DOI:10.16136/j.joel.2022.06.0769

基于多尺度特征选择性融合的遥感图像检测算法

方明帅[1],黄友锐[1,2*],韩 涛[3]

(1.安徽理工大学 计算机科学与工程学院,安徽 淮南 232001;2.皖西学院,安徽 六安 237012;3.安徽理工大学 电气与信息工程学院,安徽 淮南 232001)

摘要:遥感图像的检测在监察自然环境、军事、国土安全等方面具有极其广阔的应用前景,而遥感图像具有背景复杂、目标面积小、特征提取困难等缺点,进行检测时容易产生小目标漏检问题。本文提出一种基于多尺度特征选择性融合的遥感图像检测算法。所提算法采用改进的Resnet50作为主干网络,将Resnet50第一个卷积替换成动态卷积,并将其ConvBlock模块中的卷积替换成金字塔卷积,提高特征提取能力。同时,为了避免遗漏底层信息,在动态卷积层后加入所提有效空间通道注意力机制模块。最后,选取基于上下文信息的不同尺度特征进行融合,提高了模型对目标物体的定位能力。实验结果表明,本文算法在保证速度的同时提高了对遥感图像的检测精度,在遥感图像公开数据集RSOD和NWPUVHR-10上平均精度均值(mean average precision,*mAP*)分别达到91.88%和90.23%,检测速度达到33 FPS。

关键词:目标检测;残差神经网络;上下文信息;多尺度特征融合;注意力机制

中图分类号:TP-391.41 **文献标识码**:A **文章编号**:s1005-0086(2022)06-0629-08

Remote sensing image detection algorithm based on selective fusion of multi-scale features

FANG Mingshuai[1], HUANG Yourui[1,2*], HAN Tao[3]

(1. School of Computer Science and Engineering, Anhui University of Science and Technology, Huainan, Anhui 232001, China; 2. Wanxi University, Lu'an, Anhui 237012, China; 3. School of Electrical and Information Engineering, Anhui University of Science and Technology, Huainan, Anhui 232001, China)

Abstract: The detection of remote sensing images has a wide range of applications in monitoring the natural environment, military, homeland security and so on, while remote sensing images have the disadvantages of complex background, small target area and difficulty in character extraction. In this paper, a remote sensing image detection algorithm based on selective fusion of multi-scale features is proposed. The proposed algorithm uses the improved Resnet50 as the backbone network, replaces the first convolution of the Resnet50 with dynamic convolution, and replaces the convolution in the ConvBlock module with pyramid convolution to improve feature extraction capability. At the same time, in order to avoid missing the underlying information, the proposed effective spatial channel attention mechanism module is added after the dynamic convolution layer. Finally, the different scale features based on context information are selected to fuse and improve the model's ability to locate the target object. The experimental results show that the algorithm improves the detection accuracy of remote sensing images while ensuring speed, and the mean average precision (*mAP*) reaches 91.88% and 90.23%, respectively, on the remote sensing image disclosure data set RSOD and NWPUVHR-10, and thedetection speed reaches 33 FPS.

Key words: target detection; residual neural network; context information; multi-scale feature fusion; attention mechanism

1 引 言

目标检测是计算机视觉中最基础也是最重要的部分,其任务是搜索出图像或视频中人们感兴趣的物体,并同时预测出它们的位置和大小。传统的检测模型主要是采用手工提取特征方法,导

* E-mail:1151698189@qq.com

收稿日期:2021-11-16 修订日期:2022-12-15

基金项目:国家自然科学基金(61772033)和安徽省科技重大专项计划项目(1603091012)资助项目

致了其计算复杂、检测速度慢、鲁棒性差等缺点，目标检测发展进入瓶颈期。直到 2014 年 Ross Girshick 提出循环卷积神经网络(recurrent convolutional neural network，RCNN)框架后，目标检测进入基于深度学习的时代。常见的目标检测方法有以 SPPNet(spatial pyramid pooling Net)[1]、Faster-RCNN[2]、Mask RCNN 等[3]为代表的 Two Stage 检测方法和以 YOLO[4]等为代表的 One Stage 检测方法。这些方法都可以归类为 Anchor-based 算法，其主要思想是在候选区域设置多个锚框，然后根据提取到的特征信息对锚框进行处理。为了提高检测精度，通常会设置大量不同形状的锚框，而大部分锚框在训练时都被标记成负样本即不完全包含目标的背景框，这样就会造成正负样本不均衡导致检测速度变慢。为了解决这些问题，Anchor-free 算法开始不断地被提出，其方法不再使用锚框，而是通过确定关键点来回归目标的宽高类别和位置，大大减少了网络超参数的数量，主要代表有 CenterNet[5]、FCOS(fully convolutional one-stage)等[6]方法。

