结合数据融合与特征选择的遥感影像尺度多样目标检测

秦登达，万里，何佩恩，张轶，郭亚，陈杰

中南大学 地球科学与信息物理学院, 长沙 410083

摘 要：基于深度神经网络模型的遥感影像地物检测取得了巨大成功，很大程度上得益于大规模数据集的支撑。但是，从现有遥感影像数据集本身来看，不同类别地物的数量分布不一致，同类地物对象以不同尺寸大小呈现，是导致地物样本的尺度不均衡问题的直接因素。对此，本文采用数据集内影像加权融合与地物多尺度特征选择的策略来缓解该问题。首先，将数据集内两张影像的像素值进行加权并得到融合后的影像，从而使不同类别地物样本更加均衡且具有较高的背景多样性；其次，通过选择合适尺度的特征图预测相应尺度的目标类别，且允许同一尺度目标在相邻特征图上进行预测，这样使模型能根据目标尺度进行训练；最后，基于目标中心区域的特征图预测目标边界框，预测的边界框更符合目标本身的尺度。通过在两个遥感数据集上分别进行实验，表明训练的模型在对复杂背景下的类别不均衡目标的识别更加准确，能够适应遥感影像下不同尺度目标的识别。

关键词：图像融合增强，多尺度选择与表达，高分辨率遥感影像，目标检测，卷积神经网络

引用格式：秦登达,万里,何佩恩,张轶,郭亚,陈杰.2022.结合数据融合与特征选择的遥感影像尺度多样目标检测.遥感学报,26(8):1662-1673

Qin D D,Wan L,He P E,Zhang Y,Guo Y and Chen J. 2022. Multiscale object detection in remote sensing image by combining data fusion and feature selection. National Remote Sensing Bulletin, 26(8):1662-1673[DOI:10.11834/jrs.20221249]

1 引 言

光学遥感影像目标检测是确定给定的航空或卫星影像是否包含一个或多个属于兴趣类别的对象，并定位图像中每个预测对象的位置。遥感影像目标检测作为卫星遥感影像理解领域中最基础的任务之一，在军事领域、城市规划（Zhong和Wang，2007）和环境监测（Durieux等，2008）等诸多领域发挥着重要的作用。因此，目标检测与识别任务对于遥感影像解译具有极其重要的研究意义（冯霞 等，2014）。

虽然基于深度学习的目标检测算法取得了瞩目的结果，但还是存在一些问题亟待解决。样本不均衡问题（Shrivastava等，2016；Lin等，2017a；Li等，2019）是当前目标检测研究的热点问题之一，并且有很多相关的研究工作。在多任务遥感影像目标检测中，复杂的影像背景对检测任务造成了许多干扰（Chen等，2020），并且还存在着各种尺度大小不一的检测对象，不同地物目标的尺度都是不同的；此外，有些类别的尺度差别也很大，大的地物目标如田径场其像元个数达几万个，而小的地物目标如车辆只占几十个像元。而模型对于地物尺度的泛化性是有一定限度的，因而这种尺度的多样性和类别差异性给遥感影像目标检测任务带来了极大的挑战。

为了减轻由此带来的负面影响，Pang等（2019）提出了一个平衡学习目标检测框架Libra R-CNN。它集成了3个新的组成部分：IoU均衡采样、均衡特征金字塔和均衡L1损失函数，分别用于减少样本、特征和目标层次上的不均衡。得益于整体平衡设计，Libra R-CNN显著提高了检测性能。在线难例挖掘算法（OHEM）选择损失最大的

收稿日期：2021-04-29；**预印本**：2021-09-23

基金项目：国家重点研发计划(编号:2020YFA0713503)；国家自然科学基金(编号:42071427)；湖南省自然科学基金(编号:2020JJ4691)

第一作者简介：秦登达,研究方向为遥感影像目标检测。E-mail：qindengda@csu.edu.cn

通信作者简介：陈杰,研究方向为遥感影像智能分析与理解。E-mail：cj2011@csu.edu.cn

一些样本作为训练的样本从而改善因为样本数目不平衡而导致检测效果差的问题（Shrivastava等，2016）。Cao等（2020）提出了一种称为“主要样本注意”（PISA）的抽样和学习策略，它将训练过程的重点指向重要样本，在训练检测器时，聚焦于原始样本通常比聚焦于“难例”更有效。图像金字塔尺度归一化（SNIP）训练方案根据图像尺度的变化有选择地反向传播不同大小目标实例的梯度（Singh和Davis，2018），其核心思想是输入多尺度图像，只在该尺度图像下合适尺寸的目标样本尺寸进行训练。

光学遥感影像存在着大量不同尺度和小样本目标，以及各种复杂的背景（姚红革 等，2020）。多尺度特征融合可以有效提高小目标和不同目标的检测效果（Li等，2020）。当前基于卷积神经网络的目标检测算法可以分为两大类：其一，是把检测分为区域建议和分类回归两阶段算法，这类算法以RCNN系列（Girshick等，2014；Girshick，2015；Ren等，2017；Lin等，2017b；Cai和Vasconcelos，2018）为代表；其二，是一次性直接完成目标框回归和目标分类的单阶段算法，这类似算法以SSD（Liu等，2016）和YOLO（Redmon等，2016；Redmon和Farhadi，2017，2018）等算法为代表，相关算法在遥感上都有较多应用（江一帆 等，2020；王冰 等，2021；杨耘 等，2021）。Girshick等（2014）在2014年结合卷积神经网络提出了RCNN网络，该网络取代了传统目标检测方法。Faster RCNN（Ren等，2017）抛弃了选择性搜索算法生成候选框，而采用了RPN网络进行候选框筛选提升了检测效率和检测性能。SSD（Liu等，2016）算法通过将VGG16（Simonyan和Zisserman，2015）的多个不同尺寸特征图共同进行目标框的回归进行不同尺度的预测，最终在小目标的预测精度优于同年的YOLO（Redmon等，2016）。特征金字塔网络（FPN）网络提出了特征层融合结构（Lin等，2017b），该结构能有效提取图片的不同尺度特征信息。由于遥感影像本身存在着各种尺度的目标，多尺度融合结构在遥感目标检测取得了优秀的效果，同时该结构成为最为常用的多尺度特征提取网络。RetinaNet（Lin等，2017a）模型则采用FPN作为特征提取网络，提出Focal Loss来减轻正负样本对精度的影响，PaNet（Liu等，2018）则在FPN（Lin等，2017b）的基础上新增了一个自底向上的融合结构。于野等人在FPN的基础上融入特征的显著性图提出A-FPN（于野 等，2020）以提高浅层特征的特征表达。虽然以上多尺度方法在遥感影像上能够顾及不同尺度的目标信息，但在每一个尺度特征层上都对各尺寸的目标进行识别，而不同尺度的特征层并不是对每一种尺度的目标信息都非常清晰。所以，采用FoveaBox（Kong等，2020）在遥感影像上根据不同目标尺寸在不同的尺度特征图上进行目标识别。

