

# 一种多分辨率图像目标检测算法

杜琳琳, 张瑜, 唐宇, 马静静

(中国人民解放军军事科学院系统工程研究院, 北京 100010)

**摘要:** 为解决航天遥感图像分辨率和目标尺度变化大的挑战, 提出了一种基于多分辨率图像的目标检测算法。改进了自适应特征金字塔和轻量级的分类预测模块, 通过使用注意力机制, 从不同层次的特征图中提取语义信息。引入了一种预测目标尺度的方法, 以分析图像中目标的分布和尺度信息。将算法在 DOTA (Dataset for Object deTecton in Aerial Images) 数据集上进行了实验验证, 在 U-Net (一种基于卷积神经网络的语义分割算法) 和 ResNet-34 (一种深度残差网络) 两种不同的主干网络设置下, 召回率和检测速度均超过了 RPN (Region Proposal Network, 区域提议网络) 算法。提出的多分辨率图像目标检测算法能有效地提高检测精度, 降低计算复杂度。

**关键词:** 深度神经网络; 目标检测; 特征金字塔; 多尺度目标; 多分辨率图像

**中图分类号:** V443+.5 **文献标志码:** A **文章编号:** 2096-4080 (2024) 03-0029-08

## A Multi-Resolution Image Object Detection Algorithm

DU Linlin, ZHANG Yu, TANG Yu, MA Jingjing

(Systems Engineering Institute, Academy of Military Science, PLA, Beijing 100010, China)

**Abstract:** To address the challenges of large image resolution and target scale variations in remote sensing images, this work proposes a target detection algorithm based on multi-resolution images. The adaptive feature pyramid and lightweight classification prediction module are improved. By using attention mechanisms, we extract semantic information from feature maps at different levels, introduce a method for predicting target scale to analyze the distribution and scale information of targets in the images. The algorithm is experimentally validated on the DOTA dataset. With two different backbone network settings, U-Net and ResNet-34, the recall rates and the detection speeds both surpassing the RPN algorithm. The proposed multi-resolution image target detection algorithm effectively improves detection accuracy while reducing computational complexity.

**Key words:** Deep neural networks; Object detection; Feature pyramids; Multi-scale object; Multi-resolution image

### 0 引言

目标检测是计算机视觉领域的一个关键分支, 在对地遥测等应用中发挥着重要作用<sup>[1-4]</sup>。随着深度神经网络的不断发展, 涌现出一系列基于深度神经网络的方法, 相较于传统方法, 它们在提高目标检测准确性方面取得了显著的进步<sup>[5-9]</sup>。譬如,

基于局部区域的卷积神经网络 (CNN) 方法能够更精确地捕捉图像中的目标信息, 从而提高检测的准确性。Fast R-CNN<sup>[10]</sup>和 Faster R-CNN<sup>[11]</sup>等算法引入更高效的区域提取和候选框生成方法, 在保持高准确性的同时提高了检测速度。而 SSD (Single Shot MultiBox Detector)<sup>[12]</sup>和 YOLO (You Only Look Once)<sup>[13-14]</sup>则通过多尺度特征图和巧妙的

收稿日期: 2023-10-25; 修订日期: 2024-03-20

作者简介: 杜琳琳 (1981—), 女, 博士, 副研究员, 主要研究方向为体系架构设计

通信作者简介: 张瑜 (1987—), 男, 博士, 副研究员, 主要研究方向为体系架构设计

检测头部设计,极大地提升了目标检测速度,实现了几近实时的目标检测能力。基于深度神经网络的目标检测方法在航天应用领域对地遥测遥感、气象观测、海洋监测等任务中,具有重要的价值。

现有目标检测算法在处理不同分辨率图像时,绝大多数基于深度卷积神经网络的目标检测算法仅适用于低分辨率图像,低分辨率图像中卷积层的计算成本相对较低<sup>[15-17]</sup>。然而,直接将现今最尖端的深度卷积神经网络目标检测算法应用于更高分辨率的图像,必然需要投入大量的计算资源和时间,这是不可接受的。在分辨率图像领域中,特别是对地遥感图像,目标检测和识别扮演着至关重要的角色。遥感图像数据集与普通目标检测数据集有着明显的区别。由于遥感图像的分辨率和目标尺度之间的差异极大,这就需要目标检测算法能够应对各种尺度的目标。

