下文信息；其次，统计数据集中所有目标的尺度信息，重新设置 RPN 网络锚框；接着，使用焦点损失函数解决正负样本不平衡的问题；最后利用 ROI Align 区域聚集解决区域不匹配问题。进行消融实验验证了该算法的可行性与有效性，实验结果表明该算法优于原始 Faster R-CNN 算法，多类平均准确率 mAP 达到了 93.11%，提升了 10.59%。

[关键词] 遥感图像；Faster R-CNN；多目标检测；特征金字塔；损失函数
[中图分类号] TP391 [文献标识码] A

随着高空间分辨率遥感卫星的普及应用，遥感图像逐渐呈现出大数据的特点，仅依靠传统的遥感数据处理和信息提取技术难以满足遥感应用的需求[1]。与此同时，深度学习模型在自然图像目标检测任务中取得了巨大成功。目前基于深度学习的目标检测算法主要分为两阶段（Two-Stages）目标检测和单阶段（One-Stages）目标检测。两阶段目标检测算法通常先生成样本的候选区，再通过卷积神经网络进行样本分类，如 Faster R-CNN[2]、FPN[3] 等算法。两阶段目标检测算法准确度较高，但将检测过程一分为二导致耗时较高。与之相反，单阶段目标检测算法直接在网络中提取特征来预测物体分类和位置，在检测速度上占据优势，如 SSD[4]、YOLO[5-8] 系列算法，但在检测精度上稍有逊色。在这个过程中，人们逐渐发现使用单一尺度的特征图不能有效提取不同尺度目标的有效特征依赖信息，为解决这个问题提出了特征金字塔结构。SSD 算法使用不同尺度的特征图同时进行预测，但是没有考虑其他层次特征图的有效信息；FPN 则是十分经典的特征金字塔结构，通过将不同层次的特征图自上而下地横向串联起来，使不同层次的特征融合起来，提高了不同尺度目标的检测精度，但由于是相邻层特征融合容易导致特征层级不平衡。

上述基于深度学习的目标检测算法均在自然图像目标检测领域取得了优异的成绩，也为后续与遥感领域目标检测任务相结合奠定了基础。王俊强[9] 等改进了 SSD 算法，使用密集连接网络替换原有的 VGGNet 并构建特征金字塔进行遥感影像小目标检测，但耗时较多。董志鹏等[10] 通过目标的尺度特征改进 Faster R-CNN，提升了遥感图像目标检测的性能，但精度依旧较低。周慧等[11] 使用特征金字塔和重建图像提高舰船细粒度检测精度，但漏检率依旧较高。邓睿哲等[12] 采用对称式形变特征金字塔结构对遥感影像进行目标检测，但在目标密集区域依旧效果一般。高琛等[13] 使用密集连接特征金字塔网络提取遥感影像崩滑体，但大量使用空洞卷积导致局部信息缺失。上述方法都没有充分考虑到高分辨率遥感图像的特点。不同于普通自然光学图像，遥感图像在目标检测过程中通常会出现背景十分复杂等问题，导致很难准确完整地提取出高分辨率遥感图像目标特征，仅融合相邻层特征的特征金字塔结构亦导致特征层级的不平衡问题，进而影响目标检测与识别的最终结果。

针对上述问题，本文提出了一种改进的遥感图像多目标识别算法，以 Faster R-CNN 为基础，首先将主干网络替换为 EfficientNet-B0，期望能完整地提取图像特征；然后引入特征金字塔网络并改进，对所提取特征进行双向多尺度融合，旨在解决目标尺度差异大的问题；其次，将 RPN 网络初始锚框设置为 4 种尺度和 3 种比例，旨在提供更加准确的候选框；再者，将损失函数改为焦点损失函数，避免正负样本失衡；最后，将池化层由 ROI Pooling 替换为

[收稿日期] 2022－11－15
[第一作者] 黎 林（1997－），男，湖北黄冈人，湖北工业大学硕士研究生，研究方向为计算机视觉。

ROI Align，提高目标检测精度。通过实验结果对比验证了本文方法的有效性。

1　Faster R-CNN 网络

Faster R-CNN 是由 Ross Girshick 团队提出的一种端到端的目标检测模型。Faster R-CNN 通过 RPN(Region Proposal Network，RPN)区域候选网络代替选择性搜索 SS(Selective Search，SS)算法提取候选区域，能够极大地提升检测框的生成速度而且精度较高，达到了接近实时目标检测的性能。