针对样本类别不均衡的问题，提出了解决思路。首先，为了解决样本数目不均衡的问题，本文提出一种基于图像融合的数据增强策略，通过将两张图像融合为一张新的图像实现数据增强。由于这是针对数据层面上的处理，可以应用于任何基于深度学习的目标检测模型。考虑到光学遥感影像的特点，并且基于多尺度特征表达与选择的目标检测的策略（Kong等，2020）更加适合遥感影像目标检测，因此将该方法应用于光学遥感影像目标检测中。其次，将影像融合与多尺度特征表达与选择的目标检测进行结合，能减轻复杂背景和类别不均衡的影响。通过在两个开源数据集上验证了该方法的有效性和普适性。

2 方法原理

基于多尺度特征选择与表达的模型结合图像融合的方法对高分光学遥感影像进行目标检测。结合数据融合与特征选择的遥感影像尺度多样目标检测流程图如图1所示：首先，将用于训练的数据集进行图像融合增强，使得训练数据中不同类别更加均衡；其次，在模型训练时，训练图片先经过特征金字塔（FPN）提取5个不同尺度的特征，5个层次的特征分别预测不同尺度范围的地物目标；最后，进行类别预测与地物目标中心特征的边界框的训练和预测。

2.1 增加类别均衡性

高分遥感影像包含了丰富的地物目标和细节信息，同时影像丰富的信息对于感兴趣地物带来许多背景信息的干扰。地物目标提取的特征是否具有代表性是影响模型性能的一方面因素（Pang等，2019）。并且，地物目标自身的存在的差异性在影像上出现的概率都不尽相同，导致制作的数据集中不同类别的目标图片数量存在差异。模型

训练过程中会由于训练数据类别的不均衡而使得各类别图片训练的比重不同，这种各类别影像数量的失衡使得模型更注重于数量多的影像，而降低了对影像数量较少类别检测的敏感性，最终性能偏向于影像数量多的类别。

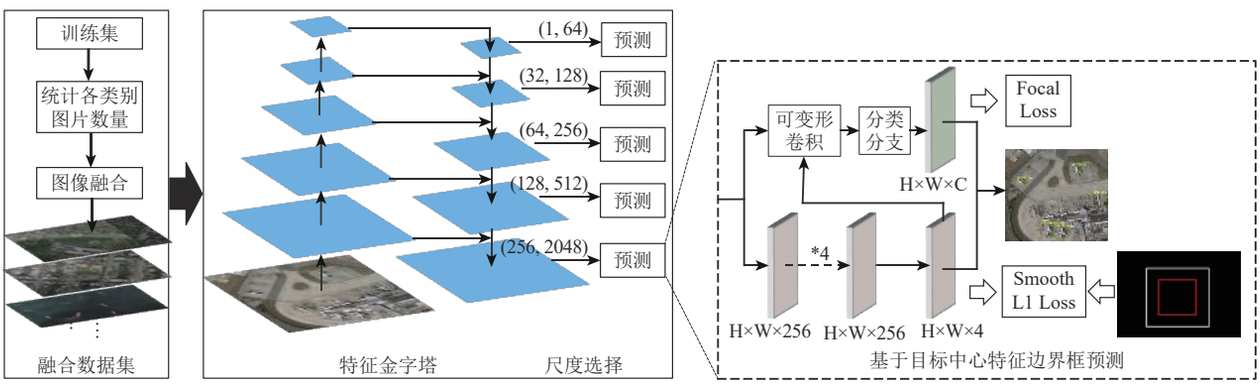


图1 算法流程图

Fig. 1 The flow chart of our method

针对上述问题，通过提出影像融合增强来削弱类别失衡和复杂背景的影响。影像融合增强通过将需要增强的训练影像与不包含任何目标的背景影像按照系数θ进行两者的融合。首先，对待增强影像与背景影像的比例进行统计，以最大长、宽作为融合影像的尺寸；其次，将目标影像像素矩阵乘以系数θ，将背景影像像素矩阵乘上系数1−θ；最后，以融合影像的长宽为基准，将得到的新的影像矩阵和新的背景矩阵赋值给融合影像，其中重合的部分则取两者之和进行目标影像与背景影像的融合得到最终融合影像。影像融合的公式可以由如下表示：

$$V = \boldsymbol{zero}(h, w, 3) \oplus \boldsymbol{I} \times \theta \oplus \boldsymbol{P}_k \times (1 - \theta) \quad (1)$$

式中，h、w代表待增强影像和背景影像的最大长宽；$\boldsymbol{I}$是待增强的影像；$\boldsymbol{P}_k$为背景影像；θ表示在［0，1］之间的系数；V表示最终影像融合结果；$\oplus$表示矩阵按对应坐标相加；×表示矩阵和数相乘。

通过上述方法进行的影像融合，在尺寸上会存在3种情况，即融合后的图像尺寸大于待增强影像、等于待增强影像以及小于待增强影像。对于大于待增强影像尺寸的情况，根据式（1）可知待增强影像目标区域的绝对坐标是没有改变的；对于大于待增强影像尺寸的情况，待增强影像目标区域的绝对坐标显然是没有改变的；同样对于小于待增强影像尺寸的情况，待增强影像目标区域的绝对坐标也是没有改变的。因此，融合后的图像标签依然可使用待增强图像I的标签。融合后的影像如图2所示。其中，图2（a）是原始影像，图2（b）、（c）、（d）分别为3张不同的背景影像；影像图2（e）、（f）、（g）分别为利用3张不同的背景图像进行融合后的结果。其结果表明，图2（e）、（f）、（g）在保留了原始地物目标情况下，场景也变得更加多样和丰富，从而在对数据样本进行扩充的同时，达到增强样本场景的多样性和模型训练后的鲁棒性。

