基于改进 RetinaNet 的遥感图像目标检测算法

程　路[1]，刘家伟[2]，周庆忠[1]，郑宇超[1]，刘伟[1]
（1. 华东交通大学软件学院，江西 南昌 330013；2. 萍乡学院数学与计算机学院，江西 萍乡 337055）

摘　要：【目的】遥感图像目标检测在智慧交通方面有广泛的应用，如路网运行状态动态监测、道路智慧执法、公路灾害智能监测等。由于遥感图像具有目标小而密集、尺度变化大且以任意方向分布等特点，通用目标检测器直接应用于遥感图像时检测效果不佳。【方法】针对以上挑战，提出了一种基于改进 RetinaNet 的遥感图像目标检测算法。本文算法结合了下采样块和卷积核动态选择的优势。首先，该模型在基础特征提取网络 ResNet50 上引入一个改进的下采样模块，对特征进行多种下采样处理，然后采用卷积核选择机制动态选择空间感受野，以此对多尺度的语义信息进行建模，最后得到目标物体的分类和回归结果。【结果】实验结果表明，该方法在大规模遥感图像目标检测数据集 DOTA 上的平均精度均值比原 Retinanet 网络提升了 3.2%。【结论】通过引入下采样模块和动态选择卷积核大小的机制，本文算法在一定程度上改进了对多尺度遥感目标的识别能力。

关键词：遥感图像；目标检测；下采样；卷积核选择
中图分类号：U1，TP753　　**文献标志码：**A

An Object Detection Algorithm for Remote Sensing Images Based on Improved RetinaNet

Lu Cheng[1], Jiawei Liu[2], Qingzhong Zhou[1], Yuchao Zheng[1], Wei Liu[1]
（1.School of　software, East China Jiaotong University, Nanchang 330013, China；2. College of　Mathematics and Computer, Xinyu University, Xinyu 338004, China)

Abstract: Remote sensing image object detection has a wide range of applications in intelligent transport, such as dynamic monitoring of road network operation status, intelligent law enforcement on roads, and intelligent monitoring of road disasters. Due to the characteristics of small and dense targets, large scale changes, and arbitrary direction distribution in remote sensing images, general object detectors have poor detection performance when directly applied to remote sensing images. To address these challenges,this paper proposes a remote sensing image object detection algorithm based on improved Retinanet. First, the model introduces Improved Downsampling Module (IDM) on the base feature extraction network ResNet50, which performs multiple down-sampling processing on features, and then dynamically selects the spatial receptive field using the convolution kernel selection mechanism to model the multi-scale semantic information of the scene. Finally, the classification and regression results of the target object are obtained. Experimental results show that the proposed method improves the mAP by 3.2% on the large-scale remote sensing image object detection dataset DOTA compared to the original Retinanet network, enabling more accurate localization and identification of remote sensing targets.

Key words: Remote sensing images; object detection; downsampling; convolution kernel selection

收稿日期：2024−03−07
基金项目：江西省自然科学基金面上项目(20232BAB203057)

【研究意义】遥感技术和计算机技术的进步促进了全球范围内及时获取和使用高空间分辨率遥感数据的广泛应用[1][2][3]。遥感图像目标检测是一种高分辨率遥感图像内容解析中的关键任务，旨在精确识别与定位遥感图像中的特定目标物体，如车辆、船舶及飞机等。这一技术在高精度遥感图像智能分析领域具有举足轻重的地位，并广泛应用于智能交通、城市规划以及地理信息系统更新等多个领域。

【研究进展】近年来，深度学习的飞速发展在通用目标检测领域取得了显著的进步。然而，在遥感图像分析这一特定领域，传统的通用目标检测方法往往难以达到预期的效果。这些方法大多依赖于人工提取的特征来进行目标识别，虽然取得了一定的成果，但在效率、鲁棒性和整体性能方面仍然存在明显的局限性。相比之下，基于卷积神经网络（Convolutional Neural Networks，CNN）的深度学习[1]目标检测框架以其强大的特征表示能力，为解决这一问题提供了新的可能。这种框架在特征图上设置了一系列的锚点，并对每个锚点进行分类和回归处理，从而能够准确地识别出目标对象的类别及其对应的边界框。