针对不同任务的目标检测数据集有多种，其中遥感图像是其重要的部分且由于遥感图像成像方式和拍摄角度的特殊性导致了具有目标尺寸差异大、样本不均衡、提取特征困难等问题。如果将检测模型直接应用到遥感图像上，检测效果并不理想，需要将检测算法与遥感图像特点相结合优化检测模型。于是，2017 年 JIANG 等[7]提出 R2CNN 方法，在 Faster-RCNN 的基础上，将原来的感兴趣区域池化(RoI pooing)改进为多尺度池化，同时为了适应遥感图像中面积较小的目标增加了两种不同尺度的池化方式，但是由于先验框的数量成倍增加，影响了检测速度。2019 年 YANG 等[8]提出 R3Det 方法，主要思想是将单阶段检测器 RetinaNet 改变成两阶段检测器，利用特征金字塔网络(feature pyramid networks，FPN)实现多尺度特征提取，在每个尺度的特征图上分别进行检测，减少了目标的漏检。本文以 CenterNet 为基础，从改进卷积结构的主干网络、基于上下文信息的特征融合和注意力机制 3 个角度出发，提出了基于 CenterNet 的多尺度特征选择性融合的遥感图像检测算法，具体来说，本文贡献主要如下：

1) 将主干网络 Resnet50 中第一个卷积替换成动态卷积(Dynamic Conv[9])，它可以提升模型表达能力且无需提升网络深度与宽度。再将 Resnet50 中的 ConvBlock 模块里的 3×3 卷积替换成金字塔化卷积(Pyramidal Conv[10])，进行多尺度特征提取，提高特征表达能力，优化输出。

2) 针对遥感图像目标小、特征提取易丢失等特点，引入 ESCA (efficient space channel attention)注意力机制并添加在底层提取特征层之间抑制无用信息。

3) 由于 Resnet 只是依次向下提取特征并没有充分利用各个特征层之间的关系，所以又增加了基于上下文的特征融合模块，将不同尺度的特征利用起来，减少了特征信息的丢失，避免小目标的漏检。

2　本文算法整体结构

本文算法整体框架如图 1 所示，以编码-解码方式的 CenterNet 作为基础模型，利用热力图获取物体中心点的位置和物体类别、中心点偏置信息以及检测框的宽高。在编码网络中将主干特征提取网络替换成Resnet50并采用动态卷积和金字塔化卷积对