针对不同的数据集，影像融合增强的目标类别是不同的，对于NWPUVHR-10数据集（Cheng等，2014，2016），增强的类别有：船只、棒球场、网球场、篮球场、港口、油桶、桥梁和车辆，这些类别的目标数相对较少。而对于RSOD（Xiao等，2015；Long等，2017）数据集，由于数据集类别只有4类，所以4个类别的训练数据都有增强，两个数据集根据8∶2划分为训练集和测试集，影像融合只对训练数据集进行操作，后续实验基于原始数据集抽取的测试集进行精度测试。两个数据集图像融合前后的数量对比如图3所示。其中，图3（a）表示RSOD数据集影像融合前后数据分布；图3（b）为NWPUVHR-10数据集增强前后各类别数量分布；通过影像融合后的两个数据集各类别图片数量相比于原始训练集更加均衡，更利于各类别图片的训练。

（a）原始影像

（a）Original image

（b）背景影像Ⅰ

（b）Background image Ⅰ

（c）背景影像Ⅱ

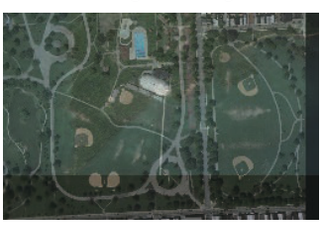
（c）Background image Ⅱ

（d）背景影像Ⅲ

（d）Background three image Ⅲ

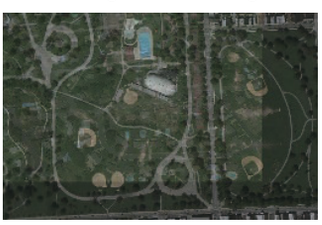
（e）融合结果Ⅰ

（e）Fusion result Ⅰ



（f）融合结果Ⅱ

（f）Fusion result Ⅱ

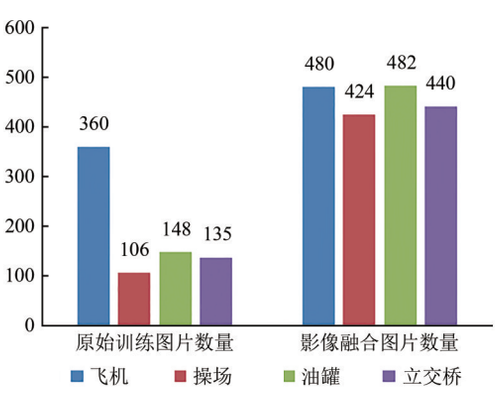


（g）融合结果Ⅲ

（g）Fusion result Ⅲ

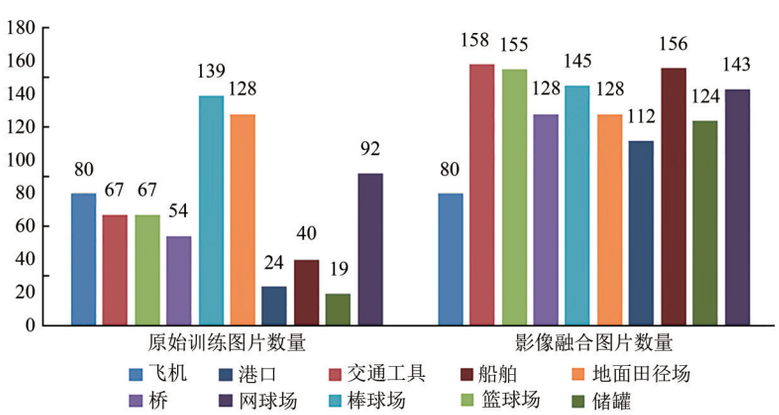
图2 影像融合前后示意图

Fig. 2 Diagram before and after image fusion



（a）RSOD数据集影像融合前后数据分布

（a）Distribution of data before and after image fusion for RSOD dataset



（b）NWPUVHR-10数据集增强前后各类别数量分布

（b）Distribution of data before and after image fusion for NWPUVHR-10 dataset

图3 影像融合前后训练集各类图片数量分布

Fig. 3 Image quantity distribution of training set before and after image fusion

2.2 多尺度特征表达与选择

尺度多样性一直是遥感影像目标检测亟待解决的问题。当前针对多尺度遥感影像目标检测常用的策略有两个方面：首先，FPN（Lin等，2017b）提出了多尺度特征预测，利用多种尺度的特征图分别进行预测。然而，大尺度的目标通常是在FPN（Lin等，2017b）的深层特征层中预测的，因此这些目标的边界可能过于模糊，无法获得准确的位置，而小尺度特征则是在浅层特征进行预测的，语义信息较少，不足以识别目标的类别。其次，Faster RCNN（Ren等，2017）通过事先设置大量的锚框。最后，利用这些锚框进行目标框的预测，而锚框的尺度设置要尽可能的覆盖数据集各个目标尺度范围，所以锚框的设置很难符合遥感影像中各种尺度的地物目标，最终影响影像的检测效果。

基于上面两点，在光学遥感影像上采用一种多尺度特征选择的训练方式和一种基于目标中心预测候选框的方法（Kong等，2020）。多尺度特征选择通过利用合适尺度的特征图来预测相应尺度的目标类别，同时允许同一尺度目标在相邻的特征图上进行预测，使得特征图可以根据目标的尺度来更好地训练。由于锚框的设置会影像检测效果，因此直接利用目标中心区域的特征预测目标的边界框，其预测的边界框框尺度更符合实际情况。