以遥感图像目标检测为背景,着眼于深度神经网络的多分辨率和多尺度目标检测算法,旨在提升深度神经网络在这种复杂环境下的检测速度

和精度。有助于应对遥感图像处理中的各种挑战,从而提高基于神经网络目标检测的实际应用价值和可靠性。

## 1 基于自适应特征融合的多尺度目标检测方法

### 1.1 自适应特征金字塔模块

自适应特征金字塔模块 (Adaptive Feature Pyramid Module, AFPM) 包含两个步骤,即特征图融合和基于通道的自适应特征提炼。

多尺度目标检测方法一般采取的是自上而下的特征融合结构,递进式将深层特征图中的语义信息传递至底层。然而,这种信息传递是单向的,无法将浅层特征图的语义信息传递至深层特征图。因此,研究了一种自适应特征融合的方法,可从各个尺度的特征图中获取不同尺度特征图所需的补充语义信息,从而增强了原始主干网络各个尺度特征图的语义内容,如图1所示。

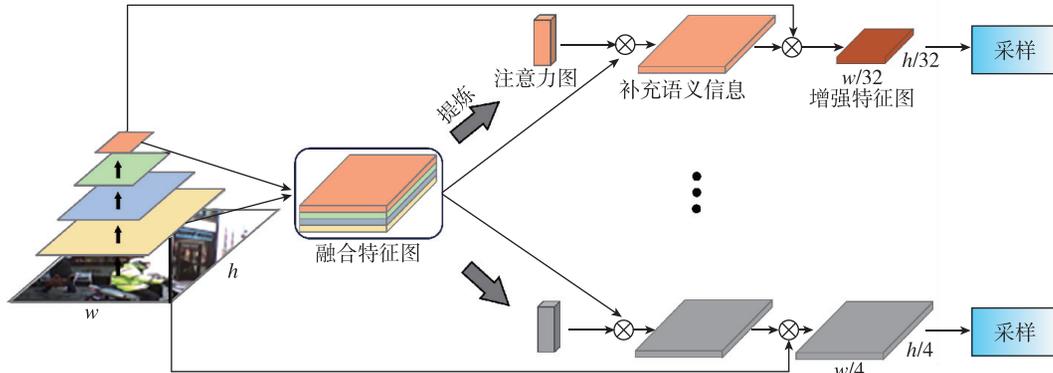


图1 自适应特征图融合模块

Fig. 1 Adaptive feature map fusion module

为有效融合多个尺度的特征图,需对多层特征图进行归一化,使其具有相同的分辨率。采用尺度归一化操作,用于重新缩放具有不同下采样比例的金字塔特征图,并将尺度归一化的多个输出特征图连接在一起。假设输入要素图的形状为  $(C \cdot r^2) \times H \times W$ , 其中  $C$  是通道数,  $r$  是采样因子,  $H$  是张量高,  $W$  是宽。尺度归一化操作是把  $r^2$  通道上相同空间的原色进行周期性重排操作,其数学公式为

$$FM_{C,y,x}^{HR} = FM_{\lfloor \frac{C}{r^2} \rfloor, y + \lfloor \frac{\text{mod}(C, r^2)}{r} \rfloor, x + \text{mod}(\text{mod}(C, r^2), r)}^{LR} \quad (1)$$

$FM^{HR}$  是高分辨率特征图,  $FM^{LR}$  是低分辨率特征图。与上采样操作相比,输出特征图仅为原

始通道的  $\frac{1}{r^2}$ 。如果上采样才做,连接后融合特征图的通道数将急剧增加,从而影响神经网络的计算速度。为了融合超分辨率特征图,选择将尺度归一化的多个输出特征图连接在一起。

在得到初步融合的特征图后,对于每个不同尺度的特征图进行下列操作。首先将融合的特征图  $ffm$  输入到全局平均池化层,其中  $Z_c^l$  表示该操作输出的特征图张量。然后,每个层次的注意力特征图  $AM_c^l$  由两个全连接层操作产生。最后,将  $AM_c^l$  与  $ffm$  之间相乘来生成补充语义信息特征图,  $fr_c^l$  表示最后得到的增强特征图。

