改进Faster R-CNN的遥感图像多尺度飞机目标检测

沙苗苗[1,2]，李宇[1]，李安[1]

1. 中国科学院空天信息创新研究院, 北京 100094;
2. 中国科学院大学 电子电气与通信工程学院, 北京 100049

摘　要：为了提高遥感图像中多尺度飞机目标的检测精度，本文提出一种基于改进Faster R-CNN的遥感图像飞机目标检测方法。该方法借助多层级融合结构，将深层次的语义特征与浅层次的细节特征相结合，生成多种尺度的既具有精确的位置信息又具有深层次的语义特征的特征图；再借助Faster R-CNN的多尺度RPN（Region Proposal Network）机制，通过对RPN中候选区域尺度的修正，从而提高遥感图像中多尺度飞机目标的定位精度；最后利用Faster R-CNN的分类回归网络，得到飞机目标检测结果。在高分辨率遥感图像中进行了实验，对3种特征提取网络ZF、VGG-16以及ResNet-50进行改进，改进后的精度分别提高了11.34%、9.87%以及1.66%，并且生成的检测框更加贴合飞机目标。实验结果表明，本文方法适用于遥感图像多尺度飞机目标检测，在提高目标定位精度的同时降低了目标漏检现象。

关键词：遥感图像，目标检测，Faster R-CNN，多层次融合结构，多尺度

引用格式：沙苗苗,李宇,李安.2022.改进Faster R-CNN的遥感图像多尺度飞机目标检测.遥感学报,26(8):1624-1635
Sha M M, Li Y and Li A. 2022. Multiscale aircraft detection in optical remote sensing imagery based on advanced Faster R-CNN. National Remote Sensing Bulletin, 26(8):1624-1635[DOI:10.11834/jrs.20219365]

1　引　言

遥感图像目标检测一直是遥感图像处理领域的一个研究热点。作为一种重要的战略目标，飞机目标的检测具有较高的研究和应用价值，引起了研究人员极大的兴趣。随着遥感图像空间分辨率的不断提高，图像内容越来越复杂多样，早期的飞机目标检测方法多利用目标的简单特征如角点（仇建斌 等，2011）、轮廓形状（蔡栋 等，2014）等，难以应对当前高分辨率遥感图像中的复杂信息。同时，飞机目标在遥感图像上形态各异且具有多种尺度，因此，在遥感图像上进行飞机目标检测十分具有挑战性。

传统的遥感图像飞机目标检测主要分为3个步骤：首先使用滑动窗口或者显著性等方法生成候选区域，然后进行候选区域的特征提取，最后将提取到的特征输入到相应的训练好的分类器中从而得到检测结果。Li等（2011）首先使用显著性方法提取遥感图像中的候选区域，然后利用飞机目标的对称性以及模板匹配的方法进行飞机目标的检测。Zhao等（2017）使用多种尺度的滑动窗口生成相应的候选区域，然后提取候选区域的集合通道特征，最后使用AdaBoost算法得到飞机目标的检测结果。然而，显著性的方法需要人工设置相应的阈值进行候选区域的生成，容易造成目标的漏检。滑动窗口的方法需要在遥感图像上进行多种尺寸的候选区域的遍历，十分耗时。同时，这类传统方法采用的特征多为形状、梯度等浅层次特征，不具有很好的区分性，无法有效地将复杂多样的飞机目标从遥感图像中检测出来。

近年来，深度学习成为人工智能领域备受瞩目的研究内容之一（张洪群 等，2017；王宇 等，2019）。在深度学习方法中，卷积神经网络CNN（Convolutional Neural Network）由于其权值共享、平移不变性等特点，在图像分类领域取得令人瞩目的成绩（Krizhevsky等，2017；张康 等，2018）。

收稿日期：2019-10-12；**预印本**：2020-02-04
基金项目：国家自然科学基金(编号：61501460)；广东省现代视听信息工程技术研究中心开放基金
第一作者简介：沙苗苗,研究方向为遥感图像处理、目标检测。E-mail：shamm2017@radi.ac.cn
通信作者简介：李宇,研究方向为遥感图像处理。E-mail:liyu@radi.ac.cn