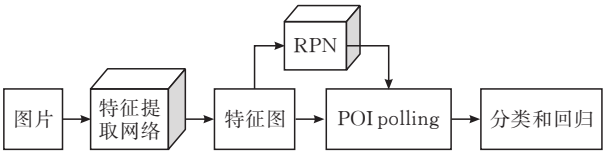


图 1　Faster R-CNN 整体网络结构

图 1 是 Faster R-CNN 的整体网络结构图，从图中可以看出，网络主要由三个部分组成，分别是特征提取主干网络、RPN 网络和目标检测网络。特征提取主干网络主要用于提取输入图像的卷积特征图信息。Faster R-CNN 的主干网络使用的是 VGG-16 网络[14]，VGG-16 网络包括 13 个卷积层和 3 个全连接层，堆叠使用多个 3×3 的卷积核来提取特征。RPN 区域候选网络代替 SS 算法，主要用于生成候选区域。RPN 网络是由全卷积网络组成的，与第三部分的目标检测网络共享卷积层，通过预先设计的 9 种锚框简化生成各种大小的候选区域，并且预测每个候选框中是否含有目标以及利用边界框回归修正锚框的位置，为目标检测网络提供比较精准的候选区域。目标检测网络包括 ROI Pooling 层和最终的分类回归层。ROI Pooling 层将候选区域映射到对应的特征图上，然后把特征图上对应位置范围分类划分出数个相等的区块，最后进行最大池化操作。之后将得到固定大小的特征送入用于分类和边界框回归的全连接层中，检测出目标的类别和边界框。

2　本文方法

Faster R-CNN 目标检测算法设计之初主要针对自然图像，为了使其能够良好地学习与耦合高分辨率遥感图像目标特征信息，本文对 Faster R-CNN 网络模型进行重构，将主干网络替换为 EfficientNet-B0 网络，引入改进后的特征金字塔，重新设置 RPN 网络锚框，使用焦点损失函数，将 ROI Pooling 层改为 ROI Align 区域聚集，本文网络整体结构如图 2 所示。

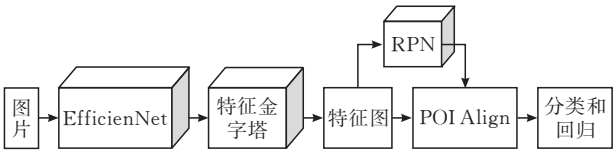


图 2　本文整体结构

2.1　替换主干网络

Faster R-CNN 使用 VGG-16 网络进行特征提取，VGG-16 网络包括 13 个卷积层、3 个全连接层和 5 个池化层，与之前的 AlexNet 和 ZF-Net 不同，不再使用大核卷积，改用堆叠多个 3×3 的卷积核来提取特征。但由于网络深度不足，对拥有复杂背景的遥感图像进行特征提取时，VGG-16 网络很难完整准确地提取出图像的特征，导致后续目标检测精度降低。为了提取更深层次的图像特征，传统的网络模型通常进行缩放，来增加网络的深度或宽度，或输入更大的图像分辨率进行训练和评估，这需要长时间的手动调优，效果也常常不理想。一般来说，增加网络的深度能够得到更加丰富、复杂的特征，但网络的深度过大会面临梯度消失，训练困难的问题；增加网络的宽度能够获得更高细粒度的特征并且也更容易训练，但网络往往很难学习到更深层次的特征；增加输入网络的图像分辨率能够获得更高细粒度的特征模板，但对于非常高的输入分辨率，准确率的增益也会减小，并且大分辨率图像会增加计算量。