$$\mathbf{Z}_c^l = \text{GAP}(f_{fm}) \quad (2)$$

$$AM_c^l = \sigma[W_2^l \delta(W_1^l \mathbf{Z}_c^l)] \quad (3)$$

$$fr_c^l = AM_c^l \otimes f_{fm} \quad (4)$$

其中,  $l$  代表特征金字塔的层数编号,  $W_1^l$  和  $W_2^l$  是用卷积核大小为  $1 \times 1$  的卷积层实现的全连接层, 并使用了批量归一化操作。  $\delta$  代表 ReLu 激活函数,  $\sigma$  是 sigmoid 激活函数,  $\otimes$  是指特征图的按位相乘运算。最后, 每一层的补充语义信息特征图会经过若干层卷积核大小为  $3 \times 3$ , 步长为 2 的卷积层, 下采样到原始第 1 层特征图的尺度大小。

## 1.2 轻量级分类预测模块

分类预测模块通常用于确定每个锚点框中是否存在目标以及目标的所属类别, 并为每个锚点框生成分类概率分布。为了提高检测的准确性, 增加模型的感知能力, 一阶段检测算法使用多个不同尺度和纵横比的锚点框, 以便更好地检测各种不同尺寸和形状的目标。一般来说, 分类预测模块在每一层的特征金字塔输出之后, 是一个卷积核大小为  $k \times k$ 、输出通道数量为  $C \times A$  的卷积层。  $A$  表示预先定义的锚点框类型数量,  $k$  表示卷积核大小。随着参数数量的急剧增加, 特别是在类别数量很大的情况下, 每个类别和每个锚点框都需要一组参数来进行分类概率的预测, 导致模型变得庞大, 难以训练和部署。

为了克服该问题, 提出了一种轻量级的分类预测模块, 见图 2。参数数量和计算复杂度在锚框数量增加时, 能平稳增加而不是急剧增加, 从而使目标检测算法能够在处理大量锚点框时仍保持性能稳定, 并同时保持输入和输出通道之间的密集连接, 以此提高网络的表达能力。将原始卷积核的卷积层划分为两组, 分别是  $1 \times 3$  和  $3 \times 1$ 。

$$\mathbf{Y} = \mathbf{W}^2 \mathbf{P} \mathbf{W}^1 \mathbf{X} \quad (5)$$

其中,  $\mathbf{Y}$  表示输出特征图,  $\mathbf{X} = [x_1^C, x_2^C, \dots, x_i^C]$ ,  $x_i^C \in \mathbf{R}^{\frac{C_{in}}{classes} \times H \times W}$  是指输入的特征图,  $C_{in}$  是输入的特征图的通道数,  $classes$  是分类的数量。而各类别的概率输出表示为  $\mathbf{Y} \in \mathbf{R}^{A \times classes \times H \times W}$ 。  $\mathbf{P}$  代表置换矩阵。  $\mathbf{W}^1 \in \mathbf{R}^{(A \times classes) \times \frac{C_{in}}{classes} \times 1 \times 3}$  和  $\mathbf{W}^2 \in \mathbf{R}^{(A \times classes) \times \frac{C_{in}}{classes} \times 3 \times 1}$  是两个结构化稀疏的卷积核, 具体定义如下

$$\mathbf{W}^i = \begin{bmatrix} \mathbf{W}_1^i & 0 & 0 & 0 \\ 0 & \mathbf{W}_2^i & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \mathbf{W}_{classes}^i \end{bmatrix} \quad (6)$$