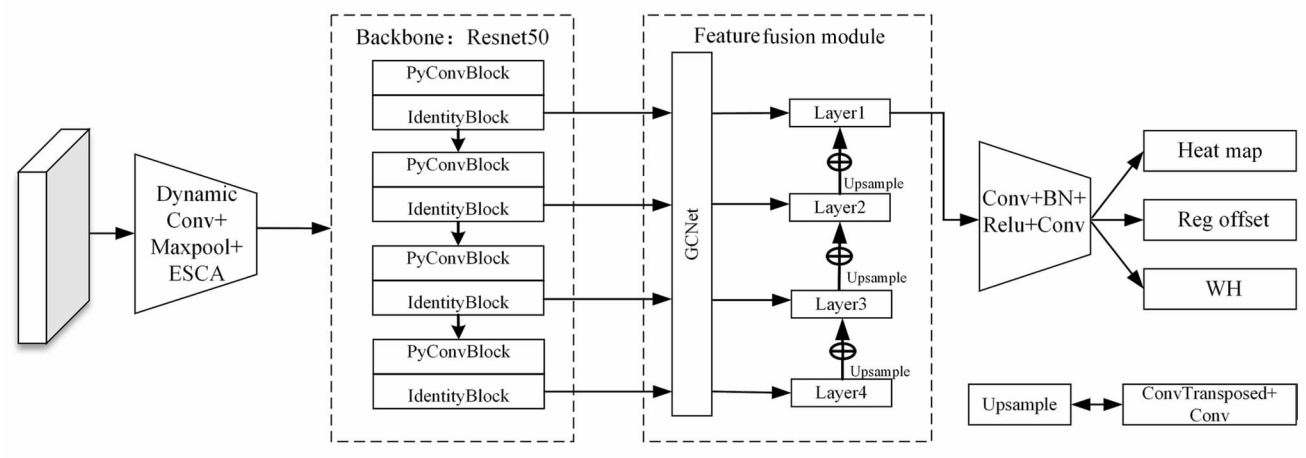

图 1　整体框架

Fig. 1　The overall frame

其进行改进，在解码网络中增加了特征融合模块，最后将得到的特征图传入分支预测网络中得到3个输出。

2.1　改进的 Resnet50 结构

针对遥感图像目标面积小、特征提取易丢失的问题，改进了 Resnet50 网络。Resnet50 网络在数据传入之后，采用的是一个 7×7、步长为 2 的卷积计算，而在最底层较大的感受野会导致一些细节特征的遗漏。所以本文将 7×7 卷积替换成 3×3、步长为 2 的动态卷积，减小感受野的同时增加模型非线性。动态卷积是在卷积层中使用多个卷积核，并且增加注意力机制去结合不同卷积核运算之后的结果，这样可以提取到更加丰富的特征。动态卷积结构如图 2 所示。

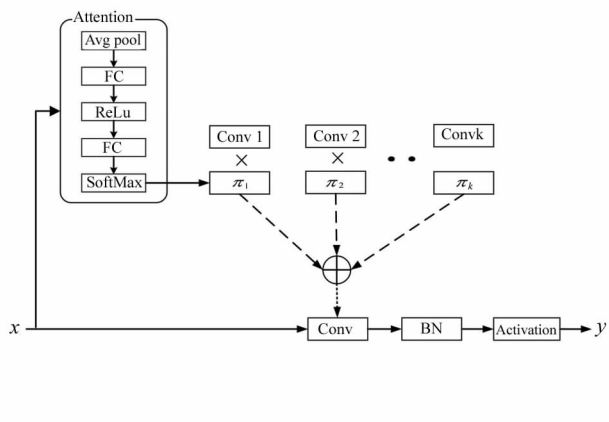

图 2　DyConv 卷积流程

Fig. 2　DyConv convolution process

Attention 模块的输入是本层的卷积核，经过 Attention 模块得到权重 $\pi_k(x)$ 与 $Conv_k$ 进行相乘，再累加获得聚合的卷积核，然后根据获取的卷积核对输入特征图 x 进行卷积运算。

另外，将 Resnet50 的 ConvBlock 模块里 3×3 卷积替换成 PyConv 变成 PyConvBlock 模块，使其每一层都拥有不同尺寸的卷积核进行多尺度特征提取，同时为了降低计算复杂度，对不同尺寸的卷积核进行分组卷积，之后再将不同组的输出进行拼接。具体操作如图 3 所示，首先将输入通道数进行 1×1 卷积变成 64 个之后进行批量归一化(batch normalization，BN)和激活(ReLu)。其次分成 4 组进行卷积且每组卷积核大小不同，分别为 9、7、5、3。接着将每组输出进行拼接之后进行 BN 和 ReLu，最后采用 1×1 卷积将通道数扩张成与原输入相同且进行相加。