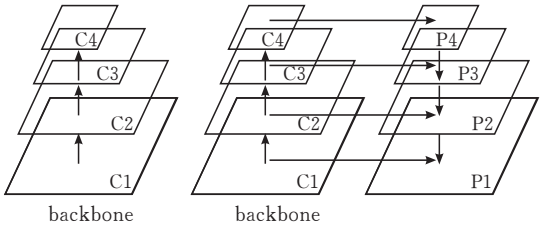
为了解决这个问题，本文对主干网络进行替换，使用 EfficientNet-B0 网络提取图像特征。EfficientNet[15]网络是一种新型的卷积神经网络模型，通过同时对网络模型的深度、宽度以及分辨率进行平衡调整，并且用轻量级的参数优化网络的性能，提升该网络的特征提取能力和目标检测准确率。不同于传统网络，该网络是一种自适应的优化网络结构，参数更少，速度更快，准确率更高。为了使模型最轻量化，本文使用 EfficientNet-B0 网络。EfficientNet-B0 网络结构分为 9 个部分，第 1 部分是一个卷积核大小为 3×3 且步距为 2 的普通卷积层，2～8 部分均为移动翻转瓶颈卷积 MBConv(mobile inverted bottleneck convolution，MBConv)卷积模块，第 9 部分则由一个普通的 1×1 的卷积层、一个平均池化层和一个全连接层组成。该网络主要由 MBConv 卷积模块进行特征提取，MBConv 结构主体引入了倒置残差和残差模块，首先经由 1×1 的普通卷积进行升维操作，然后对特征图进行深度可分离卷积，加入注意力模块获取不同通道的权重，最后再使用 1×1 的卷积对特征图进行降维，输出最终的特征图。

2.2 引入改进后的特征金字塔结构

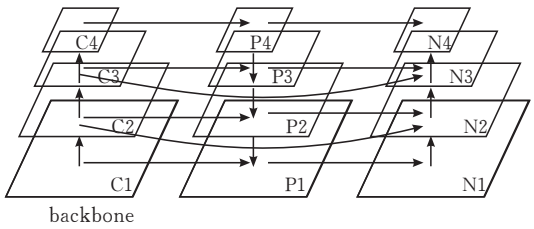
在普通的卷积神经网络中，浅层网络的感受野小，语义信息表征能力弱，但特征图的分辨率高，几何信息的表征能力强；而深层网络与之相反，感受野大，虽然几何细节信息表征能力弱，分辨率低，但是语义信息表征能力强。原 Faster R-CNN 网络仅仅使用主干网络深层提取的单一尺度特征图作为 RPN 网络的输入进行进一步的分类和定位，通过不断下采样获得图像的高层语义信息，如图 3a 所示。这种方法的缺陷十分明显，由于部分目标本身较小，所含的像素信息十分有限且下采样过程中很容易被丢失，导致模型对小目标的检测效果不佳。而特征金字塔结构可以较好的解决该问题，故本文将引入并改进特征金字塔结构以提高模型对小目标的检测精度，改进前后的特征金字塔结构分别如图 3b、c 所示。

特征金字塔结构是一种检测不同尺度目标的网络组件，可使不同目标在相应的尺度下拥有合适的特征表示，通过融合多尺度信息，在不同尺度下对不同大小的目标进行预测。其工作原理是将低分辨率的深层特征与高分辨率的浅层特征相融合，丰富每层网络的特征信息，从而提升目标识别的准确率。普通的特征金字塔结构如图 3b 所示。C1、C2、C3、C4 是主干网络向前传播得到的不同层级的特征图；P4、P3、P2、P1 则是先将高层特征图进行 2 倍上采样，然后以按元素相加方式与不同层级的特征图进行特征融合得到的。由图可见，特征金字塔结构主要分为三个部分，一是主干网络提取特征，二是自上而下传递语义信息，三是将以上两部分的特征图进行融合。但普通的特征金字塔结构往往是相邻两层特征图进行融合，导致特征层级的不平衡问题，会使得当前层的特征更加关注相邻层的语义信息，而非相邻层的依赖关系在目标识别当中往往起着重要的作用，这极大地限制了特征学习和表示，而且最高层特征并未接受来自其他层的特征信息，只包含单尺度的上下文信息，缺乏其他尺度的特征信息。