遥感影像中存在着众多尺度不一的地物目标，选择FPN特征提取出来的多个级别的特征图P_i（i=3，4，…，7）进行预测，每个级别的特征图的长宽依次增加一倍，这5个级别的特征图分别并行地进行预测。为将合适尺度的特征图来预测相应尺度的目标，根据5个尺度大小的特征图将其划分5个尺度的目标大小范围，这5个尺度的地物目标范围的并集会包含数据集所有地物目标的尺度范围。如图1所示，5种不同尺度的特征图负责预测相应尺度等级的目标，并且各个尺度范围有一定的重叠度。具体地，根据数据集中训练目标的大致尺度范围，划分每个级别特征图预测的目标尺度范围；同时，各尺度区间范围之间有一定的重合，使得同一目标能在相邻尺度特征下进行预测。尺度范围的划分可以由$[S_i/2, S_i\cdot 2]$表示，其中S_i表示特征图P_i的基础像素面积，其值可以由如下公式计算：

$$S_i = 4^i \cdot S \tag{2}$$

式中，4^i表示的是每个级别的特征图面积相差大小，S为最小特征图的面积大小。

以上过程划分了各个特征图所负责预测的尺度范围，在训练时网络忽略那些目标大小在相应尺度范围之外的实例，由于最终划分的尺度区间包含了数据集中各类目标的所有尺度，因此一个目标至少会在一个层次的特征图上进行预测。

2.3 基于目标中心区域的边界框预测

在Faster RCNN（Ren等，2017）中，通过人为设置9种固定尺度的锚框，然后训练这些锚框偏置值使预测框尽可能接近真实的标注框。然而，人为设置的锚框并不能很好的和真实框相吻合，也不利于后续的训练。因此，采用目标中心区域的特征进行目标边界框和目标类别预测，可以获取任意尺度的候选框。并且，预测结果是根据目标特征得到，预测的边界框会与真实的标注框会更加吻合，从而更有利于后续的训练。模型对于结果的训练和预测并不是基于目标中心点，而是基于目标中心一定范围区域的特征进行预测。图4为基于目标中心区域的候选框预测示意图，其中红色框表示真实的标注框，黄色框表示根据真实框进行训练和预测的范围框。中心区域的训练范围可以由目标检测数据集中训练图片的标注框形状和位置确定。首先将真实框映射到各个级别的特征图P_i中，并且确定真实框中心在原图的位置，该过程可以由如下公式表示：

$$\begin{aligned} x_{f1} &= x_1/2^i, y_{f1} = y_1/2^i \\ x_{f2} &= x_2/2^i, y_{f2} = y_2/2^i \\ c_x &= x_{f1} + 0.5(x_{f2} - x_{f1}) \\ c_y &= y_{f1} + 0.5(y_{f2} - y_{f1}) \end{aligned} \tag{3}$$

式中，x_1、y_1、x_2、y_2表示真实框在原图上的两个顶点坐标，x_{f1}、y_{f1}、x_{f2}、y_{f2}表示真实框映射到特征图上的两个顶点坐标，2^i表示特征图下采样步长，c_x和c_y表示真实框映射到特征图上的中心点坐标。

得到中心点坐标后，据此获取目标中心区域范围（x_{p1}，y_{p1}，x_{p2}，y_{p2}），此区域的特征将用来进行候选框的训练和预测，其过程可以由如下公式表示：

$$\begin{aligned} x_{p1} &= c_x - 0.5(x_{f2} - x_{f1})\mu \\ x_{p2} &= c_x + 0.5(x_{f2} - x_{f1})\mu \\ y_{p1} &= c_y - 0.5(y_{f2} - y_{f1})\mu \\ y_{p2} &= c_y + 0.5(y_{f2} - y_{f1})\mu \end{aligned} \tag{4}$$

式中，x_{p1}，y_{p1}，x_{p2}，y_{p2}表示用于预测的特征范围的左上角和右下角坐标，μ是一个控制这个区域大小的参数，当μ大于1时，预测区域会大于真实框区域，当μ小于1时，预测区域会小于真实框。由于真实框是目标的外接矩形框，所以会包含一些背景信息。模型使用目标中心区域的特征来进行训练和预测，不仅可以提高准确率，也可以提高模型对地物目标提取的特征表达能力，因此μ的设置会小于1，即训练区域会小于真实框。

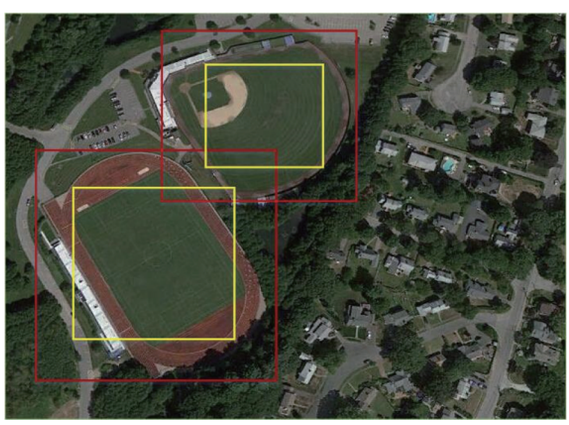


图4 目标中心区域的边界框预测示意图

Fig. 4 Diagram of bounding box prediction for object center area

3 实 验

3.1 实验数据集及评价指标

文中的方法主要在两个具有挑战性的公开遥感影像目标检测数据集上评估所提出的方法。分别是RSOD-Dataset和NWPUVHR10-Dataset。

（1）RSOD-Dataset（Xiao等，2015；Long等，2017）是由武汉大学团队标注，包含飞机、操场、立交桥、油桶4类目标。

（2）NWPUVHR10-Dataset（Cheng等，2014，2016a，2016b）是由西北工业大学团队标注，共包含10类目标，这10类物体分别是飞机、轮船、储罐、棒球场、网球场、篮球场、地面田径场、港口、桥梁和车辆。这些图像是从谷歌地球和瓦辛根数据集中裁剪出来的，然后由专家手工标注。

实验采用平均查准率（AP）和平均准确度（mAP）这两个常用的评价指标评估模型在上述两种数据集上的效果。平均查准率是指精度和召回率曲线下的面积，它是一种结合了精度和召回率的度量；平均准确度是多类别平均查准率的平均值，它是评价多类目标检测最重要的指标。这两个指标越大越好。召回率（Recall）是测试集所有正样本样例中，被正确识别为正样本的比例，其表达式为：