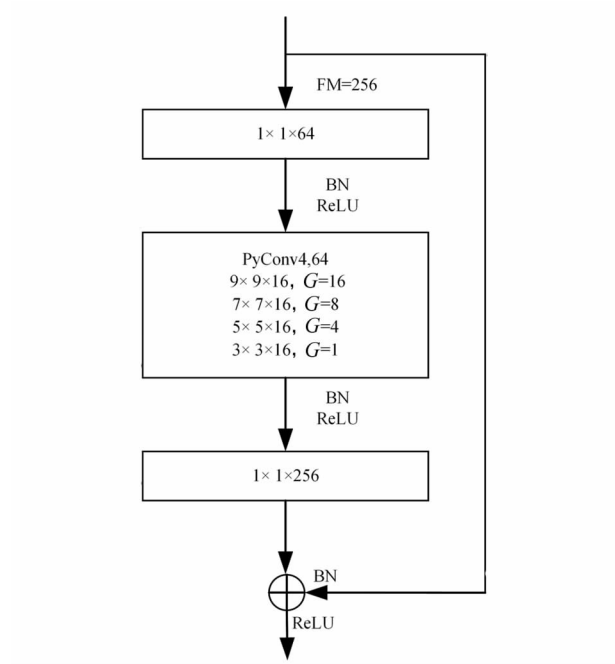

图 3　PyConv 卷积流程

Fig. 3　PyConv convolution process

2.2　ESCA 注意力机制

Efficient Channel Attention(ECA)[11]是一种改进 SE-Attention(squeeze-and-exciation)[12]的注意力机制，提出计算所有通道之间的注意力是非必要的。ECA 在不改变通道数的全局平均池化之后通过一维卷积来计算相邻 k 个通道的注意力，并将其与输入进行像素级相乘，但是 ECA 只对通道做了一个局部注意力。本文提出 ESCA 注意力模块，对 ECA 增加了空间注意力。首先，用一个全局平均池化加上 1×1 卷积将输入变成 1×1×16。接着进行 repeat 操作变成 $H\times W\times 1$ 的权重，最后将得到的权重与原输入相乘再与 ECA 输出进行拼接。ESCA 结构如图 4 所示，$k=3$ 代表计算相邻 3 个通道的注意力，⊗代表相乘，⊕代表相加。

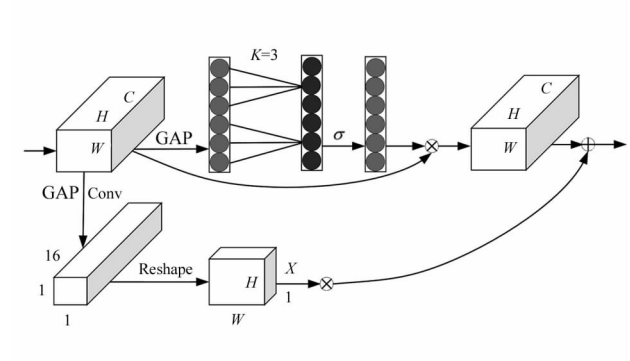

图 4　ESCA 注意力结构

Fig. 4　ESCA attention structure

2.3　基于上下文信息的特征融合模块

获取目标长距离依赖关系对视觉场景进行全局理解是有效的，GCNet[13] 是将查询到的上下文信息聚合到每个查询位置上来获取依赖关系。GCNet 上下文构建模型主要分为以下 3 步：

（1）上下文模型构造模块，聚集所有位置上的特征，生成一个全局背景特征；

（2）特征转换模块，用来捕获各个通道之间的相互依赖关系；

（3）融合模块，将全局上下文特征合并为所有的位置特征。

多尺度特征融合是提高检测性能的重要方法，浅层感受野小包含更多的细节信息，对小目标敏感，但噪声具多。深层感受野大，可以具有较强的语义信息，泛化性能好，但容易遗漏细节特征。

本文利用 GCNet 对不同尺寸的特征图进行全局上下文特征聚合，提出了基于上下文信息的特征融合方法。首先选取 Resnet50 中 4 个 Identity Block 模块里的输出特征图作为特征融合的基础，传入 GCNet 中聚合上下文信息，分别记为{C2，C3，C4，C5}。其次采用反卷积将 C5 变成与 C4 相同尺寸的特征图 P5，在此过程中为了减轻反卷积带来的“棋盘效应”，除了将卷积核尺寸调整至能被步长整除之外，还在反卷积后跟上一个 1×1 的卷积调整通道数。接着将 P5 与 C4 相加得到 P4，依次执行相同操作，得到输出特征图 P2。最后将添加了 ESCA 注意力机制的底层特征图与特征融合的输出 P2 进行拼接，传入解码网络中对其进行预测。具体流程如图 5 所示。

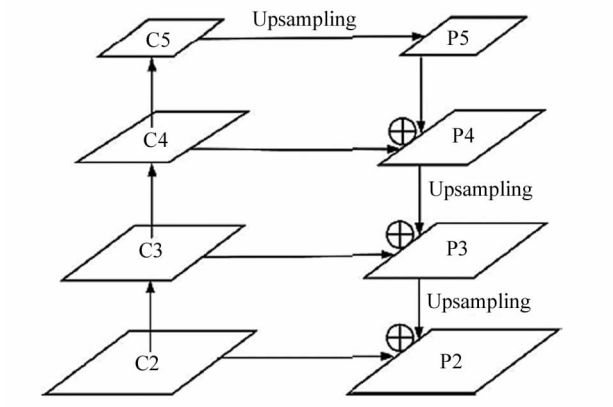

图 5　基于上下文信息融合模块

Fig. 5　A context-based information fusion module

2.4　损失函数

因为在解码网络中产生了 3 个分支，所以损失函数包含了中心点预测损失（L_k）、中心点偏置损失（L_o）以及宽高损失（L_s）。为使模型在反向传播时参数收敛更快，根据各个分支的重要程度对分支损失函数设置了权重，分别是 $\lambda_s=0.1$，$\lambda_o=1$，总体的损失函数为：