改进后的特征金字塔结构如图 3c 所示。C1、C2、C3、C4 是主干网络提取的不同层级的特征图；P4、P3、P2、P1 则是通过上采样和低层特征元素相加后生成的，与普通特征金字塔结构相同。改进后的特征金字塔的生成过程是：从最底层的 P1 开始，每个低层级的 P 图通过一个步长为 2 的 3×3 卷积进行池化操作来减小该特征图的分辨率大小，使其与高一层的特征图分辨率保持一致，将得到的特征图与高一层的特征图按通道拼接并且与跳跃连接得到的 C 图进行特征融合，最终得到 N1、N2、N3、N4 构成的双向金字塔，其中 N1 即 P1，因为 P1 已经融合了全局信息所以不再做多余处理，在此过程中每个对应的 C、P、N 特征图的通道数和分辨率一致。与普通的特征金字塔结构只有一条从上到下的融合通道不同，改进后的特征金字塔结构在原有基础上增加了一个自下而上的融合通道，将最底层与前一层的特征图再次重新融合，且用跳跃连接额外增加一个从输入直通输出的通道，极大地丰富了当层特征图的语义信息和几何信息，使得网络的特征融合更加准确，更加关注全局特征信息，不再局限于相邻层融合的影响，使得不同大小的目标都可以在相应的尺度下拥有合适的特征表示。



(a)原Faster R-CNN单一尺度特征 (b)普通特征金字塔



(c)改进的特征金字塔

图 3 特征金字塔结构

2.3 重新设置 RPN 网络锚框

本文使用 NWPU VHR-10 数据集[16-18]，该数据集是由西北工业大学标注的航天遥感目标检测数据集。本文统计了原始 NWPU VHR-10 数据集中的目标信息，共 800 张图像，包含 10 个类别，总计 3775 个实例目标，目标的尺度范围均在 17×30 像素～418×303 像素之间。

重新设置目标感兴趣区域提取网络的初始锚框如图 4 所示。原 RPN 网络使用 3 种尺度(128,256,512)和 3 种比例(1∶1,1∶2,2∶1)，共生成 9 种初始锚框用于生成候选区域，9 种初始锚框生成目标感兴趣区域大小如图 4 右侧矩形框内所示。在 NWPU VHR-10 数据集中，9 种初始锚框只覆盖了 392 个目标，占总目标数的 10.38%，这表明 9 种初始锚框难以耦合该数据集中目标尺寸。本文对 RPN 网络的初始锚框进行改进，设置为 4 种尺度(32,64,128,256)和 3 种比例(1∶1,1∶2,2∶1)共生成 12 种初始锚框，12 种初始锚框生成目标感兴趣区域大小如图 4 左侧矩形框内所示。改进后的 RPN 网络的感兴趣区域可以覆盖 3673 个目标，占

总目标数的 97.29%，几乎所有的目标尺寸都被包含在其中。由统计结果可知，本文设置的 4 种尺度(32，64，128，256)和 3 种比例(1∶1，1∶2，2∶1)的初始锚框能更有效地耦合该数据集的目标尺度范围。因此，本文在对 Faster R-CNN 模型进行重构时，RPN 网络使用 4 种尺度(32，64，128，256)和 3 种比例(1∶1，1∶2，2∶1)生成训练和测试过程中的目标感兴趣区域。