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{5}$$

准确度（Precision）指预测为正样本是正样本所占的比例，其表达式为：

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{6}$$

式中，TP表示被正确地划分成正例的个数，FP表示被错误地划分为正例的个数，FN表示被错误地划分为负例的个数，即实际为正例但被分类器划分为负例的数量。

3.2 类别均衡性实验设置

类别均衡实验在RSOD和NWPUVHR10数据集上对比了OHEM（Shrivastava等，2016）、Libra R-CNN（Pang等，2019）、旋转和翻转常规增强（表中名称为FoveaBox & aug）几种方法。两个数据集以8∶2的比例随机分为训练集和测试集，其中，模型的精度评价基于未使用影像融合的测试集。在RSOD数据集上所有模型采用1000×900的图片尺寸训练，NWPUVHR10数据集训练和测试尺寸为1024×512，并且两个数据集都进行120个epoch的训练，初始学习率为0.01，分别在40、80、100个epoch进行衰减率为0.1的学习率衰减。

3.3 多尺度特征选择与表达实验设置

多尺度特征选择与表达的实验对比同时在RSOD和NWPUVHR10数据集上对比Faster RCNN（Ren等，2017）、SSD300（Liu等，2016）、FPN（Lin等，2017b）、RetinaNet（Lin等，2017a）、FoveaBox（Kong等，2020）方法。训练集和数据集RSOD数据集的训练和测试的尺寸为1000×900，NWPUVHR10数据集使用1024×512进行模型的训练和测试，而SSD300的训练图片尺寸为300×300。其他模型设置均采用最大训练120个迭代次数，并

且设置0.001为初始学习率，学习率在训练中衰减3次，分别在40、80、100次迭代时学习率衰减为原来的学习率的0.1倍。RetinaNet训练与测试时的网络设置与FPN网络一致。多尺度特征选择与表达的模型设置与上述网络一致，网络中控制训练区域的参数μ设置为0.4。特征提取网络都采用ImageNet的预训练模型进行训练。

3.4 定量分析

为验证影像融合增强对结果的有效性，将文中使用的影像融合和特征选择的多尺度网络分别在RSOD和NWPUVHR10两个数据集上进行实验对比。值得注意的是，用于模型的训练数据和测试数据集是以8∶2的比例从整体数据集中随机抽取，并且只对训练数据进行影像融合增强。相关模型的精度值见表1、表2。从表1的RSOD数据集中精度对比可以看出：结合数据融合与特征选择多尺度方法相较于在线难例挖掘和平衡学习目标检测框架的方法分别有2.69%和2.38%的平均准确度的提升；且影像融合增强方法相较于旋转、翻转的常规增强方法有0.88%的平均精度优势。在表2的NWPUVHR10数据集上的精度表明：结合数据融合与特征选择多尺度方法比另两种均衡方法性能更具优势；且相对于旋转、翻转常规数据增强平均精度提升了3.96%。两个数据集的结果说明影像数据融合增强对网络性能有较强的促进作用，使得模型具有更好的性能与泛化能力。

表1 RSOD数据集类别均衡实验的AP50精度对比

Table 1 The AP50 accuracy comparison for category balance experiment in RSOD dataset

/%

类别	方法			
	OHEM（Shrivastava等，2016）	Libra R-CNN（Pang等，2019）	FoveaBox & aug	FoveaBox &fusion
Aircraft	81.49	81.60	**89.67**	89.55
Overpass	87.47	**89.08**	85.47	88.92
Playground	98.55	97.65	99.41	**99.79**
Oiltank	90.33	**90.76**	90.53	90.35
平均精度mAP	89.46	89.77	91.27	**92.15**

注：加黑代表对应行中精度最高。

表2 NWPUVHR10数据集类别均衡实验的AP50精度对比

Table 2 The AP50 accuracy comparison for category balance experiment in NWPUVHR10 dataset

/%

类别	方法			
	OHEM（Shrivastava等，2016）	Libra R-CNN（Pang等，2019）	FoveaBox & aug	FoveaBox &fusion
Airplane	98.66	**99.99**	**99.99**	99.32
Ship	75.15	71.59	80.42	**95.75**
oil tank	84.47	91.61	90.92	**92.56**
BD	98.30	**99.78**	99.01	98.95
TC	93.35	95.65	96.94	**99.82**
BC	96.35	98.64	94.60	**100**
GTF	99.92	**100**	**100**	**100**
Harbor	95.56	96.15	**100**	**100**
Bridge	83.09	84.00	87.65	**92.00**
Vehicle	68.80	68.86	84.20	**94.92**
平均精度mAP	89.37	90.73	93.37	**97.33**

注：加黑代表对应行中精度最高。

为验证影像融合和特征选择的多尺度网络在遥感影像上的有效性，分别在RSOD和NWPUVHR10两个数据集上进行实验对比。如表3所示，RSOD数据集中的精度表明基于影像融合和特征选择的多尺度网络整体性能更加优秀。虽然对比于未进行融合的多尺度特征选择与表达模型只提升了0.12%，但由于RSOD数据集中只包含有4个类别，训练和预测过程比大型数据集更容易。而且，每个类别的可用的训练图像数量比例相差不大，所以在多尺度特征选择模型的训练和预测时并没有很好的体现图像融合的优势。

表3 RSOD数据集AP50精度对比

Table 3 The AP50 accuracy comparison in RSOD dataset

/%

数据集	方法					
	Faster RCNN（Ren等，2017）	SSD300（Liu等，2016）	FPN（Lin等，2017b）	RetinaNet（Lin等，2017a）	FoveaBox（Kong等，2020）	FoveaBox &fusion
Aircraft	90.02	77.07	81.59	86.88	89.30	**89.55**
Overpass	66.32	80.42	79.57	81.98	**89.21**	88.92
Playground	89.32	**100**	**100**	99.79	99.01	99.79
Oiltank	90.25	90.27	90.45	90.61	**90.59**	90.35
平均精度mAP	83.98	86.94	87.90	89.81	92.03	**92.15**

注：加黑代表对应行中精度最高。

表4中NWPUVHR10数据集各类别识别精度可以看出：基于影像融合和特征选择的多尺度网络对比于其他几种主流方法精度有显著提升，并且经过影像融合增强的船只、棒球场、网球场、篮球场、港口、桥梁和车辆等这些类别在精度上有较大提升，达到了几种方法中最好的精度。整体表明特征选择与表达的网络在包含了各种尺度大小目标的遥感影像下的地物识别能取得较高的精度。影像融合增强能够一定程度消除训练数据中类别不均衡的问题，几种典型的目标检测网络的数据融合增强对比可以发现图像融合增强的策略具有更强的普适性，对模型的性能以及鲁棒性都有一定的提升。