鉴于卷积神经网络强大的特征提取能力，研究人员将其应用到目标检测领域。其中，以基于区域的卷积神经网络R-CNN（Girshick等，2013）在VOC2012数据集上取得最高的检测精度为里程碑，基于卷积神经网络的目标检测真正的活跃起来。这种方法通过使用卷积神经网络进行候选区域特征提取，大幅提高目标检测精度，但是该方法依然存在以下问题：（1）每个候选区域都要分别进行特征提取，检测效率低；（2）需要分别进行分类器以及边框回归的训练；（3）候选区域的生成与特征提取割裂开来，无法满足实时的检测需求。针对第一个问题，He等（2014）提出的基于空间金字塔池化的卷积神经网络SPP（Spatial Pyramid Pooling）使用感兴趣区域RoI（Region of Interest）从整幅特征图中“裁剪”出候选区域对应的特征，从而大幅提高检测效率。针对第2个问题，Girshick（2015）提出的Fast R-CNN通过使用多任务损失函数，同时进行分类以及边框回归的训练，从而将目标检测集成为两个阶段：候选区域的生成以及使用卷积神经网络进行特征提取、分类和边框回归。随后Ren等（2017）提出的Faster R-CNN，通过共享特征提取网络，在经过卷积池化后的最后一个特征图上使用RPN直接生成多种尺度以及纵横比的候选区域，将目标检测的多个步骤统一到一个网络框架中，实现端到端的目标检测，检测精度以及效率大幅提升。鉴于Faster R-CNN比传统的目标检测方法在检测精度上有很大的提高，研究人员将其应用到遥感图像飞机目标检测中。Wang等（2017）基于Faster R-CNN，使用聚类的方法确定候选区域的尺度继而进行遥感图像飞机目标检测。Ren等（2018）通过在Faster R-CNN的特征提取网络中加入上下文信息，从而提高遥感图像中飞机目标尤其是小目标的检测精度。Li等（2019）基于Faster R-CNN，通过设置更小的候选区域尺度从而提高遥感图像飞机目标的检测精度。然而，上述方法均是在单一尺度的特征图上进行目标检测，不适用于遥感图像多尺度飞机目标。并且，特征图在经过卷积神经网络的多次池化之后，一方面其精确的细节信息丢失，另一方面尺度较小的目标对应特征图中的区域较小，直接在池化后的单一尺度特征图上进行目标检测可能造成目标定位精度不高以及目标漏检的现象。

针对上述问题，本文提出一种基于改进Faster R-CNN的多尺度飞机目标检测方法，通过在Faster R-CNN的特征提取网络中加入多层级融合结构构建多尺度特征提取网络，同时，针对飞机目标选取合适的候选区域生成网络参数，从而适应于遥感图像多尺度飞机目标检测。除此之外，由于网络中新加入的结构单元将高层次的语义信息与低层次的细节信息相结合，改进后的网络所生成的多尺度特征图既具有较高的定位精度又具有很好的区分性，从而在提高多尺度飞机目标检测精度的同时，提升了目标的定位精度、降低了目标的漏检现象。

2 模型方法

本文提出的遥感图像多尺度飞机目标检测流程图如图1。遥感图像多尺度飞机目标检测主要分为3个部分：特征提取网络、候选区域生成网络RPN以及分类回归网络。对于卷积神经网络，通常有许多连续的卷积层输出相同大小的特征图，则称这些卷积层处于同一网络层级（Lin等，2017）。在进行检测时，首先，使用特征提取网络进行图像的特征提取，通过多层级融合结构将高层级得到的特征图进行上采样，再将其与较低层级得到的特征图进行融合，生成一系列不同尺度的特征图F5、F4、F3以及F2。然后，在不同尺度的特征图上分别使用RPN进行候选区域的生成。最后，使用分类回归网络将不同尺度的候选区域对应到相应尺度的特征图进行分类与位置回归，从而得到最终的飞机目标检测结果。

2.1 特征提取网络

在对遥感图像进行飞机目标检测时，特征提取的好坏在很大程度上决定了最终的检测精度。本文通过对Faster R-CNN的特征提取网络进行改进，在网络中加入多层级融合结构从而生成多种尺度的特征图，对不同尺度的目标使用不同尺度的特征图进行特征提取，使其适应于遥感图像多尺度飞机目标检测。

图2为多层级融合结构的示意图。在进行多层级融合时，首先对高层级的特征图进行1×1的卷积得到固定通道数的特征图，然后对其进行2倍上采样生成更高分辨率的特征图，最后通过和经过1×1