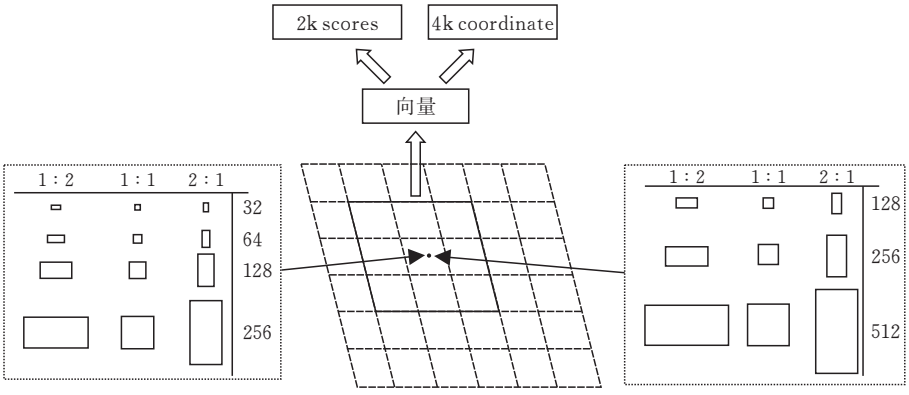


图 4　RPN 网络的初始锚框

RPN 网络的输入特征图为主干网络提取到的特征图，在得到的特征图上使用滑动窗口，每个窗口位置会使用 k 个尺寸和长宽比不同的初始锚框产生感兴趣区域，在本文即使用 4 种尺度(32，64，128，256)和 3 种比例(1∶1，1∶2，2∶1)共计 12 种初始锚框，并且计算出滑动窗口中心点对应原始图像上的中心点，将这 12 种初始锚框映射到原始图像上，即可以获得不同的原始候选区，然后将由滑动窗口产生的向量通过两个全连接层，获得 $2k$ 个分数来区分该区域属于前景还是背景以及包括 $4k$ 个参数的锚框区域向量(xy 坐标以及长宽)进行后续的目标感兴趣区域的回归训练。

2.4　改变损失函数

由于正负样本的不均衡性，很容易出现下列现象：当某单个类别的样本数特别多的时候，训练出来的网络对该类的检测精度往往会比较高；而当某一类训练样本数较少的时候，模型对该类目标的检测精度就会有所下降。通常遥感图像的复杂程度远高于自然图像，在不同类别目标的数量与分布密度上充满不确定性，极易出现正负样本不均衡的问题。绝大多数负样本都是易分类的背景样本，易分类样本过多导致损失函数无法收敛至最优结果。所以为了平衡易分类和难分类样本比例，减小易分类样本在总损失中的占比权重，让模型更加关注难分类样本以提升检测精度，现将原交叉熵损失函数替换为焦点损失函数[19]。

焦点损失函数公式如下：

$$FL(p,y) = -y(1-p)\log(p) - (1-y)p\log(1-p) \quad (1)$$

式中：y 代表真实值，通常为 1 或者 0；p 代表预测概率，取值一般在 0～1 之间；γ 为指数，通过 γ 值控制调制因子 $(1-p)^{\gamma}$ 和 p^{γ} 的大小，从而控制易分类和难分类样本比例。$\gamma=2$ 时，对于正样本预测结果为 0.95，调制因子 $(1-0.95)^2$ 会使损失函数值变得更小，这样就减少了简单样本的影响。焦点损失函数的代价函数如图 5 所示。当 $\gamma=0$ 时，蓝色曲线代表的是交叉熵损失函数，易分类样本损失占总损失的比例较大；由图可以看出，随着 γ 值增大，易分类样本损失占总损失的比例变小，难分类样本损失占总损失的比重变小，使得模型更加专注于难分类样本。

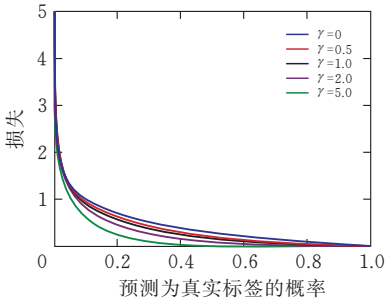


图 5　焦点损失函数的代价函数

2.5　区域聚集

在 Faster R-CNN 中，RPN 网络提取出一系列的候选框后，由于候选区域尺寸不一样，需要将不同尺寸的候选区域经过 ROI Pooling 特征池化后得到相同大小。ROI Pooling 的作用是根据候选框的位置坐标在特征图中将相应区域池化为固定尺寸的特征图，以便进行后续的分类和包围框回归操作。首先将候选框映射到特征图上，然后特征图等分为各个单元，将大小不同的候选区域池化为尺寸相同的特征图，这其中包含了两次量化取整的操作，其分别是 1、将候选区域为浮点数的边界长度通过向下取整变成一个整数；2、将量化后的边界所包含的区域

均匀分成多个相同大小的单元。这两次取整都会使原本映射在特征图上的候选区域会产生偏差，最后降低目标检测的精度，尤其对小目标检测的影响巨大。这种现象也被称为为区域不匹配问题。