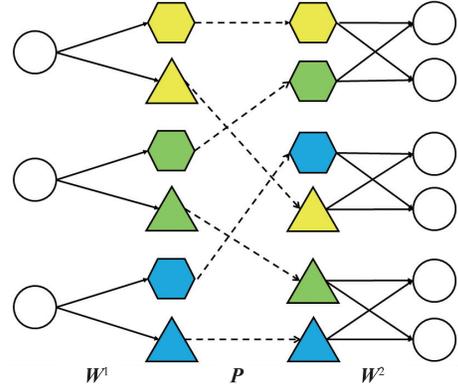


图 2 轻量级分类预测模块

Fig. 2 Lightweight classification prediction module

## 2 针对多分辨率图像的目标检测算法

### 2.1 多分辨率图像尺度预测神经网络

多尺度预测神经网络为多分辨率图像上的每个像素位置提供关于前景目标存在的概率估计和目标的尺度信息, 即为每个像素位置预测前景目标存在的概率值和目标的大致尺度范围。

如图 3 (a) 所示, 首先, 不同尺度的目标像素大小根据需要划分为多个区间, 把每个像素点的目标尺度预测问题转换为对每个像素点的目标尺度分类问题, 预测目标尺度大小属于各个尺度区间范围的概率值, 通过使用标注框的长宽得到目标的尺度大小, 然后再分配不同的尺度类别标签。如划分为  $[0, 16)$ ,  $[16, 64)$ ,  $[64, 128)$ ,  $[128, 256)$ ,  $[256, +\infty)$  这 5 个区间。对于图像  $i$  中的第  $j$  个目标, 通过使用标注框的长宽得到目标的尺度大小, 然后根据该大小分配不同的尺度类别标签, 具体定义如下。其中,  $obj$  表示图像目标。

$$scale\_Label_i(j) = \begin{cases} 0 & 0 \leq obj_h^j \times obj_w^j < 16 \\ 1 & 16 \leq obj_h^j \times obj_w^j < 64 \\ 2 & 64 \leq obj_h^j \times obj_w^j < 128 \\ 3 & 128 \leq obj_h^j \times obj_w^j < 256 \\ 4 & 256 \leq obj_h^j \times obj_w^j \end{cases} \quad (7)$$

考虑到下采样和上采样的需求, 以及如何在 不丢失关键信息的情况下降低计算复杂度, 如图 3 (b) 所示, 对下采样模块采用基本模块设计, 用于逐渐减小特征图的空间尺寸, 并提取高级语义特征。图 3 (b) 中包括最大值池化层和两层卷积层, 其中卷积核的大小为  $3 \times 3$ 。如图 3 (c) 所示, 上采样模块中的基本模块设计则包括双线性插值

层和两层卷积层，用于将下采样阶段输出的特征图的空间尺寸增加，以便进行更精细的特征提取。整个网络包含4个下采样阶段，每个阶段将图像的分辨率降低，最终变为原始图像的1/32。整个网络包含两个上采样阶段，每个阶段将下采样阶段输出的特征图进行上采样，以恢复部分空间细节。因此，最终得到的特征图的大小是原始图像的1/4。这有助于保留大部分目标的信息特征，同时降低了计算复杂度和内存占用。另外，对相同尺度特征图的两个阶段加入短连接，允许将下采样阶段提取的特征信息直接反馈到网络较深的输出，这有助于提高小目标的尺度预测能力，并确保梯度能够有效传播到浅层网络，从而促进网络的训练。

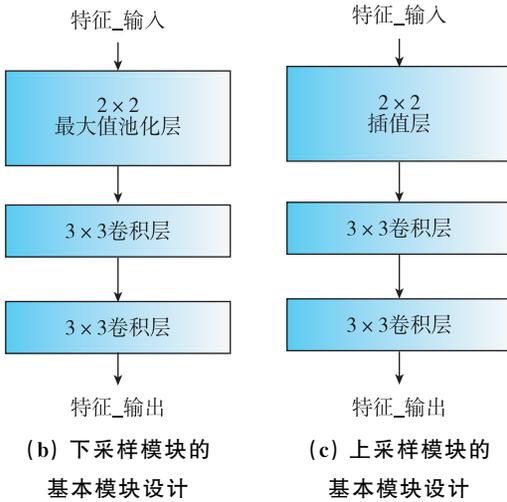
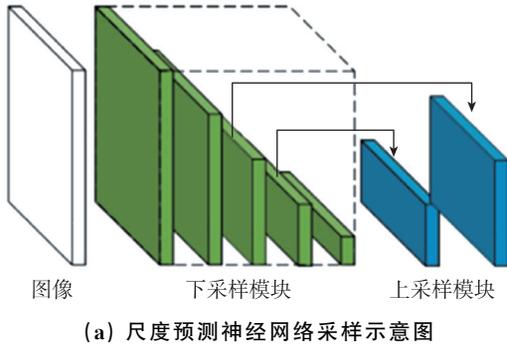