卷积的低层级特征图进行融合，从而得到既具有深层次的语义特征又具有浅层次的空间信息的特征图。对于卷积神经网络，将每个网络层级得到的最后一个特征图作为此结构的特征图映射集。由于网络的第一个层级输出的特征图提取到的特征较浅且占用的内存较大，因此，不将其纳入到映射集中。

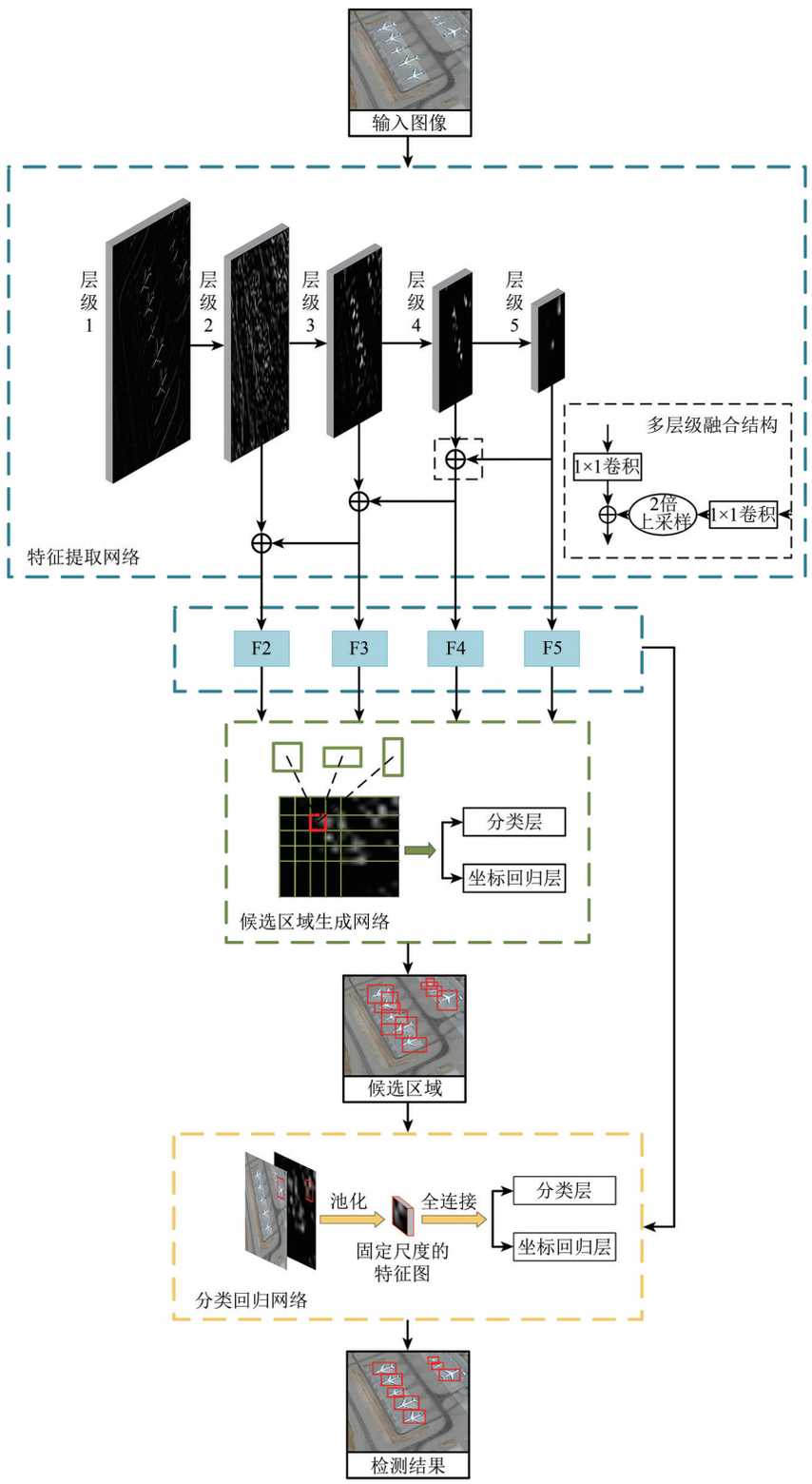

图1 遥感图像多尺度飞机目标检测流程图

Fig. 1 The flow chart of multi-scale aircraft detection in optical remote sensing imagery