基于 CNN 的目标检测算法已经成为当前主流的目标检测算法，主要分为：双阶段检测算法和单阶段检测算法两类。双阶段算法 RRCNN[5]通过增加旋转的 RoI 池化层以及不同类间的非极大抑制实现 RoI 特征与目标方向更好的对齐，以提高检测效率；R2CNN[6]采用小尺寸的锚框设计，提升了小目标检测能力。尽管这些方法有效地检测了旋转目标，但他们需要预设大量的密集排布的旋转锚点，存在冗余计算和类别不平衡问题。为了缓解上述问题，RoI Transformer[7]采用全连接层学习生成旋转候选框。Oriented RCNN[8]和 Gliding Vertex[9]则设计边界框编码方式，减轻了由旋转角度周期性造成的训练损失不稳定。单阶段算法 R3Det[10]通过多个优化模块的级联进行精细化回归，并采用基于水平锚的特征重构和对齐的特征插值技术，实现目标的高效检测。DRN[11]使用特征选择模块和动态细化检测头改善了遥感图像中目标密集且方向任意的问题。

【关键问题】虽然上述方法在一定程度上提高了遥感图像目标检测的性能，但在应对具有明显尺度差异或目标小而密集的场景时，其性能仍然有待进一步提升。

【创新特色】为解决上述问题，本文在 RetinaNet[12]的基础上，引入了改进的下采样模块，将其嵌入到 ResNet50[13]骨干网络中，融合多种下采样方法提取到的特征来生成下采样特征图，然后利用卷积核选择机制动态选择空间感受野，从而更加关注重要的信息并抑制次要的信息。在公开的数据集 DOTA[14]上的实验结果表明，本文设计的模型能够有效应对目标对象尺寸小且尺度变化大的挑战，提高了在遥感图像上的目标检测能力。

1 相关工作

1.1 特征下采样

下采样是 CNN 中的一个关键步骤，其主要目的是在降低骨干网络中特征图的分辨率同时，保留特征图中的重要信息。此外，下采样还有助于提高神经网络模型的计算效率并增强模型的泛化能力。常见的下采样方法包括最大池化[15]、平均池化和步幅卷积[16]等。最大池化是一种广泛使用的下采样方法，它在特征图中的预定义窗口内选取最大值。与之相似，平均池化则是从同样的窗口中选取平均值。这两种方法在计算上都很高效，并且有助于降低特征图的维度。然而上述做法可能会导致一些关键的空间信息丢失，降低模型的泛化性。步幅卷积则通过增大跨步，在卷积过程中跳过某些像素，缩减特征图的尺寸，但这可能导致特征图稀疏以及信息密度减小，遗漏某些关键信息。LIP[17]通过局部重要性池化自适应地调整每个局部区域内池化的权重，有效丢弃缺乏信息的特征。SoftPool[18]将输入的数据映射到一个连续的值空间，然后对这个连续值进行池化操作。这种操作方式使得 SoftPool 可以更好地参考区域内的激活值分布，因为它的输出服从一定的概率分布，而最大池化和平均池化的输出是无分布的。

1.2 核选择机制

卷积核选择机制是一种动态上下文建模的自适应技术，模型根据不同的上下文信息动态地调整其注意力，从而产生具有不同感受野大小的神经元，实现高性能的特征提取。SENet[19]对每个卷积层的特征图进行压缩，通过激活函数重新加权特征通道。SGE[20]通过对特征图上各个空间位置生成注意力因子来调节每个空间位置处特征的重

要性。Dynamic Convolution[21]根据不同的输入特征动态聚合多个平行卷积核，这些内核通过注意力以非线性方式聚合，模型能够表现出更强的表示能力。SKNet[22]是一种使用多分支卷积网络、组卷积、空洞卷积和注意力机制的卷积网络，是在网络的不同层使用不同的核函数。SCNet[23]引入了自校准卷积层，通过自校准操作融合来自两个不同空间尺度的信息。上述算法的灵活性有待进一步提高，以使卷积核大小的自适应选择，能够根据输入内容动态调整其感受野大小。为此，本文算法在预测网络中引入卷积核选择机制，在不同尺度上动态调整不同核的权重，从而使得网络能够聚焦于不同尺度的特征。