ROI Align 区域聚集[20]操作很好地解决了区域不匹配问题。ROI Align 的核心原理是双线性插值算法。ROI Align 具体思路是取消量化操作，首先将候选框映射到特征图上，坐标保持浮点数的形式，然后使用双线性内插的方法获得坐标为浮点数的像素点上的对应图像坐标，然后进行最大池化操作，从而将整个特征聚集过程转化为一个连续的操作。ROI Align 无需量化且保留浮点数的操作更加完整地保留了原图的特征信息，尽可能地减小候选框映射偏差，从而提高整个网络的检测准确率。

3 实验与分析

3.1 实验数据集

本文使用公开的遥感图像目标检测数据集 NWPU VHR-10 来验证所提出方法的有效性。该数据集由西北工业大学于 2014 年发布，是遥感图像目标检测领域经典的数据集。该数据集包括 800 张图像，其中正样本共 650 张，负样本 150 张，图像分辨率大致为 850×900 像素。该数据集共标记了 10 类目标，所标注的实例包括飞机、舰船、油罐、棒球场、网球场、篮球场、田径场、港口、桥梁、车。由于原始 NWPU VHR-10 数据集规模较小，对原始数据集进行图像增广，增广后的总样本数为 2750 张图像，最大限度强化模型的泛化性。

3.2 评价指标

本文采用目标检测中常用的评价指标平均准确率 AP、多类平均准确率 mAP 来验证本文方法的有效性。

交并比 IOU 是指预测框面积和真实框面积的交集和并集的比值，计算公式如式(2)所示，其中 area(C) 表示预测框面积，area(G) 表示真实框面积。

$$\mathrm{IOU}=\frac{\mathrm{area}(C)\cap\mathrm{area}(G)}{\mathrm{area}(C)\cup\mathrm{area}(G)} \tag{2}$$

以预先设置的阈值和计算的交并比 IOU 为基准，可以将图像分为是三个类别：被正确划分为正例的个数 T_P 、被错误划分为正例的个数 F_P 以及被错误划分为负例的个数 T_N 。由此可以得出查准率和召回率，分别如(3)式和(4)式所示：

$$\mathrm{precision}=\frac{T_P}{T_P+F_P} \tag{3}$$

$$\mathrm{recall}=\frac{T_P}{T_P+F_N} \tag{4}$$

根据查准率和召回率可以得到 P-R 曲线，从而得到平均准确率 AP 和多类平均准确率 mAP。AP 和 mAP 计算公式分别如 v(5)和式(6)所示：

$$\mathrm{AP}=\int_0^1 p(r)\,dr \tag{5}$$

$$\mathrm{mAP}=\frac{\sum_{i=1}^{C}AP_i}{C} \tag{6}$$

式(5)和式(6)中，p 代表查准率 precision，r 代表召回率 recall，C 代表类别数。

3.3 实验结果

本文设计的各个模块之间的消融实验结果对比如表 1 所示。由表 1 可以看出，本文提出的每一个改进模块均提升了目标检测的精度，每一个改进模块对提升多类平均准确率 mAP 都有一定的作用，证实了本文的可行性和有效性。原始 Faster R-CNN 算法的特征提取网络为 VGG-16 网络，在 NWPU VHR-10 数据集中的多类平均准确率 mAP 为 83.12%。进行消融实验时，首先仅改变主干网络，在将主干网络替换为 EfficientNet 网络后，其余模块不进行改动，由于主干网络提取到了更丰富的图像特征，mAP 取得 86.87%的显著提高，比使用 VGG-16 网络提升了 3.75%。之后再逐步改进其他模块，在引入改进的特征金字塔后，多类平均准确率 mAP 达到 90.39%，在 EfficientNet 模块基础上又提升了 3.52%，mAP 比原始 Faster R-CNN 算法提高了 7.27%。在上述基础上，再重新设置 RPN 网络锚框，使多类平均准确率 mAP 达到 91.23%，最后引入焦点损失函数并且使用 ROI Align 区域聚集后，多类平均准确率 mAP 最终提升至 93.05%(表 1)，比原始 Faster R-CNN 算法多类平均准确率 mAP 提升了 9.93%。