表4 NWPUVHR10数据集AP50精度对比

Table 4 The AP50 accuracy comparison in NWPUVHR10 dataset

/%

数据集	方法					
	Faster RCNN（Ren等，2017）	SSD300（Liu等，2016）	FPN（Lin等，2017b）	RetinaNet（Lin等，2017a）	FoveaBox（Kong等，2020）	FoveaBox &fusion
Airplane	97.83	98.35	99.33	99.32	**100**	99.32
Ship	78.66	71.02	71.65	72.41	89.58	**95.75**
oil tank	90.68	80.35	87.12	79.63	**93.02**	92.56
BD	89.99	88.40	97.98	99.10	98.36	**98.95**
TC	80.85	89.27	94.28	94.07	94.08	**99.82**
BC	58.80	70.91	83.45	96.43	97.05	**100**
GTF	95.47	99.45	**100**	**100**	**100**	**100**
Harbor	80.68	85.94	**100**	98.79	92.76	**100**
Bridge	63.33	63.92	**92.00**	87.83	71.58	**92.00**
Vehicle	73.09	54.61	79.08	74.04	89.07	**94.92**
平均精度mAP	80.94	80.22	90.49	90.16	92.55	**97.33**

注：加黑代表对应行中精度最高。

3.5 定性分析

图5显示的是RSOD数据集上不同模型的可视化结果，图6是NWPU VHR-10预测的可视化结果。图5中可以看到，RetinaNet模型对于排列复杂密集的飞机影像识别效果不理想，基于影像融合和特征选择的多尺度网络的方式对复杂背景下的4种地物类别有更好的识别效果，并且具有更少的误检框，说明该方式应用于遥感影像能具有比较好的鲁棒性和性能优势。

（a）真实标签

（a）Ground truth

（b）FPN网络预测结果

（b）FPN network prediction results

（c）RetinaNet预测结果

（c）RetinaNet prediction results

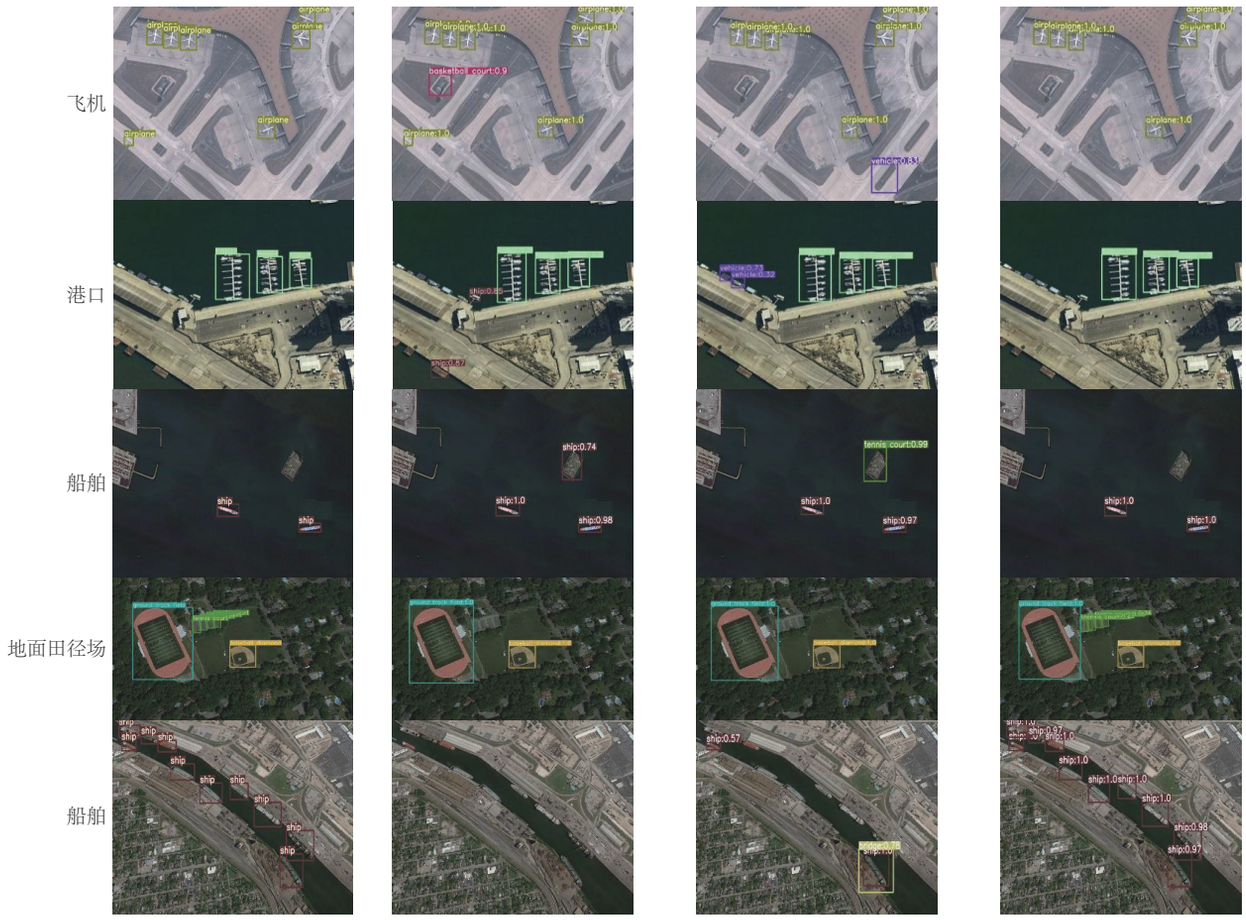
（d）基于影像融合和特征选择的多尺度网络的结果

（d）The results of FoveaBox and fusion

图5 RSOD数据集测试结果可视化

Fig. 5 Visualization of test results in RSOD dataset

从图6的可视化结果可以看出，使用图像融合增强的多尺度选择与表达的模型后，在飞机类别、船舶、海港、田径场不同尺度大小的目标上，相对于其他的多尺度网络有更好的识别效果。对于田径场相对大尺度场景下，另外两个方法难以识别出更小的网球场；在岸边包含船舶的影像上，FPN以及RetinaNet很难将河岸和船舶很好地区分开（图6第5行），而采用基于影像融合和特征选择的多尺度网络的方法对复杂背景下的目标的识别也相对更加准确，说明图像融合增强了样本场景的多样性，并且模型结果整体表明在光学遥感影像中不同尺度的目标都能够合理的预测出来。