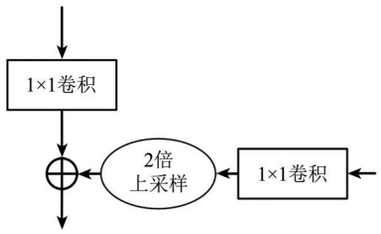

图2 多层级融合结构示意图

Fig. 2 The illustration of the multiple stages fusion structure

在选取基本的特征提取网络时，本文选取当前具有代表性的3种网络：ZF（Zeiler和Fergus，2014）、VGG-16（Simonyan和Zisserman，2015）以及ResNet-50（He等，2016），其中ZF以及VGG-16均为原始的Faster R-CNN特征提取网络，而ResNet-50则为网络层次更深，性能更优的特征提取网络。将上述3种特征提取网络分别加入多层级融合结构进行相应改进，图3展示了改进后的ResNet-50网络模型。在对ResNet-50进行改进时，首先将1×1的卷积作用于第五层级特征图conv5_3，从而得到特征图F5。然后，在该卷积的基础上，使用线性插值的方法对其进行2倍上采样。接着，对conv4_6特征图同样进行1×1的卷积，再将其与F5上采样生成的特征图进行融合得到特征图F4，F3、F2以此类推。对于ZF以及VGG-16网络，生成多尺度特征图的过程基本一致。使用这种结构，可以充分利用卷积神经网络各个层级提取到的特征，融合生成的特征图具有更丰富的语义信息。