表 1 各个模块之间的消融实验结果对比

Faster R-CNN	√					
EfficientNet		√	√	√	√	√
改进的特征金字塔			√	√	√	√
重新设置 RPN 网络锚框				√	√	√
焦点损失函数					√	√
ROI Align 区域聚集						√
mAP/%	83.12	86.87	90.39	91.23	92.58	93.05

表 2 为四种算法在 NWPU VHR-10 数据集中的多类平均准确率 mAP 结果对比。根据结果可以看出单阶段目标检测 SSD 算法的 mAP 最低，仅为 82.52%，虽然在检测速度上较快，但在检测精度上略逊一筹，在小尺度的目标检测上精度较差，还比不

上 Faster R-CNN 算法，主要是因为小尺度目标多用较低层级的锚框来训练，而 SSD 算法较低层级的特征非线性程度不够，无法训练到足够的精确度。Faster R-CNN 算法在将主干网络替换为 ResNet50 时，在检测精度上有一定的提高，ResNet50 使用残差模块确保网络性能不会随着网络深度增加而退化，但在目标是舰船、油罐、桥梁、车等小尺度目标时，检测精度依旧不尽人意，AP 分别为 84.23%、86.46%、74.74%和 81.79%，而 FPN 网络则一定程度上解决了小尺度目标检测上的错检和漏检问题，mAP 达到了 87.21%。本文将主干网络替换为 EfficientNet，并引入改进后的特征金字塔结构，且重新设置 RPN 网络锚框、引入焦点损失函数、使用 ROI Align 区域聚集，更好地解决了上述问题，mAP 达到了 93.11%，分别比 SSD 算法、Faster R-CNN（ResNet50）算法和 FPN 算法提升了 10.59%、7.59%和 5.9%。通过上述结果可以验证本文的可行性与有效性。

表 2　4 种算法的 mAP 结果对比

类别	Faster R-CNN (ResNet50)	FPN	SSD	本文算法
飞机	93.46	95.57	95.02	97.87
舰船	84.23	89.94	83.37	92.48
油罐	86.46	87.28	83.31	92.60
棒球场	90.74	93.23	91.45	94.36
网球场	86.02	79.97	85.97	93.41
篮球场	87.77	88.39	83.25	92.57
田径场	89.93	95.12	89.09	92.59
港口	80.01	84.54	81.75	91.78
桥梁	74.74	73.56	70.35	91.21
车辆	81.79	84.51	71.62	92.25
mAP	85.52	87.21	82.52	93.11

本文算法结果可视化如图 6 所示。

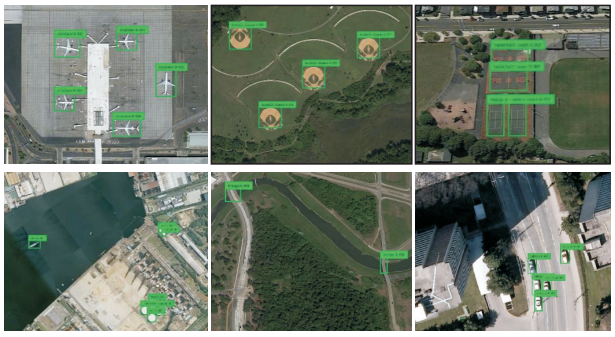


图 6　本文算法结果可视化

4　结束语

目前遥感图像领域的部分目标识别算法不能完整准确地提取出目标特征，并且无法较好地解决目标尺度差异大的问题，导致这部分算法的识别精度较低且鲁棒性较差。针对以上问题，本文提出了一种新的遥感目标检测算法，对原始 Faster R-CNN 网络进行重构，并采用双向深度特征融合的特征金字塔结构使模型更加关注全局特征信息，加强了模型的目标特征提取能力，提升了模型对不同尺寸目标实例的检测精度，最后对实验数据集进行图像增广防止模型在训练时产生过拟合，提高模型的鲁棒性，使其有效地解决传统自然图像目标检测模型移植到遥感图像时的缺陷。实验结果显示，该方法目标检测效果较好，能够在复杂背景下准确识别目标位置，具有较高的检测精度。

[参 考 文 献]

[1] 骆剑承，吴田军，夏列钢. 遥感图谱认知理论与计算[J]. 地球信息科学学报，2016，18(05)：12.