图3 尺度预测神经网络

Fig. 3 Scale prediction neural network

以5个尺度区间划分为例，初步的尺度预测神经网络最终的输出张量包含6个通道，其中第一个通道代表属于前景的分类概率，剩下5个通道代表各个尺度大小的分类概率。

关于尺度预测神经网络的训练，主要介绍损失函数的设计。对于前景/背景分类任务，采用与

YOLOv3<sup>[18]</sup>一样的方法，每个位置使用一位表示objectness代表存在目标的概率，对其使用sigmoid激活函数映射到[0, 1]区间。objectness的损失函数（标准的二值化交叉熵损失函数）可以用公式表示为

$$Loss_{obj} = \frac{1}{N} \sum_{x,y} I_{(x,y)} \log(p_{x,y}^{obj}) + (1 - I_{(x,y)}) \log(1 - p_{x,y}^{obj}) \quad (8)$$

其中，

$$p_{x,y}^{obj} = \text{sigmoid}(s_{x,y}^F)$$

$$s_{x,y}^F = F(\text{img}, \theta)$$

$I_{(x,y)}$  是指示函数，指代在当前位置是否属于目标的检测框内。 $F$  代表尺度预测神经网络， $\theta$  代表该网络的可学习参数。 $p_{x,y}^{obj}$  是特征图上处于位置 $(x,y)$  存在目标的概率值，由预测值 $s_{x,y}^F$  经过激活函数得到。

对于尺度类别预测任务，首先根据公式(7)得到图像上每个位置的尺度类别标签，然后通过公式(8)将离散特征转换为二维特征的独热编码，最后通过交叉熵损失函数得到该部分的损失值，计算出最终的尺度预测网络损失函数。

## 2.2 基于最远点采样的候选检测区域生成算法

在得到图像的尺度预测结果之后，需要根据预测结果得到候选检测区域。图4是尺度预测结果的可视化图，结果以彩色掩码区域的形式呈现，每种颜色代表一个尺度范围，掩码区域的强度表示预测置信度。对于同一目标实例，每个实例在一定区域范围内都会得到同样的尺度结果，对于每个目标实例，其他相邻点的预测结果被抑制，