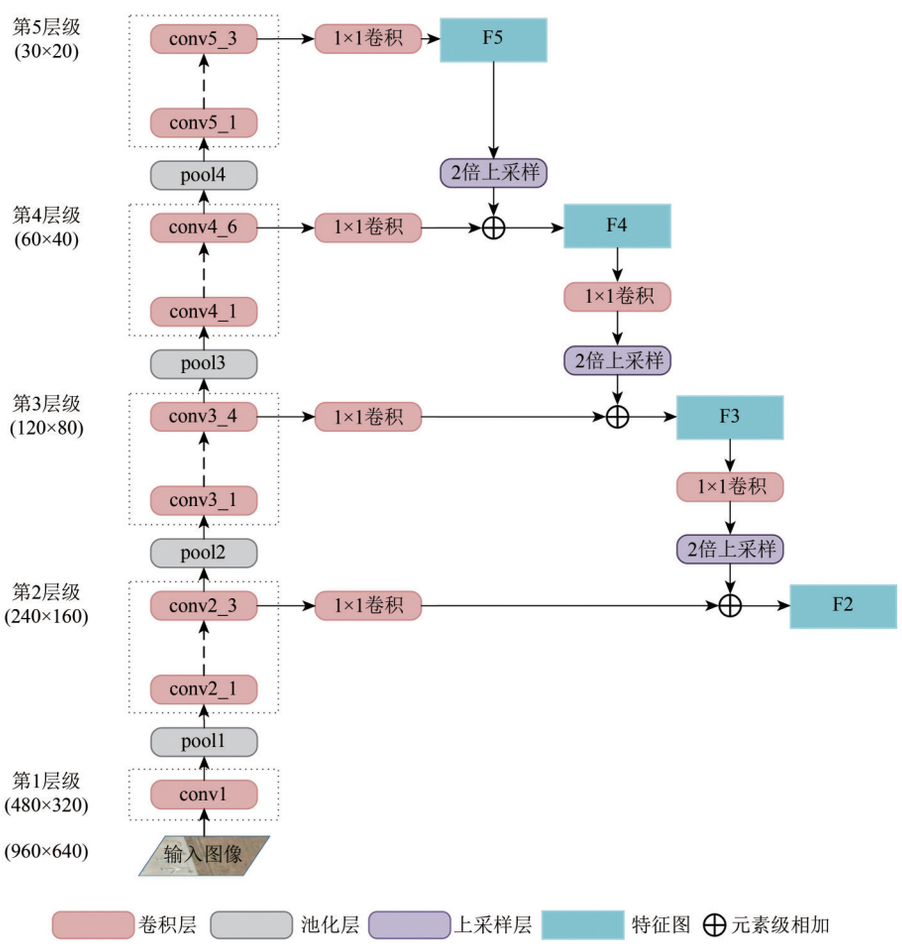

图3 改进后的ResNet-50网络模型

Fig. 3 Proposed network structure based on ResNet-50

2.2 候选区域生成网络

在RPN出现之前，候选区域的生成与特征提取网络割裂开来，造成目标检测的效率较低，不适用于多尺度飞机目标检测。RPN通过和检测网络共享特征提取网络，能够显著提高目标检测的效率以及精度，并且RPN可以生成多种尺度以及纵横比的候选区域，十分适合遥感图像多尺度飞机目标检测。原始的Faster R-CNN是对自然图像

目标进行检测，使用的候选区域尺度较大，与自然图像相比，遥感图像中飞机目标尺度较小，需要为其设置相应的小尺度候选区域。本文在对遥感图像飞机目标进行检测时，根据遥感图像中飞机目标的特点，使用多种尺度的特征图F2、F3、F4和F5，并对每个尺度的特征图设置相应尺度的候选区域，对于高分辨率的特征图F2设置小尺度的候选区域，对于较高分辨率的特征图F3设置较小尺度的候选区域，F4、F5以此类推。

2.2.1 RPN结构

如图4所示，RPN通过对卷积神经网络各个层级生成的特征图Fi（i=2，3，4，5）使用滑动窗口进行滑动，在每个滑动窗口的位置上，RPN同时进行多种尺度以及纵横比候选区域的生成，并且将滑动窗口经过的每个位置映射为固定维数的特征向量（根据选择网络的不同，维数也不同，ZF网络生成的维数为256，VGG-16和ResNet-50生成的维数为512），然后将该特征向量输入到两个全连接层中：一个是边框回归层，另一个是分类层。将特征图每个位置生成的候选区域的最大数量记为k，则每个边框回归层有4k个输出（每个位置上输出每个边框的中心点坐标以及长宽共4k个参数），同理，每个分类层输出2k个参数（每个位置上输出每个边框为目标类以及非目标类的概率）。同一个位置的每个候选区域对应原像素空间同一个位置的某个参考区域，这个参考区域就被称为基准矩形框，也叫锚点（Anchor）。锚点的设置可以使预测框更精确的回归到标签框，得到质量更优的候选框。本文在对候选区域参数进行设置时，保留与原Faster R-CNN同样的候选区域纵横比1∶2，2∶1以及1∶1，并设置更小尺度的候选区域，从而适应于遥感图像目标检测。

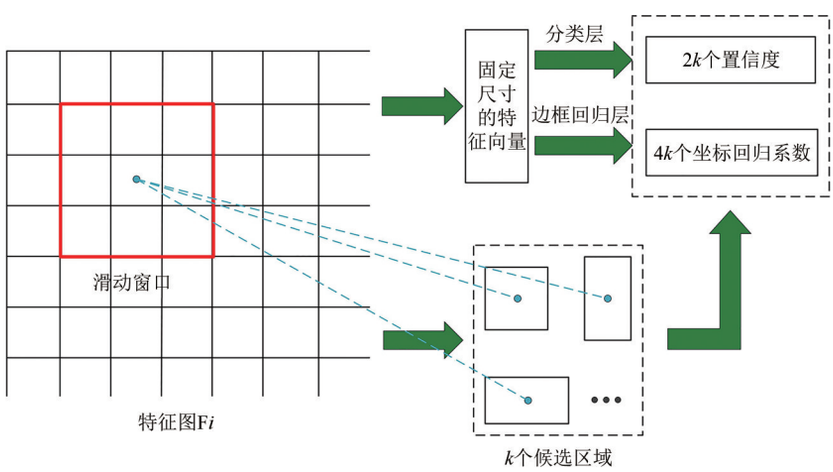

图4 RPN结构示意图

Fig. 4 Schematic diagram of RPN

2.2.2 损失函数

训练RPN时，需要为每个基准矩形框设置一个二值分类标签（是否为飞机），其中，将以下两类基准矩形框标定为正样本：

（1）与某个目标标签框具有最高的交并比IoU（Intersection over Union）；

（2）与任意目标标签框的IoU超过0.7。

将与所有目标标签框的IoU小于0.3的基准矩形框标定为负样本。其他的基准矩形框不参与RPN的训练过程。

候选区域生成网络的损失函数是一个多任务损失函数，该函数同时进行分类与坐标回归的训练任务，函数如式（1）所示：

$$L\left(\{p_i\},\ \{t_i\}\right)=\frac{1}{N_{\text{cls}}}\sum_i L_{\text{cls}}\left(p_i,\ p_i^*\right)+\lambda\frac{1}{N_{\text{reg}}}\sum_i p_i^* L_{\text{reg}}\left(t_i,\ t_i^*\right) \tag{1}$$