[2] REN S，HE K，GIRSHICK R，et al. Faster R-CNN：Towards real-time object detection with region proposal networks[C]// NIPS，2016，39(06)：1137-1149.

[3] LIN T Y，DOLLÁR P，GIRSHICK R，et al. Feature pyramid networks for object detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition，2017：2117-2125.

[4] LIU W，ANGUELOV D，ERHAN D，et al. Ssd：Single shot multibox detector[C] // European Conference on Computer Vision，Springer，Cham，2016：21-37.

[5] REDMON J，DIVVALA S，GIRSHICK R，et al. You only look once：Unified，real-time object detection[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition，2016：779-788.

[6] REDMON J，FARHADI A . YOLO 9000：Better，Faster，Stronger[C]// IEEE Conference on Computer Vision & Pattern Recognition，IEEE，2017：6517-6525.

[7] REDMON J，FARHADI A. Yolov3：An incremental improvement[EB/OL]. (2018-04-08).[2022-11-15]. https ：// arxiv.org/abs/1804.02767.

[8] GE Z，LIU S，WANG F，et al. Yolox：Exceeding yolo series in 2021[EB/OL]. (2021-08-05).[2022-11-15]. https ：// arxiv.org/abs/2107.08430.

[9] 王俊强，李建胜，周学文，等. 改进的 SSD 算法及其对遥感影像小目标检测性能的分析[J]. 光学学报，2019，39(06)：10.

[10] 董志鹏，王密，李德仁，等. 遥感影像目标的尺度特征卷积神经网络识别法[J]. 测绘学报，2019，48(10)：11.

[11] 周慧，严凤龙，褚娜，等. 基于特征金字塔模型的高分辨率遥感图像船舶目标检测[J]. 大连海事大学学报，

2019，45(04)：8.

[12] 邓睿哲，陈启浩，陈奇，等. 遥感影像船舶检测的特征金字塔网络建模方法[J]. 测绘学报，2020，49(06)：11.

[13] 高琛，冯德俊，胡金林，等. 改进特征金字塔网络的遥感影像崩滑体提取[J]. 测绘科学，2021，46(11)：8.

[14] SIMONYAN K，ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2015-04-10).[2022-11-15]. https：//arxiv.org/abs/ 1409.1556.

[15] TAN M，LE Q. Efficientnet：Rethinking model scaling for convolutional neural networks[C]// International Conference on Machine Learning. PMLR，2019：6105-6114.

[16] 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(dec.)：119-132.

[17] CHENG G，HAN J. A survey on object detection in optical remote sensing images[J]. Isprs Journal of Photogrammetry & Remote Sensing，2016，117：11-28.

[18] GONG C，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.

[19] LIN T Y，GOYAL P，GIRSHICK R，et al. Focal loss for dense object detection[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence，2017，PP(99)：2999-3007.

[20] HE K，GKIOXARI G，DOLLÁR P，et al. Mask r-cnn[C]// Proceedings of the IEEE International Conference on Computer Vision. 2017：2961-2969.

An Improved Algorithm for Multi-object Recognition in Remote Sensing Images

LI Lin，YAO Ying，GUO Xinyi，YE Yunlong

(*School of Computer Science，Hubei Univ. of Tech.，Wuhan* 430068，*China*)

Abstract：Aiming at the problems of complex background and large difference of target scale in the process of remote sensing image object recognition，this paper proposes an improved remote sensing image multi-object recognition algorithm. Based on Fast R-CNN network，this paper uses EfficientNet as the backbone network to extract more complete image features；Then the improved feature pyramid structure is introduced to obtain more context information；Secondly，the scale information of all objects in the data set is counted，and the RPN network anchor frame is reset；Then，the focus loss function is used to solve the problem of imbalance between positive and negative samples；Finally，ROI align region aggregation is used to solve the region mismatch problem. In this paper，ablation experiments are carried out to verify the feasibility and effectiveness of the algorithm. The experimental results show that the algorithm in this paper is superior to the original Fast R-CNN algorithm，and the average multi class accuracy (mAP) reaches 93.11%，increasing by 10.59%.

Keywords：remote sensing image；Faster R-CNN；multi-object detection；characteristic pyramid；loss function

[责任编校：裴 琴]