图4 尺度预测结果

Fig. 4 Scale prediction result

只需要根据每个实例的预测结果生成候选检测区域即可。因此，关键在于如何选择中心点和设置阈值，根据该点的尺度预测结果生成固定大小的

检测框。

根据上述分析，提出了基于最远点采样的候选检测区域生成算法 (FPS\_grouping)，见图 5。

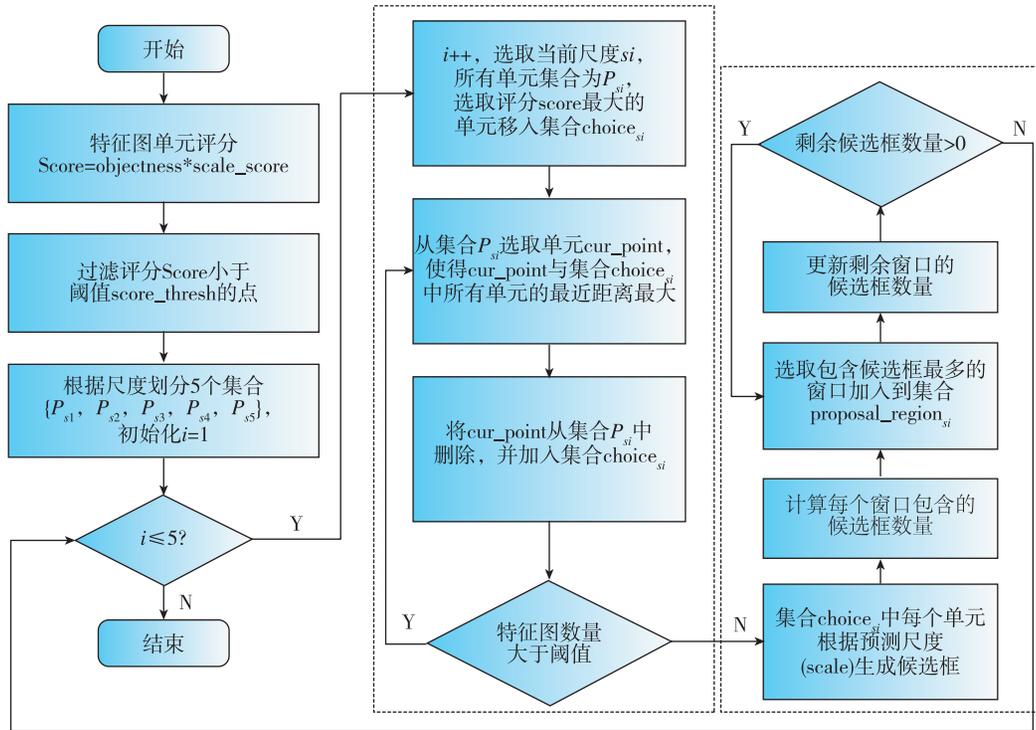


图 5 候选检测区域生成算法

Fig. 5 Candidate detection region generation algorithm

如图 6 所示，以蓝色代表的预测尺度为例，经过检测之后可以得到红色圆点代表的目标检测框中心采样点，而红色虚线框代表对应的目标检测框。其中绿色虚线框代表不同的滑动窗口，通过最后候选检测区域生成步骤，最终筛选得到的部分绿色滑动窗口即候选检测区域的集合。

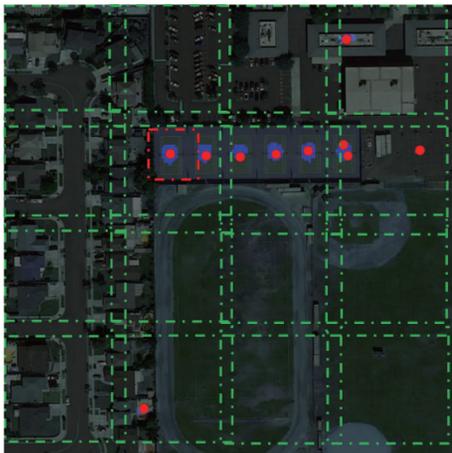


图 6 滑动窗口生成方法

Fig. 6 Sliding window generation method

### 3 实验与分析

#### 3.1 数据集

数据集来源于 DOTA 数据集<sup>[19]</sup>，包括来自不同传感器和平台的遥感图像，总计 2 806 张图像。每张图像的大小约为 4 000 × 4 000 像素。目标实例具有多样性，包括不同的尺寸、方向和形状。数据集中的图像经过专业的遥感图像解释专家标注，共涵盖了 15 种常见对象类别，包括建筑物、车辆、船只、飞机等。每个目标实例都有一个任意四边形标记，描述了目标的形状和方向。DOTA 包含了共计 188 282 个目标实例的完全标记数据。

#### 3.2 网络结构

在实验中，使用了两种不同的主干神经网络，即 ResNet-34<sup>[20]</sup> 和 U-Net<sup>[21]</sup>，用于 RPN 算法和尺度预测神经网络。为了进行公平比较，当使用 ResNet-34 作为主干网络时，添加了 U-Net 算法中的上采样阶段的部分神经网络结构，以输出两个尺度的特征图。对于 RPN 算法，使用修改后的

ResNet-34 和 U-Net 神经网络获得两个尺度的特征图，然后分别用于候选框的预测。针对这两层的预测结果，采用非极大值抑制 (NMS)<sup>[22]</sup> 算法进行后处理，最终输出 1 000 个候选框。

对于基于尺度预测神经网络的候选检测区域生成算法，使用修改后的 ResNet-34 和 U-Net 神经网络的最后一层特征图进行前景/背景预测和尺度预测。

尺度预测神经网络的训练采用了批量随机梯度下降方法，动量值为 0.9，权重衰减系数为 0.000 1，批大小为 4。初始学习率为 0.005，训练周期为 30 轮，学习率在第 18 和 24 轮进行 1/10 的衰减。对于 RPN 算法，初始学习率为 0.02，训练周期为 12 轮，学习率在第 8 和 12 轮进行 1/10 的衰减。在不同的尺度预测神经网络中，有效预测区域 valid\_ratio 超参数默认设置为 0.4， $\lambda$  比例因子设置为 5.0。

### 3.3 评价标准

对于 RPN 算法和基于尺度预测神经网络的候选区域生成算法，采用召回率指标进行精度的比较，采用检测速度指标进行速度的比较。在计算召回率的时候，本节只使用 IoU (交并比) 大于 0.5 作为候选检测框是否正确的评价阈值，且将目标标注框作为前景类别。

在比较 RPN 算法和基于尺度预测神经网络的候选区域生成算法时，采用召回率指标来衡量精度。召回率是指检测到的正样本数量与总正样本数量之比，通常用于评估算法的覆盖率；用 FPS 指标来衡量速度。在计算召回率时，仅使用 IoU 大于 0.5 的检测框作为正确的评价阈值，并将目标标注框视为前景类别，即只有检测到与目标标注框有足够重叠的检测框才会被视为正确的检测。