2 模型设计

本文设计的目标检测网络在 RetinaNet 上进行改进，由三部分组成：骨干网络、特征金字塔和分类回归子网。本文在 RetinaNet 的骨干网络中引入改进的下采样模块，增强模型捕获复杂细节的能力；在目标检测网络中加入核选择模块，增强网络提取并融合多尺度特征信息的能力。

2.1 基于改进 ResNet50 的特征提取网络

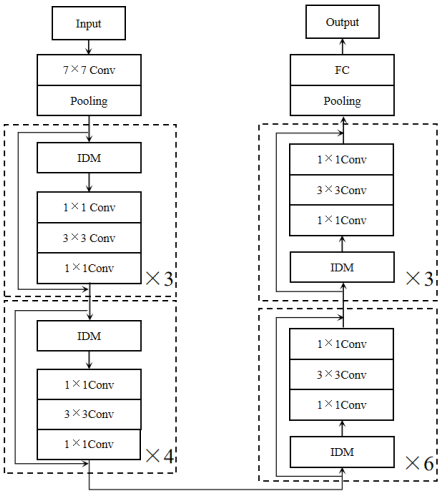

图 1　改进 ResNet50 的网络结构示意图

Fig.1　Illustration of the improved network structure of ResNet50

在遥感图像中，目标尺度变化较大，且小目标的数量占比很高，原始的ResNet50网络采用的下采样方法主要依靠卷积层进行，这可能会导致一些关键的语义信息被遗漏，同时难以充分挖掘和保留细粒度的特征信息。为解决这一问题，本文引入了一种改进的下采样模块(Improved Downsampling Module，IDM)，其在网络中的位置如图1所示。

特征提取网络ResNet50包括一系列堆叠的残差模块，每个残差块包含多个卷积层和恒等映射，在进行残差学习时采用 IDM 模块进行下采样，其结构如图 2 所示。本文使用三个分支对输入的特征进行处理，实现了多尺度特征的提取与融合，增强了特征的表示能力，从而减少了模型在小目标检测时的细节丢失。

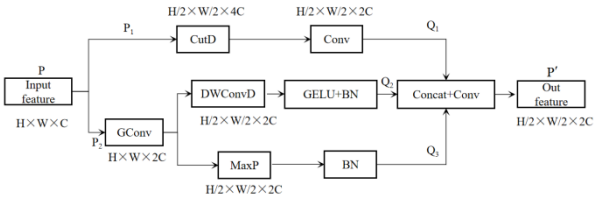

图2　下采样模块IDM结构图

Fig.2　Structure diagram of IDM

IDM 将输入的图像特征$P \in R^{H\times W\times C}$复制为 P_1 和 P_2,其中W、H和C分别表示特征的宽度、高度和通道数量。首先，对 P_1 进行切片下采样，得到C_1、C_2、C_3和C_4 4 个空间降采样后的特征图。切片下采样的过程，如图 3 所示。图中x_{ij}表示 P_1 在空间位置（i，j）处的特征。在通道维度，拼接C_1、C_2、C_3和C_4得到新的特征图，使特征图的通道数量由原来的C增加到$4C$。接着，使用步长为 1 的 1×1 卷积运算将拼接后的特征图通道数量压缩为$2C$，得到特征Q_1。特征图通道数的减半，可以使模型的计算量减小。

$$P_1=\begin{pmatrix} x_{11} & x_{12} & \cdots & \cdots \\ x_{21} & x_{22} & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & x_{(H)(W)} \end{pmatrix}$$

$$C_1=\begin{pmatrix} x_{11} & x_{13} & \cdots & \cdots \\ x_{31} & x_{33} & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & x_{(H-1)(W-1)} \end{pmatrix}_{(\frac{H}{2})(\frac{W}{2})} \quad C_2=\begin{pmatrix} x_{12} & x_{14} & \cdots & \cdots \\ x_{32} & x_{34} & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & x_{(H-1)(W)} \end{pmatrix}_{(\frac{H}{2}\times\frac{W}{2})}$$

$$C_3=\begin{pmatrix} x_{21} & x_{23} & \cdots & \cdots \\ x_{41} & x_{43} & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & x_{(H)(W-1)} \end{pmatrix}_{(\frac{H}{2})(\frac{W}{2})} \quad C_4=\begin{pmatrix} x_{22} & x_{24} & \cdots & \cdots \\ x_{42} & x_{44} & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & x_{(H)(W)} \end{pmatrix}_{(\frac{H}{2}\times\frac{W}{2})}$$