$$L_{det}=L_k+\lambda_s L_s+\lambda_0 L_0。\tag{1}$$

中心点预测损失函数 FocalLoss 的引入，主要是用来处理样本不均匀的问题，是针对 CenterNet 算法改进，L_k 中心点损失函数为：

$$L_k=\frac{-1}{N}\sum_{xyc}\begin{cases}(1-\hat{Y}_{xyc})^{\alpha}\log_2(\hat{Y}_{xyc}),\text{if}\quad Y_{xyc}=1\\(1-Y_{xyc})^{\beta}(\hat{Y}_{xy}{}^{\alpha}\log_2(1-\hat{Y}_{xyc}),\text{otherwise}\end{cases},\tag{2}$$

式中，N 为当前检测图片中目标个数，xyc 表示热力图 xy 位置上当前类别为 c，将 α，β 分别设置为 2 和 4，可以更好地提高算法检测精度。

L_s 偏置损失函数为：

$$L_s=\sum_{k=1}^{N}|\hat{S}_{pk}-S_k|,\tag{3}$$

式中，k 代表当前检测目标，$\hat{S}_{pd}$ 是预测出来的宽高信息，S_k 是真实宽高信息。

经过特征提取网络之后对图像进行了下采样，将得到的热力图映射到原始图像上时中心点位置会有微小的误差，为了消除误差就引入了 L_o 偏置损失：

$$L_o=\frac{1}{N}\sum_{p}|\hat{O}_p-(\frac{p}{R}-[\frac{p}{R}])|,\tag{4}$$

式中，p 代表目标中心点坐标，$\hat{O}$ 是预测出来的中心点坐标，R 为下采样因子。

3　实验结果与分析

3.1　数据集

为了检验改进算法的有效性，本文采用武汉大学发布的 RSOD 遥感图像数据集进行消融实验，主要针对该数据集图片模糊噪声大，而且在飞机类别上小目标多。另外采用复杂背景下多类别、尺寸差异大的航天遥感图像数据集 NWPUVHR-10 进行验证，数据集格式均采用 VOC 格式进行处理。

RSOD 包含标注图片 976 张，7 400 个目标信息，由 4 类目标组成分别为：飞机（4 993 架）、操场（191 个）、油桶（1 586 个）和立交桥（180 座）。如表 1 所示，对数据集目标框的尺寸进行了统计，结果显示 100 pixel 以下占 74.43%。

NWPUVHR-10 总共有 800 张图片，其中标注图片 650 张，背景信息图片 150 张。

目标分为 10 类：棒球场、网球场、篮球场、田径

场、飞机、舰船、油罐、港口、桥梁和车辆。表 2 对数据集目标框的尺寸进行了统计，结果显示大部分目标都集中在100 pixel以下，少量分布在300 pixel左右。

表 1　数据集 RSOD 目标尺寸统计表

Tab. 1　Data set RSOD target size statistics table

Scales /pixel	0-10	10-40	40-100	100-300	300-500	>500
Width	0	0.21	0.53	0.22	0.023	0.007
Height	0	0.27	0.49	0.18	0.003	0.007

表 2　数据集 NWPUVHR-10 目标尺寸统计表

Tab. 2　Data set NWPUVHR-10 target size statistics table

Scales /pixel	0-10	10-40	40-100	100-300	300-500	>500
Width	0	0.13	0.69	0.15	0.012	0
Weight	0	0.14	0.72	0.12	0.010	0

本文中采用“目标-图像比”定义图片中小目标的个数，即 $\mathrm{Ratio}_s = S_t/S$，其中 S_t 是目标框的面积，S 是目标所在图片的面积，当 $\mathrm{Ratio}_s \in 0, 0.01]$时，就将该目标定义为小目标。表 3 列出了两个数据集中小目标的确切数量，RSOD 数据集中小目标多集中在飞机目标，NWPUVHR-10 数据集中小目标多集中在油桶目标上，部分数据集示例如图 6 所示。