式中，i是一个批处理子集中基准框的序号，p_i是第i个基准框内有飞机目标的预测概率。p_i^*是第i个基准框的标签，当基准框内是正样本时，$p_i^*=1$；当基准框内是负样本时，$p_i^*=0$。t_i是一个矢量，这个矢量代表着预测框的4个参数化坐标向量，$t_i=\left(t_i^x,\ t_i^y,\ t_i^w,\ t_i^h\right)$，$t_i^*$是基准框为正样本时标签框

的4个参数化坐标向量，$t_i^* = \left(t_i^{x^*},\ t_i^{y^*},\ t_i^{w^*},\ t_i^{h^*}\right)$，具体形式为

$$t_i^x = \left(x - x_a\right)/w_a,\ t_i^y = \left(y - y_a\right)/h_a \tag{2}$$

$$t_i^w = \log\left(w/w_a\right),\ t_i^h = \log\left(h/h_a\right) \tag{3}$$

$$t_i^{x^*} = \left(x^* - x_a\right)/w_a,\ t_i^{y^*} = \left(y^* - y_a\right)/h_a \tag{4}$$

$$t_i^{w^*} = \log\left(w^*/w_a\right),\ t_i^{h^*} = \log\left(h^*/h_a\right) \tag{5}$$

式中，x，y，w，h分别表示预测框的中心横坐标、中心纵坐标、宽度和高度。x^*，y^*，w^*，h^*分别表示标签框的中心横坐标、中心纵坐标、宽度以及高度。x_a，y_a，w_a，h_a分别表示基准矩形框的中心横坐标、中心纵坐标、宽度和高度。N_{cls}和N_{reg}分别是分类以及坐标回归的归一化系数。λ用于调节分类损失和坐标回归损失的相对重要程度。L_{cls}是分类的损失函数，该损失函数是一个二分类的逻辑回归损失函数，其表达式如式（6）：

$$L_{\text{cls}}\left(p_i,\ p_i^*\right) = -\log\left(p_i^* p_i + \left(1 - p_i^*\right)\left(1 - p_i\right)\right) \tag{6}$$

L_{reg}是坐标回归的损失函数，其具体的表达式为：

$$L_{\text{reg}}\left(t_i,\ t_i^*\right) = \sum_{S \in \{x,\ y,\ w,\ h\}} \text{smooth}_{L_1}\left(t_i^s - t_i^{s^*}\right) \tag{7}$$

式中，smooth_{L_1}函数为：

$$\text{smooth}_{L_1}(x) = \begin{cases} 0.5x^2, & |x| < 1 \\ |x| - 0.5, & |x| \geqslant 1 \end{cases} \tag{8}$$

2.3 分类回归网络

使用RPN得到一系列尺度、纵横比不同的候选区域之后，如图5所示，首先找到候选区域对应在特征图中的位置，进行感兴趣区域RoI（Region of Interest）投影，从而提取候选区域对应的特征，这里的感兴趣区域就是通过RPN得到的候选区域。本文中，根据生成的候选区域尺度将其投影到不同的特征图。将高度为h宽度为w的感兴趣区域投影到Fi特征图，其中：

$$i = \left\lfloor i_0 + \log_2\left(\sqrt{wh}/128\right)\right\rfloor \tag{9}$$

式中，“⌊ ⌋”表示向下取整；i_0是w=128，h=128的RoI所映射的特征图编号。本文中将i_0设置为4，即宽为128，长为128的RoI映射到F4特征图进行RoI投影，宽为64，长为64的RoI映射到F3特征图进行投影，其他尺度以此类推。获得RoI在其对应特征图中的特征之后，进行RoI池化，将尺度纵横比不同的感兴趣区域对应的特征区域均池化为固定尺度的特征图。然后，再将该特征图输入到两个全连接层中得到相应的RoI特征向量$\boldsymbol{R}$。最后将$\boldsymbol{R}$分别输入到两个全连接层中，使用softmax函数计算该区域是飞机目标以及非飞机目标的概率，同时进行边框回归得到飞机目标的边框参数。