图 3 切片下采样示意图

Fig.3 Schematic diagram of slice downsampling

如图 2 所示，对于图像特征 P_2 采用两个分支进行处理。在其中一个分支，使用步长为 1、尺寸为 3×3 的分组卷积 GConv 处理，然后使用步长为 2 的 3×3 卷积进行下采样，并使用 G+ELU 激活函数和归一化层（Batch Normalization，BN）得到特征Q_2。在另一个分支上，使用与上述结构一致的分组卷积处理 P_2，并做最大池化和归一化处理，得到特征Q_3。

在下采样模块中，三个分支对应的变换公式可以如下表示：

$$Q_1 = \mathrm{Conv}(\mathrm{CutD}(P_1)) \quad (1)$$

$$Q_2 = \mathrm{GELU}(\mathrm{BN}(\mathrm{DWConvD}(\mathrm{GConv}(P_2)))) \quad (2)$$

$$Q_3 = \mathrm{BN}(\mathrm{MaxP}(\mathrm{GConv}(P_1))) \quad (3)$$

式中，Conv、CutD、GELU、BN、DWConvD、GConv、MaxP 分别表示卷积、切片处理、GELU 激活函数、批处理归一化、深度卷积、分组卷积和最大池化操作。

在通道方向上拼接特征Q_1、Q_2和Q_3，并在拼接结果上使用 1×1 卷积层，得到一组通道数翻倍、尺寸减半的特征图 P′。此过程用如下公式表示：

$$P' = \mathrm{Conv}(\mathrm{Concat}(Q_1, Q_2, Q_3)) \quad (4)$$

其中，Concat 表示在通道方向上连接特征。

2.2 基于核选择机制的预测网络

如图 4 所示，为提高模型对不同尺度目标的检测能力，本文采用了核选择模块，根据输入图像的特性动态选择多种不同的卷积核融合特征，从而提高模型的表达能力。

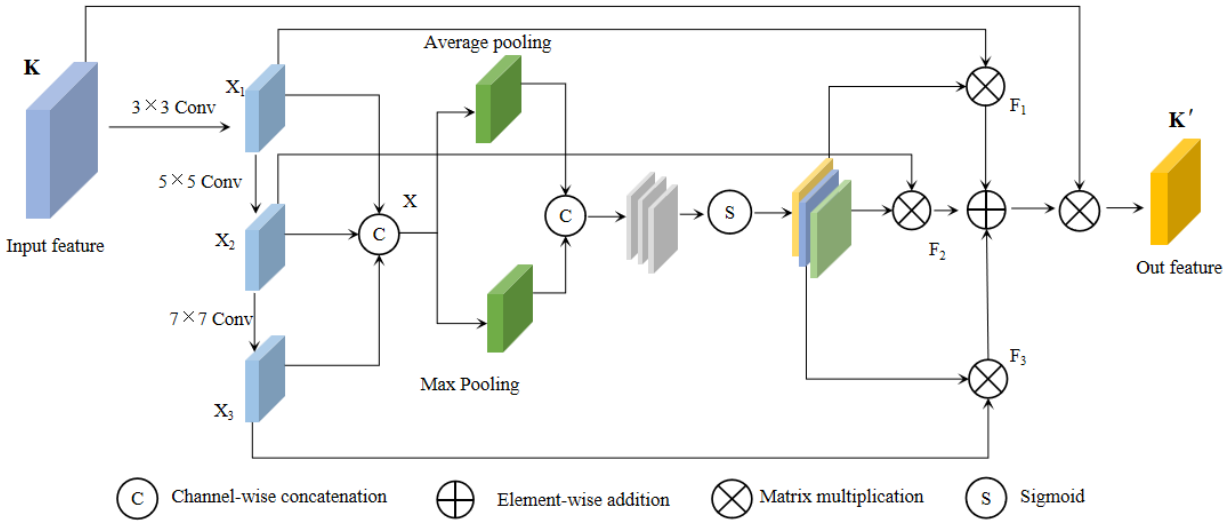

图 4 核选择模块

Fig. 4 Kernel selection module

在目标检测网络的检测任务头中，对于输入的特征 K，利用卷积核大小为 3×3、5×5、7×7 三个空洞卷积来学习多尺度空间信息，得到三个不同尺度感受野的特征图$X_1 \in R^{H\times W\times C}$、$X_2 \in R^{H\times W\times C}$和$X_3 \in R^{H\times W\times C}$，其中：