表 3　两个数据集中小目标的类别与数量

Tab. 3　The categories and number of small targets in both datasets

Dataset	Plane	Oil tank	Ship	Total
RSOD	1 503	2	—	1 505
NWPUVHR10	7	91	5	201

图 6　数据集示例

Fig. 6　Example of a dataset

对数据集中的图片进行数据增强操作，包括随机左右翻转，上下翻转以及比例放缩等，网络训练的优化器为 Adam，初始学习率设置为 0.01，训练轮数 200，批处理大小设置为 8 张图片，使用预训练网络迁移学习。图像预处理阶段对输入图像进行统一缩放操作，进入特征编码网络的图像尺寸为 512×512，训练集与验证集划分比例为 8∶2。

3.2　实验环境

实验设备为 Windows10 和 Ubuntu18.04 系统计算机，运行环境 Pytorch1.8.0，Cuda 版本为 11.2，GPU 型号为 RTX3060，CPU 型号为 Intel(R)Core(TM)i5-10600KF，内存为 16 G。

3.3　评价指标

本次实验采用 mAP 作为评价指标，mAP 指的是数据集中每个类别的平均精度(average precision，AP)之和比上类别数，而 AP 则是由精确率(*Percision*)和召回率(*Recall*)所确定的。精确率是针对模型预测结果而言的，它代表的是预测为正的样本中真实正样本所占比例。召回率是针对原来的样本而言的，它表示的是样本中的正例有多少被预测正确了。计算如式(5)—(7)所示：

$$Percision = \frac{TP}{TP + FP}, \tag{5}$$

$$Recall = \frac{TP}{TP + FN}, \tag{6}$$

式中，TP(true positives) 正确预测且判断为正样本，TN(true negatives) 正确预测且判断为负样本，FP(false positives) 错误预测且判断为正样本，FN(false negatives) 错误预测且判断为负样本。

$$mAP = \frac{1}{N}\int_0^1 P(R)\mathrm{d}R。 \tag{7}$$

mAP 是由横坐标为召回率和纵坐标为精确率所构成的 PR 曲线的面积比上种类个数 N。

本文根据预测框(predection result，PR)与真实框(ground truth，GT) 的交并比(Iou)来判断是否有效检测到物体，如式(8)所示：

$$Iou = \frac{PR \cap GT}{PR \cup GT}。 \tag{8}$$

3.4　对比实验

数据集 RSOD 中只含了 4 类目标，缺乏目标多样性，因此为了检验本文改进算法的可靠性，在数据集 NWPUVHR-10 上将本文算法与其他算法进行对比实验。相较于 RSOD 数据集，NWPUVHR-10 数据集包含种类多，特征差异明显，尺度变化大，检测难度高。

由表4可知：本文算法在数据集NWPUVHR-10上平均检测精度达90.2%，不过在篮球场这类目标上由于背景复杂且训练数据较少导致检测精度低，仅有50.77%，除此之外，在其他7类目标上的检测效果优于其他算法，证明了本文算法相比于其他算法在遥感图像检测中具有优势；平均检测精度分别高出Faster-RCNN[5]，MaskRCNN[6]，YOLOv4[14]，YOLOX-S[15]算法的13.8%，6.3%，3.5 %和3.8%。相比于其他针对遥感图像检测的算法，如Sig-NMS[16]，RICAOD[17]，YOLO-RS[18]算法，平均检测精度分别提高了7.3%，3.1%，4.1%。

表4 不同算法在NWPUVHR-10数据集上的对比实验

Tab. 4 Comparative experiments on NWPUVHR-10 datasets by different algorithms

Target and *mAP*	*AP*/%								
	Faster-RCNN	Mask RCNN	RFBNet 300	YOLOv4	Sig-NMS	RICAOD	YOLO-RS	YOLOX-S	Improved algorithm
Plane	82.8	93.2	97.2	94.9	90.8	99.7	98.9	86.3	99.91
Warship	77.5	75.5	77.4	78.6	80.5	90.8	86.6	84.7	88.02
Oil tank	52.5	92.9	59.8	95.4	59.2	90.6	94.3	91.2	98.79
Baseball field	96.3	90.4	97.7	98.3	90.8	92.9	97.3	90.3	98.90
Tennis court	62.9	90.3	81.6	88.2	80.9	90.3	83.9	79.6	96.26
Basketball field	68.8	91.2	93.8	67.5	90.9	80.3	71.7	75.9	50.77
Athletics field	98.4	95.2	96.5	99.3	99.3	90.8	97.7	90.7	99.98
Port	82.5	75.2	98.5	80.7	90.3	80.3	84.1	86.3	99.99
Bridge	78.8	60.6	97.6	95.9	67.8	68.5	62.3	88.2	90.13
Vehicle	63.8	74.2	55.2	67.7	78.1	87.1	86.9	91.3	79.06
mAP/%	76.4	83.9	85.5	86.7	82.9	87.1	86.1	86.4	90.2