### 3.4 实验结果分析

DOTA 数据集包含 15 个目标类别，而且图像分辨率更大，目标尺度差异范围更大。由于存在分辨率大于  $2\ 048 \times 2\ 048$  像素的图像，为了公平比较 RPN 算法和多分辨率图像目标检测算法，对这些高分辨率的图像进行了预处理。预处理包括将图像划分为多个分片，以便输入到神经网络中进行训练或测试。

DOTA 数据集中具有更多的目标类别，其数据来源是真实的遥感图像数据，因此更加符合多分辨图像的应用场景。从表 1 中可以看出，在分辨率为  $800 \times 600$  的时候，U-Net 和 ResNet-34 两个不同的主干网络设置，SANet 的召回率均超过了 RPN 算法。这与 CPD<sup>[23]</sup> 和 WP<sup>[24]</sup> 数据集上的结果不一样，原因可能是 DOTA 数据集本身的图像分辨率较大，将图像压缩为  $800 \times 600$  会损失小目标特征的

表 1 DOTA 数据集上的召回率与速度测试结果

Tab. 1 Recall and speed test results on DOTA dataset

算法	主干网络	分辨率	召回率/%	检测速度/(f/s)
SANet	U-Net	$800 \times 600$	61.69	44.6
SANet+FPS_grouping	U-Net	$800 \times 600$	59.74	42.3
RPN	U-Net	$800 \times 600$	61.24	34.2
SANet	U-Net	1/2 分辨率	64.23	19.4
SANet+FPS_grouping	U-Net	1/2 分辨率	63.41	17.2
RPN	U-Net	1/2 分辨率	63.28	9.0
SANet	ResNet-34	$800 \times 600$	63.42	25.5
SANet+FPS_grouping	ResNet-34	$800 \times 600$	61.59	23.6
RPN	ResNet-34	$800 \times 600$	62.98	17.4
SANet	ResNet-34	1/2 分辨率	67.13	14.9
SANet+FPS_grouping	ResNet-34	1/2 分辨率	66.21	13.0
RPN	ResNet-34	1/2 分辨率	66.02	5.7

图像信息,从而降低了 RPN 算法的检测精度。而 SANet 受此因素影响较小,可能是由于尺度分类任务与位置回归任务相比,更加易于网络的学习。值得注意的是,当图像分辨率为原始输入的 1/2 时,同时使用 SANet 和 FPS\_grouping 算法可以达到比 RPN 更高的召回率结果,充分显示了该算法对多分辨率图像检测的优势。总的来说,当单独使用 SANet 算法的情况下,算法的召回率结果均比 RPN 算法更高;当同时使用 SANet 和 FPS\_grouping 算法时,召回率结果基本与 RPN 算法持平,但在检测速度上更快。

对于基础的尺度预测神经网络进行了 3 项修改,分别为有效预测区域设置、中心偏移率预测和前景/背景损失函数平衡。为了验证各项修改对原始算法的作用,设置了对比实验进行验证。在该实验中,使用 U-Net 作为主干网络,输入大小设置为原始图像的 1/2,并同时使用 SANet 和 FPS\_grouping 算法进行候选检测区域的预测。实验使用 DOTA 数据集进行训练和测试,统计并比较各实验的召回率结果。如表 2 所示,表格中的 Valid\_ratio、Centerness 和 Loss<sub>obj</sub> 分别代表 3 种修改方式。结果表明 3 种方法对于基准实验都有不同程度的提升。其中单独使用 Loss<sub>obj</sub> 时的提升最为明显,其主要原因是在单独使用有效预测区域和中心偏移率预测的时候,前后背景不平衡的问题依然存在,影响了神经网络的训练效果。当同时使用 3 项修改时,尺度预测神经网络的召回率相比其他设置都有明显的提升,召回率在 DOTA 测试集中达到 63.41%。

表 2 尺度预测神经网络各模块的测试实验

Tab. 2 Test experiment of each module of scale prediction neural network

基准	Valid_ratio	Centerness	Loss <sub>obj</sub>	召回率/%
✓				59.37
✓	✓			60.49
✓		✓		60.55
✓	✓	✓		61.62
✓			✓	61.46
✓	✓	✓	✓	63.41