$$X_2 = \mathrm{DWConv}(X_1) \quad (5)$$

$$X_3 = \mathrm{DWConv}(X_2) \quad (6)$$

式中，DWConv表示空洞卷积。接着，使用通道拼接融合X_1、X_2和X_3以获得具有不同感受野尺寸的特征信息，得到特征X，并在通道方向上拼接X的平均池化和最大池化的结果。然后，相继使用卷积和Sigmoid函数获取独立的空间选择掩码。接着，使用空间选择掩码对X_1、X_2和X_3分别加权，分别得到特征F_1、F_2和F_3。最后，对F_1、F_2和F_3逐元素相加，得到带有注意力的融合特征，并将融合特征和输入特征K进行逐元素相乘，获得特征K'。核选择模块的计算公式如下所示：

$$K' = K * (\sum\nolimits_i (X_i F_i)) \quad (7)$$

3 实验

3.1 DOTA[14]数据集处理

DOTA 是用于目标检测任务的大规模高分辨率航拍图像公共数据集，由 2 806 张大尺寸图像组成，包含了不同尺度、方向和形状的物体。DOTA 包含 15 个对象类别，包括飞机(PL)、棒球场(BD)、桥梁(BR)、田径场(GTF)、小型车辆(SV)、大型车辆(LV)、船舶(SH)、网球场(TC)、篮球场(BC)、储油罐(ST)、足球场(SBF)、环路(RA)、港口(HA)、游泳池(SP)和直升机(HC)。图像的分辨率在 800×800 到 4000×4000 之间。本文以步幅 200 将原图像裁剪成 1 024×1024 大小。将裁剪后的图像在数量上按 2:1 划分，得到训练集和测试集的大小分别为 21 046 和 10 833。测试结果提交至 DOTA 评测服务器。

3.2 实验设置

实验使用一块显存为 24 GB 的 GeForce RTX3090 的显卡训练和测试算法。训练的 batch size 和 epoch 分别设置为 2 和 12 使用 SGD 作为优化器，初始学习率和动量系数分别为 0.0025 和 0.9。采用平均准确率（Average Precision，AP）和全类平均准确率（mean Average Precision，mAP[24]）作为检测评价指标。此外，使用 Params（模型参数的总数）和 Flops（浮点运算次数）[2]衡量模型的计算复杂度和参数数量。

表 1 下采样模块的消融实验对比

Table1 Ablation studies of Downsampling module

Method	Convolutional Downsampling	Slice downsampling	Maxpooling	mAP/%
Baseline				70.88
	√			69.97
		√		69.51
			√	69.76
	√	√		69.34
		√	√	71.11
	√		√	70.81
Ours	√	√	√	**71.63**

3.3 消融实验

本文分析了不同下采样模块对模型的贡献。如表 1 所示。各模块可以使模型的精度得到不同程度的提升，同时使用这三种下采样策略，模型的性能最优。

此外，本文也研究了核组成对实验结果的影响。大尺度感受野的特征图可以直接通过大型卷积核处理或者由多个小型空洞卷积核逐层处理这两种方式获得。如表 2 所示，当卷积运算后均得到感受野尺寸为 29 的特征图时，以三个小型空洞卷积核组合获得大尺度感受野特征图时，模型的计算复杂度最低，参数总量最少。

表 2 不同核组成的实验结果

Table2 Experimental results of different kernel composition

（Kernel size，dilation rate）	Receptive field	mAP/%	Params/M	Flops/M
（29,1）	29	69.34	243.99	40.66
（5,1）+（7,4）	29	70.32	239.71	40.46
（3,1）+（5,2）+（7,3）	29	70.88	239.57	40.43

最后，本文也验证了核选择模块中融合特征的分支数对模型造成的影响。结果如表 3 所示，本文在多种设置下融合不同尺度的感受野特征图。通过对比这些实验结果可以发现，采用 3×3、5×5 和 7×7 组合时，模型表现出最佳的性能。