3.5 消融实验

为了检验本文改进算法的性能，针对本文所采用的RSOD数据集进行消融实验，依次增加新的卷积结构的Resnet50、ESCA注意力机制、基于上下文信息的特征融合模块，选取AP_{50}作为性能评价指标，由表5可见：通过对比方法1和2表明，引入新的卷积结构改进后的算法，相对于原算法检测精度提升了5.68%，表明主干网络在引入动态卷积和金字塔化卷积替代标准卷积后，网络的特征提取能力明显加强；通过对比方法2和3表明，引入ESCA注意力机制后，检测精度提高了2.59%，验证了在低层增加注意力抑制无用信息是有效的；通过对比方法3和方法4表明，引入上下文特征融合模块后，检测精度提高了3.86%。消融实验表明，本文采用的改进方法对检测效果均有提升。

图7为消融实验不同方法在数据集RSOD上对部分图片的检测结果，方法4为本文最终使用方法。

由图7可见：4种方法在检测飞机目标上差异较大且本文方法在飞机目标上没有漏检，在油桶目标上都有目标漏检，不过本文方法效果最好，从图中对应位置操场和立交桥目标的置信度上看，本文方法的检测置信度最高。总体可见，本文提出的改进方法针对RSOD数据集在漏检问题上有明显改善，但立交桥目标由于背景复杂且背景区域中有相似目标干扰，所以检测效果不理想。

表5 消融实验

Tab. 5 Ablation experiment

Method	Improved Resnet50	ESCA attenation	FPN	Oil tank /%	Play ground /%	Plane /%	Overpass /%	*mAP*	Detection speed/FPS
1	×	×	×	88.35	86.94	87.38	56.34	79.75	55
2	√	×	×	95.47	92.62	89.01	64.62	85.43	35
3	√	√	×	98.54	93.52	90.98	69.07	88.02	35
4	√	√	√	99.83	95.75	94.17	77.15	91.88	33