（a）真实标签 （b）FPN网络预测结果 （c）RetinaNet预测结果 （d）基于影像融合和特征选择的多尺度网络的结果

（a）Ground Truth （b）FPN network prediction results （c）RetinaNet prediction results （d）The results of FoveaBox and fusion

图6 NWPU VHR-10数据集测试结果可视化

Fig. 6 Visualization of test results in NWPU VHR-10 dataset

4 结 论

本文将多尺度特征选择的模型应用在了遥感影像上，通过多尺度特征的表达与选择能更加契合于复杂背景下遥感数据的不同尺度大小的目标。而且，提出了图像融合增强的策略。相较于之前的数据增强方式，文中提出的方法具有以下几点优势：（1）不会改变图像的现状大小以及目标的绝对位置。（2）由于采用的是同一样本库的图像进行融合，因此不会改变样本库的分布。（3）多尺度特征表达与选择和图像增强融合可以应对遥感影像中相对复杂背景的影像，减轻类别不均衡的影响，更加符合遥感影像使用的场景。

遥感影像的俯视成像使得影像中的目标具有密集且方向任意的特点，这些特点对目标检测的性能同样存在影响。但在本文中还未结合影像中目标的这些特点。在未来的研究中，将从卷积神经网络的特征提取的特性出发，结合更多遥感影像中目标的特性，完善高分遥感目标检测模型。

参考文献（References）

Cai Z W and Vasconcelos N. 2018. Cascade R-CNN: Delving into high quality object detection//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT: IEEE: 6154-6162 [DOI: 10.1109/CVPR.2018.00644]

Cao Y H, Chen K, Loy C C and Lin D. 2020. Prime sample attention in object detection//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA: IEEE: 11580-11588 [DOI: 10.1109/CVPR42600.2020.01160]

Chen J, Wan L, Zhu J R, Xu G and Deng M. 2020. Multi-scale spatial

and channel-wise attention for improving object detection in remote sensing imagery. IEEE Geoscience and Remote Sensing Letters, 17(4): 681-685 [DOI: 10.1109/LGRS.2019.2930462]

Cheng G and Han J W. 2016a. A survey on object detection in optical remote sensing images. ISPRS Journal of Photogrammetry and Remote Sensing, 117: 11-28 [DOI: 10.1016/j.isprsjprs.2016.03.014]

Cheng G, Han J W, Zhou P C and Guo L. 2014. Multi-class geospatial object detection and geographic image classification based on collection of part detectors. ISPRS Journal of Photogrammetry and Remote Sensing, 98: 119-132 [DOI: 10.1016/j. isprsjprs. 2014. 10.002]

Cheng G, Zhou P C and Han J W. 2016b. Learning Rotation-Invariant convolutional neural networks for object detection in VHR optical remote sensing images. IEEE Transactions on Geoscience and Remote Sensing, 54(12): 7405-7415 [DOI: 10.1109/TGRS. 2016. 2601622]

Durieux L, Lagabrielle E and Nelson A. 2008. A method for monitoring building construction in urban sprawl areas using object-based analysis of Spot 5 images and existing GIS data. ISPRS Journal of Photogrammetry and Remote Sensing, 63(4): 399-408 [DOI: 10.1016/j.isprsjprs.2008.01.005]

Feng X, Qiu K, Cui W H, Chen Y X and Li X H. 2014. Multiscale description and recognition of target shape in high-resolution remote sensing images. Journal of Remote Sensing, 18(1): 90-104 (冯霞, 秦昆, 崔卫红, 陈一祥, 李向辉. 2014. 高分辨率遥感影像目标形状特征多尺度描述与识别. 遥感学报, 18(1): 90-104) [DOI: 10.11834/jrs.20133056]

Girshick R, Donahue J, Darrell T and Malik J. 2014. Rich feature hierarchies for accurate object detection and semantic segmentation// 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH: IEEE: 580-587 [DOI: 10.1109/CVPR. 2014.81]

Girshick R. 2015. Fast R-CNN//2015 IEEE International Conference on Computer Vision (ICCV). Santiago: IEEE: 1440-1448 [DOI: 10.1109/ICCV.2015.169]

Jiang Y F, Yu H Y, Li C L, Liu P and Zhang H Y. 2020. Oil storage tank detection and information extraction from high-resolution remote sensing imagery based on improved Faster R-CNN. Remote Sensing Information, 35(4): 89-96 (江一帆, 于海洋, 李朝亮, 刘鹏, 张慧勇. 2020. 基于改进faster R-CNN的高分遥感影像储油罐检测与信息提取. 遥感信息, 35(4): 89-96) [DOI: 10.3969/j.issn.1000-3177.2020.04.013]

Kong T, Sun F C, Liu H P, Jiang Y N, Li L and Shi J B. 2020. FoveaBox: Beyound anchor-based object detection. IEEE Transactions on Image Processing, 29: 7389-7398 [DOI: 10.1109/TIP. 2020. 3002345]

Li B Y, Liu Y and Wang X G. 2019. Gradient harmonized single-stage detector. Proceedings of the AAAI Conference on Artificial Intelligence, 33(1): 8577-8584 [DOI: 10.1609/aaai.v33i01.33018577]

Li C Y, Liu J, Hong H L, Mao W J, Wang C J, Hu C D, Su X and Luo B. 2020. Object detection based on OcSaFPN in aerial images with noise. arXiv:2012.09859

Lin T Y, Dollár P, Girshick R, He K M, Hariharan B and Belongie S. 2017b. Feature pyramid networks for object detection//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI: IEEE: 936-944 [DOI: 10.1109/CVPR.2017.106]

Lin T Y, Goyal P, Girshick R, He K M and Dollár P. 2017a. Focal loss for dense object detection//2017 IEEE International Conference on Computer Vision (ICCV). Venice: IEEE: 2999-3007 [DOI: 10.1109/ICCV.2017.324]

Liu S, Qi L, Qin H F, Shi J P and Jia J Y. 2018. Path aggregation network for instance segmentation//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT: IEEE: 8759-8768 [DOI: 10.1109/CVPR.2018.00913]

Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C Y and Berg A C. 2016. SSD: single shot MultiBox detector//14th European Conference on Computer Vision. Amsterdam: Springer: 21-37 [DOI: 10.1007/978-3-319-46448-0_2]

Long Y, Gong Y P, Xiao Z F and Liu Q. 2017. Accurate object localization in remote sensing images based on convolutional neural networks. IEEE Transactions on Geoscience and Remote Sensing, 55 (5): 2486-2498 [DOI: 10.1109/TGRS.2016.2645610]

Pang J M, Chen K, Shi J P, Feng H J, Ouyang W L and Lin D H. 2019. Libra R-CNN: towards balanced learning for object detection// 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA: IEEE: 821-830 [DOI: 10.1109/CVPR.2019.00091]

Redmon J, Divvala S, Girshick R and Farhadi A. 2016. You only look once: unified, real-time object detection//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV: IEEE: 779-788 [DOI: 10.1109/CVPR.2016.91]

Redmon J and Farhadi A. 2017. YOLO9000: better, faster, stronger// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI: IEEE: 6517-6525 [DOI: 10.1109/CVPR.2017.690]

Redmon J and Farhadi A. 2018. YOLOv3: an incremental improvement. arXiv:1804.02767

Ren S Q, He K M, Girshick R and Sun J. 2017. Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(6): 1137-1149 [DOI: 10.1109/TPAMI.2016.2577031]

Shrivastava A, Gupta A and Girshick R. 2016. Training region-based object detectors with online hard example mining//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV: IEEE: 761-769 [DOI: 10.1109/CVPR.2016.89]

Simonyan K and Zisserman A. 2015. Very deep convolutional networks for Large-Scale image recognition. arXiv:1409.1556

Singh B and Davis L S. 2018. An analysis of scale invariance in object detection - SNIP//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT: IEEE: 3578-

3587 [DOI: 10.1109/CVPR.2018.00377]

Wang B, Zhou Y, Zhang H N and Zhao K. 2021. Aircraft detection method based on SSD framework for remote sensing images. Fire Control and Command Control, 46(1): 14-19 (王冰, 周焰, 张怀念, 赵凯. 2021. 基于改进SSD框架的遥感影像飞机目标检测方法. 火力与指挥控制, 46(1): 14-19) [DOI: 10.3969/j.issn.1002-0640.2021.01.003]

Xiao Z F, Liu Q, Tang G F and Zhai X F. 2015. Elliptic Fourier transformation-based histograms of oriented gradients for rotationally invariant object detection in remote-sensing images. International Journal of Remote Sensing, 36(2): 618-644 [DOI: 10.1080/01431161.2014.999881]

Yang Y, Li W L, Gao S Y, Bai H and Jiang W C. 2021. Objects detection from high-resolution remote sensing imagery using training-optimized YOLOv3 network. Laser and Optoelectronics Progress, 58(16): 147-153 (杨耘, 李龙威, 高思岩, 柏晗, 江万成. 2021. 基于YOLOv3网络训练优化的高分辨率遥感影像目标检测. 激光与光电子学进展, 58(16): 147-153) [DOI: 10.3788/lop202158.1601002]

Yao H G, Wang C, Yu J, Bai X J and Li W. 2020. Recognition of small-target ships in complex satellite images. Journal of Remote Sensing, 24(2): 116-125 (姚红革, 王诚, 喻钧, 白小军, 李蔚. 2020. 复杂卫星图像中的小目标船舶识别. 遥感学报, 24(2): 116-125) [DOI: 10.11834/jrs.20208238]

Yu Y, Ai H, He X J, Yu S H, Zhong X and Zhu R F. 2020. Attention-based feature pyramid networks for ship detection of optical remote sensing image. Journal of Remote Sensing, 24(2): 107-115 (于野, 艾华, 贺小军, 于树海, 钟兴, 朱瑞飞. 2020. A-FPN算法及其在遥感图像船舶检测中的应用. 遥感学报, 24(2): 107-115) [DOI: 10.11834/jrs.20208264]

Zhong P and Wang R S. 2007. A multiple conditional random fields ensemble model for urban area detection in remote sensing optical images. IEEE Transactions on Geoscience and Remote Sensing, 45(12): 3978-3988 [DOI: 10.1109/TGRS.2007.907109]

Multiscale object detection in remote sensing image by combining data fusion and feature selection

QIN Dengda, WAN Li, HE Peien, ZHANG Yi, GUO Ya, CHEN Jie

School of Geosciences and Info-physics, Central South University, Changsha 410083, China

Abstract: Remote sensing image object detection based on depth neural network model has achieved great success largely due to the support of large-scale data sets. However, from the perspective of the existing remote sensing image datasets, the number distribution of different types of ground objects is inconsistent, and the same type of ground objects is presented in different sizes. This factor leads to the scale imbalance of ground object samples.

To alleviate this problem, the strategy of weighted image fusion and multiscale feature selection is adopted. First, the pixel values of the two images in the data set are weighed to obtain the fused image. Therefore, the different types of feature samples are more balanced and have higher background diversity. Second, the target category of the corresponding scale is predicted by selecting the appropriate scale feature map, and the same scale target can be predicted on the adjacent feature map. Thus, the model can be trained according to the target scale. Finally, the bounding box of the target is predicted based on the feature map of the target center area, and the result is more consistent with the scale of the target itself.

The image fusion and multiscale feature selection network is experimentally compared in the paper on two datasets, RSOD and NWPUVHR10. The results show that the trained model is more accurate in the recognition of imbalanced objects in complex background and can adapt to the recognition of different scale objects in remote sensing images. In addition, the proposed method enhances the diversity of sample scenes and can be adapted to different scales of targets. Qualitative analysis shows that the proposed method has better robustness and performance advantages when applied to remote sensing images.

The proposed method can cope with remote sensing images with complex backgrounds, mitigate the effects of category imbalances and better fit the scenarios in which remote sensing images are used. Mitigation of category imbalances and Scale Selection enhance the performance of remote sensing image object detection.

Key words: image fusion enhancement, multi scale selection and expression, high resolution remote sensing image, object detection, convolutional neural network

Supported by National Key Research and Development Program of China (No. 2020YFA0713503); National Natural Science Foundation of China (No. 42071427); Natural Science Foundation of Hunan Province, China (No. 2020JJ4691)