表 3 网络中不同卷积核设置的实验结果

Table3 Experimental results of different network designs

	3×3	5×5	7×7	9×9	mAP/%
Baseline					68.43
	√				69.95
	√	√			70.45
Ours	√	√	√		**70.88**
	√	√	√	√	70.23

3.4 对比实验

为了验证本文方法的优越性，开展实验对比分析了本文与其它遥感图像目标检测算法。如表格 4 所示，本文达到了 71.63%的 *mAP*，超过其他单阶段和双阶段模型。与基准模型相比，在大型车辆、船舶、海港、环岛等目标类别的检测精度方面明显提高。在对比算法中，表现最好。与 DRN[11]相比，本文算法关于 *mAP* 有 0.93%的提升。实验结果表明，本文提出的模型能够有效提升尺度变化大的物体的检测精度。

为了定性对比基线方法和本文方法的效果，本文从数据集中随机挑选了 3 张图片，测试并可视化。如图 5 所示，相比基线模型，本文算法能更加准确地定位识别海港、大型车辆、环岛等尺度变化大的目标，而基线模型则可能会出现漏检或误检。

表 4　不同算法在 DOTA 数据集上的结果

Tab.4　Result of different algorithms on DOTA dataset

Method	PL	BD	BR	GTF	SV	LV	SH	TC	BC	ST	SBF	RA	HA	SP	HC	mAP/%
ICN[25]	81.36	74.30	**47.70**	70.32	64.89	67.82	69.98	90.76	79.06	78.20	53.64	62.90	**67.02**	64.17	50.23	68.16
RetinaNet(baseline)	**89.41**	76.82	40.91	67.61	**77.51**	62.66	77.54	90.88	82.34	81.99	58.15	61.55	56.46	63.71	38.96	68.43
R3det[10]	88.9	75.25	44.96	66.27	75.16	72.53	79.35	90.88	79.88	83.22	49.42	61.63	63.84	62.84	35.55	68.59
RoI Transformer	88.64	78.52	43.44	**75.92**	68.81	73.68	83.59	90.74	77.27	81.46	58.39	53.54	62.83	58.93	47.67	69.56
CAD-Net[26]	87.83	82.37	49.43	73.51	71.08	63.48	76.59	90.89	79.23	73.35	48.42	60.87	62.14	67.12	62.32	69.91
DRN[11]	88.91	**80.22**	43.52	63.35	73.48	70.69	84.94	90.14	**83.85**	84.11	50.12	58.41	67.62	**68.60**	**52.50**	70.70
Ours	88.43	77.72	46.81	67.96	77.27	**74.8**	**85.37**	**90.91**	78.46	**84.71**	**59.82**	**64.02**	65.59	66.03	46.53	**71.63**

图 5　在 DOTA 数据集上的可视化结果

Fig.5　Visualization results of the proposed algorithm on DOTA dataset

4　结论

针对遥感图像目标检测中存在目标间尺度差异大以及目标方向任意的问题，本文提出了基于改进 RetinaNet 的遥感图像目标检测算法。

1)　通过在 RetinaNet 的特征提取网络 ResNet50 中加入改进的下采样模块来融合不同下采样策略提取的特征图，有效地保留目标特征的边缘信息；在预测网络中采用卷积核选择机制动态选择空间感受野，提升了模型对关键信息的提取能力。

2)　通过在 DOTA 数据集上的实验，本文方法在一定程度上解决了遥感图像中目标方向任意且尺度变化大的问题，并取得了良好的检测效果。

3) 本文算法适用于基于 CNN 的目标检测网络。作者拟在本文研究的基础上进一步基于 Transformer 算法多尺度目标检测。

参考文献：

[1] LIU W, LIU J, LUO Z, et al. Weakly supervised high spatial resolution land cover mapping based on self-training with weighted pseudo-labels[J]. International Journal of Applied Earth Observation and Geoinformation, 2022, 112: 102931.

[2] LIU W, LIN Y, LIU W, et al. An attention-based multiscale transformer network for remote sensing image change detection[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2023, 202: 599-609.

[3] YU Q, LIU W, GONÇALVES W N, et al. Spatial Resolution Enhancement for Large-Scale Land Cover Mapping via Weakly Supervised Deep Learning[J]. Photogrammetric Engineering & Remote Sensing, 2021, 87(6): 405-412.