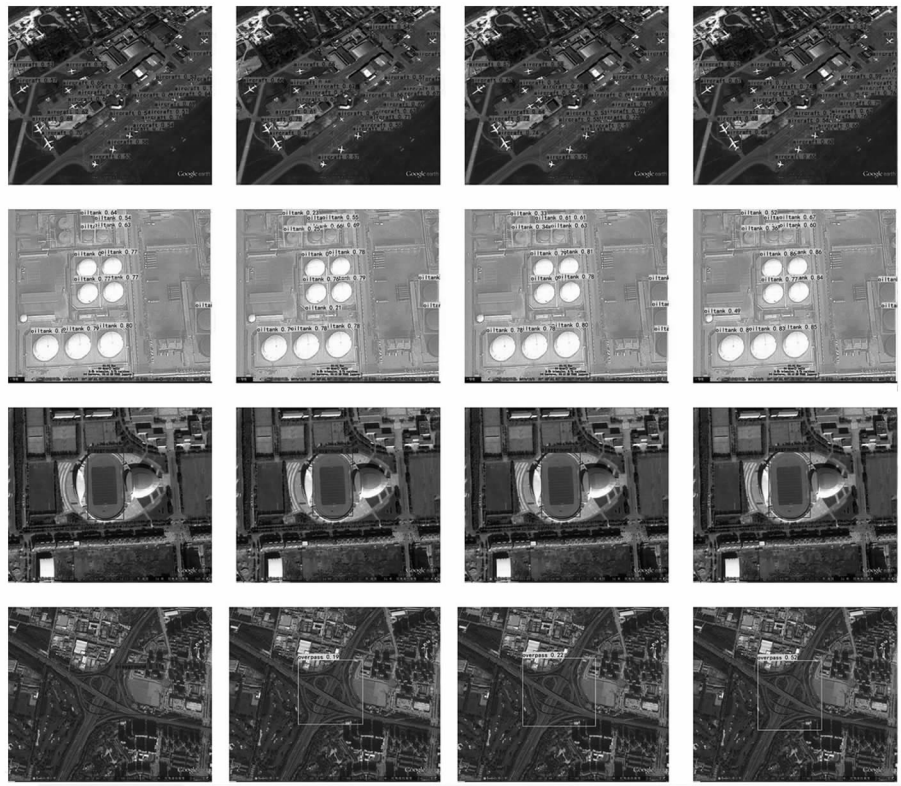

图 7　消融实验不同方法检测图片可视化

Fig. 7　The ablation experiment detects picture visualization in different ways

4　结　论

针对遥感图像背景复杂，待检测目标小且分辨率低等问题，本文提出了基于多尺度特征选择性融合的检测算法。使用金字塔化卷积和动态卷积来代替部分标准卷积，提高特征表达能力。使用了 ECSA 注意力机制，抑制无用信息。同时增加了上下文信息的多尺度特征融合模块，充分利用不同尺度特征，避免小目标漏检。最终，所提方法分别在 ROSD、NWPUVHR-10数据集上进行实验达到了91.88%和90.23%的检测精度，充分证明了算法的可行性和可靠性。由于数据集较小，还有硬件设施等问题，所以检测效果仍存在一定的局限性，下一步将使网络更轻量化提高实时性。

参考文献：

[1] HE K，ZHANG X，REN S，et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence，2014，37(9)：1904-1916.

[2] 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.

[3] HE K，GKIOXARI G，DOLLAR P，et al. Mask R-CNN[C]// 2017 IEEE International Conference on Computer Vision (ICCV)，October 22-29，2017，Venice，Italy. New York：IEEE，2017：2980-2988.

[4] 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，2016：779-788.

[5] ZHOU X，WANG D，KRHENBUHL P. Objects as points [EB/OL]. (2019-04-25) [2021-11-16]. https://arxiv.org/abs/1904.07850v2.

[6] TIAN Z，SHEN C，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，2019：9626-9635.

[7] JIANG Y，ZHU X，WANG X，et al. R2 CNN：rotational region CNN for arbitrarily-oriented scene text detection [C]//2018 24th International Conference on Pattern Rec-

ognition (ICPR), August 20-24, 2018, Beijing, China. New York: IEEE, 2018: 3610-3615.

[8] YANG X, LIU Q, YAN J, et al. R3det: refined single-stage detector with feature refinement for rotation object[EB/OL]. (2020-12-08) [2021-11-16]. https://arxiv.org/abs/1908.05612v6.

[9] CHEN Y, DAI X, LIU M, et al. Dynamic convolution: attention over convolution kernels[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE, 2020: 11027-11036.

[10] Duta I C, LIU L, ZHU F, et al. Pyramidal convolution: rethinking convolutional neural networks for visual recognition[EB/OL]. (2020-06-20) [2021-11-16]. https://arxiv.org/abs/2006.11538v1.

[11] WANG Q, WU B, ZHU P, et al. ECA-Net: efficient channel attention for deep convolutional neural networks[C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE, 2020: 11531-11539.

[12] 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, 2018: 7132-7141.

[13] CAO Y, XU J, LIN S, et al. GCNet: Non-local networks meet squeeze-excitation networks and beyond [C]// 2019 IEEE/CVF International Conference on Computer Vision Workshop, October 27-28, 2019, Seoul, Korea (South). New York: IEEE, 2019: 1971-1980.

[14] BOCHKOVSKIY A, WANG C Y, LIAO H. Yolov4: optimal speed and accuracy of object detection[EB/OL]. (2020-04-23) [2021-11-16]. https://arxiv.org/abs/2004.10934v1.

[15] ZHENG G, SONG T L, FENG W, et al. YOLOX: exceeding YOLO series in 2021[EB/OL]. (2021-08-06) [2021-11-16]. https://arxiv.org/abs/2107.08430v2.

[16] DONG R, XU D, ZHAO J, et al. Sig-NMS-based faster R-CNN combining transfer learning for small target detection in VHR optical remote sensing imagery[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57 (11): 8534-8545.

[17] LI K, CHENG G, BU S, et al. Rotation-insensitive and context-augmented object detection in remote sensing images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2017: 2337-2348.

[18] ZHANG Y, YANG H T, LIU X Y. Research on remote sensing image object dectection method based on densely connected multi-scale features[J]. Journal of China Academy of Electronics and Information Technlongy, 2019, 14 (5): 530-536.
张裕, 杨海涛, 刘翔宇. 基于多尺度特征稠密连接的遥感图像目标检测方法[J]. 中国电子科学研究院报, 2019, 14(5): 530-536.

作者简介:

黄友锐 (1971—), 男, 博士, 教授, 博士生导师, 研究方向为智能控制和矿山物联网.