#### 4 结束语

将多分辨率图像目标检测算法在 DOTA 数据

集上进行验证,使用两个不同的主干网络(U-Net 和 ResNet-34)进行设置,实验结果表明,该算法的召回率分别达到了 61.6%和 63.4%,检测速度分别达到了 44.6 f/s 和 25.5 f/s。这表明该算法在提高检测精度和检测速度方面取得了显著的进步。

#### 参考文献

- [1] 李喆,李建增,扈琪.基于候选区域的无人机侦察影像目标检测技术研究[J].战术导弹技术,2019(2):99-105.
- [2] 张鹏,徐丹.基于光电侦察的“低慢小”目标检测综述[J].中国电子科学研究院学报,2022,17(11):1042-1048.
- [3] 寇墨林,卓力,张菁,等.无人机影像目标检测技术研究进展[J].测控技术,2020,39(8):47-61.
- [4] 蒋兴浩,赵泽宇,许可.基于视觉的飞行器智能目标检测对抗攻击技术[J].空天防御,2021,4(1):8-13.
- [5] 王军敏,樊养余,李祖贺.基于深度卷积神经网络和迁移学习的纹理图像识别[J].计算机辅助设计与图形学学报,2022,34(5):701-710.
- [6] 翟宏亮.基于轻型卷积神经网络的无人机多目标检测系统研究[J].地理空间信息,2022,20(12):81-83,96.
- [7] 付苗苗,邓森磊,张德贤.深度神经网络图像目标检测算法综述[J].计算机系统应用,2022,31(7):35-45.
- [8] 杜紫薇,周恒,李承阳,等.面向深度卷积神经网络的小目标检测算法综述[J].计算机科学,2022,49(12):205-218.
- [9] He K M, Zhang X Y, Ren S Q, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.
- [10] Girshick R. Fast R-CNN[C]//Proceedings of the IEEE International Conference on Computer Vision (ICCV), Santiago, Chile: IEEE Press, 2015: 1440-1448.
- [11] 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.
- [12] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[C]//European Conference on Computer Vision. Cham: Springer, 2016: 21-37.
- [13] 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 Vi-

- sion and Pattern Recognition (CVPR). Las Vegas, USA; IEEE Press, 2016: 779-788.
- [14] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA; IEEE Press, 2017: 6517-6525.
- [15] 扶明, 郑霖, 杨超, 等. 基于深度卷积神经网络的慢动目标检测[J]. 传感器与微系统, 2022, 41(2): 111-114.
- [16] 范丽丽, 赵宏伟, 赵浩宇, 等. 基于深度卷积神经网络的目标检测研究综述[J]. 光学精密工程, 2020, 28(5): 1152-1164.
- [17] 黄健, 张钢. 深度卷积神经网络的目标检测算法综述[J]. 计算机工程与应用, 2020, 56(17): 12-23.
- [18] Redmon J, Farhadi A. YOLOv3: an incremental improvement[EB/OL]. (2018).: 1804.02767. <http://arxiv.org/abs/1804.02767v1>.
- [19] Xia G S, Bai X, Ding J, et al. DOTA: a large-scale dataset for object detection in aerial images[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA; IEEE Press, 2018: 3974-3983.
- [20] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Las Vegas, USA; IEEE Press, 2016: 770-778.
- [21] Ronneberger O, Fischer P, Brox T. U-Net:convolutional networks for biomedical image segmentation[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer, 2015: 234-241.
- [22] Neubeck A, Van Gool L. Efficient non-maximum suppression[C]//Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06). Hong Kong, China; IEEE Press, 2006: 850-855.
- [23] Wu Z, Su L, Huang Q M. Cascaded partial decoder for fast and accurate salient object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). Long Beach, USA; IEEE Press, 2019: 3902-3911.
- [24] 李叶, 毛伊敏, 陈志刚. 基于 Winograd 卷积的并行深度卷积神经网络优化算法[J]. 信息与控制, 2023, 52(4): 466-482.

引用格式: 杜琳琳, 张瑜, 唐宇, 等. 一种多分辨率图像目标检测算法[J]. 宇航总体技术, 2024, 8(3): 29-36.

Citation: Du L L, Zhang Y, Tang Y, et al. A multi-resolution image object detection algorithm [J]. Astronautical Systems Engineering Technology, 2024, 8(3): 29-36.