[4] 张长乐，金钧. 基于深度学习的绝缘子故障检测仿真研究[J]. 华东交通大学学报，2023，40（5）：41-48. ZHANG C L，JIN J. Simulation study on insulator fault detection based on deep learning[J]. Journal of East China Jiaotong University，2023，40（5）：41-48.

[5] LIU Z，HU J，WENG L，et al. Rotated region based CNN for ship detection[C]// Beijing: 2017 IEEE International Conference on Image Processing (ICIP). IEEE， 2017：900-904.

[6] JIANG Y，ZHU X，WANG X，et al. R2CNN：Rotational region CNN for orientation robust scene text detection[J]. arXiv preprint arXiv：1706.09579， 2017.

[7] DING J, XUE N, LONG Y, et al. Learning RoI transformer for oriented object detection in aerial images[C]// California: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 2849-2858.

[8] XIE X, CHENG G, WANG J, et al. Oriented R-CNN for object detection[C]// Seoul: Proceedings of the IEEE/CVF international conference on computer vision. 2021: 3520-3529.

[9] XU Y, FU M, WANG Q, et al. Gliding vertex on the horizontal bounding box for multi-oriented object detection[J]. IEEE transactions on pattern analysis and machine intelligence, 2020, 43(4): 1452-1459.

[10] YANG X, YAN J, FENG Z, et al. R3det: Refined single-stage detector with feature refinement for rotating object[C]// Vancouver: Proceedings of the AAAI conference on artificial intelligence. 2021, 35(4): 3163-3171.

[11] PAN X，REN Y，SHENG K，et al. Dynamic refinement network for oriented and densely packed object detection[C]//Seattle: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020： 11207-11216.

[12] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]//Proceedings of the IEEE international conference on computer vision. 2017: 2980-2988.

[13] HE K， ZHANG X， REN S， et al. Deep residual learning for image recognition[C]//Las Vegas: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016： 770-778.

[14] XIA G S, BAI X, DING J, et al. DOTA: A large-scale dataset for object detection in aerial images[C]// Utah: Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 3974-3983.

[15] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[C]// Advances in neural information processing systems, 2012, 25.

[16] SZEGEDY C，LIU W，JIA Y，et al. Going deeper with convolutions[C]// Hawaii: Proceedings of the IEEE conference on computer vision and pattern recognition. 2015：1-9.

[17] GAO Z, WANG L, WU G. Lip: Local importance-based pooling[C]// Seoul: Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 3355-3364.

[18] STERGIOU A, POPPE R, KALLIATAKIS G. Refining activation downsampling with SoftPool[C]// Seoul:Proceedings of the IEEE/CVF international

conference on computer vision. 2021: 10357-10366.

[19] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]// Utah: Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7132-7141.

[20] LI X, HU X, YANG J. Spatial group-wise enhance: Improving semantic feature learning in convolutional networks[J]. arXiv preprint arXiv:1905.09646, 2019.

[21] CHEN Y, DAI X, LIU M, et al. Dynamic convolution: Attention over convolution kernels[C]// Seattle: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 11030-11039.

[22] LI X，WANG W，HU X，et al. Selective kernel networks[C] // Los Angeles: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019：510-519.

[23] LIU J J，HOU Q，CHENG M M，et al. Improving convolutional networks with self-calibrated convolutions [C]// Seattle: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020：10096-10105.

[24] BHATTACHARYYA A. On a measure of divergence between two statistical populations defined by their probability distribution[J]. Bulletin of the Calcutta Mathematical Society, 1943, 35: 99-110.

[25] AZIMI S M，VIG E，BAHMANYAR R，et al. Towards multiclass object detection in unconstrained remote sensing imagery[C]// Kanagawa: Asian conference on computer vision. Ch-am： Springer International Publishing， 2018： 150-165.

[26] ZHANG G， LU S， ZHANG W. CAD-Net： A context-aware detection network for objects in remote sensing imagery[J]. IEEE Transactions on Geoscience and Remote Sensing，2019，57(12)：10015-10024.

第一作者：程路（2000—），男，硕士研究生，研究方向为目标检测。E-mail：2459650517@qq.com。

通信作者：刘伟（1986—），男，副教授，博士，硕士生导师，研究方向为遥感图像解析、机器学习和计算机视觉。

E-mail：weiliu@ecjtu.edu.cn。