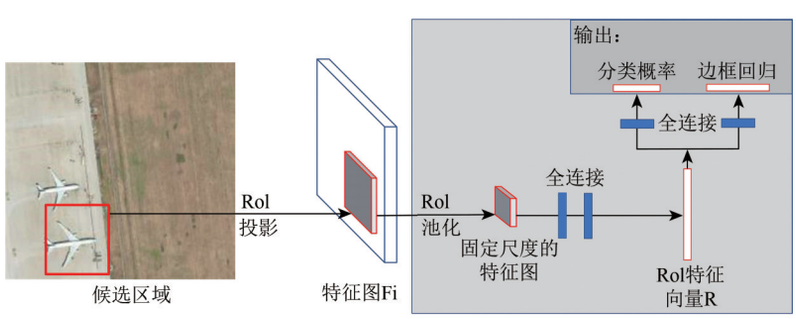

图5 分类回归网络结构示意图

Fig. 5 Schematic diagram of classification and regression network

3 实验和分析

本文实验主要是在基于Linux系统的Caffe框架下完成，服务器处理器为Intel Xeon（R） CPU E5620 @ 2.40 GHz，使用GPU加速，显卡为NVIDIA TITAN Xp，使用Python进行编程。训练时，各模型迭代40000次，其中，前20000次的学习率设置为0.001，后20000次的学习率设置为0.0001，动量设置为0.9，权值衰减参数为0.0001。

3.1 实验数据

实验数据选择RSOD数据集（Long等，2017），该数据集由武汉大学团队标注，数据集来源主要是Google Earth和天地图，分辨率为0.5—2.0 m。本文仅使用其中的飞机目标数据集，共有446张宽高在像素值1000左右的飞机图像，图像中共有4993个飞机目标。其中，将60%的图像作为训练以及验证数据集，将其余40%的图像用于测试。由于深度学习的方法进行目标检测时需要大量的

训练数据，因此，对于参与训练以及验证的图像使用水平镜像以及将图像进行90°、180°以及270°旋转的方法，获得原始训练以及验证影像8倍的图像。RSOD数据集的部分样本图像如图6所示。

（a）样本图像1 （b）样本图像2

（a）The first sample image （b）The second sample image

图6 RSOD数据集部分样本图像

Fig. 6 Image sample of the RSOD dataset

3.2 评价准则

为评估本文算法进行遥感图像飞机目标检测的有效性，将两种广泛使用的标准度量方法：精度—召回率曲线图PRC（Precision-Recall Curve）以及平均精度AP（Average Precision）作为本文飞机目标检测的评价标准。其中，PRC是以召回率（recall）为横坐标，精度（precision）为纵坐标，记录随着阈值变化时，precision与recall值变化关系的曲线。平均精度AP就是当recall从0到1变化时precision的平均值，也就是PRC曲线与横纵坐标围成的面积。precision以及recall的具体计算公式如式（10）、（11）所示：

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{10}$$

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{11}$$

式中，TP（True Positive）表示预测的目标实际也是目标；FP（False Positive）表示预测的目标实际是背景；FN（False Negative）表示预测的背景实际是目标。本文将与目标标签框的IoU值大于等于0.5的预测框作为TP，反之，该预测框为FP。

3.3 RPN候选区域尺度的设置

本文先利用ZF、VGG-16以及ResNet-50网络对RSOD数据集进行特征提取，由于遥感图像飞机目标具有多种尺度，在上述特征提取网络的最后一个特征图上使用RPN进行候选区域生成时，需要为其设置相应尺度的候选区域，相应的检测精度（%）如表1所示。

表1 不同尺度的候选区域检测精度对比

Table 1 Comparison of detection performance under different anchor scales

候选区域尺度	特征提取网络		
	ZF	VGG-16	ResNet-50
（32×32，64×64，128×128）	78.17	80.08	88.87
（64×64，128×128，256×256）	78.32	80.44	85.66
（128×128，256×256，512×512）	70.52	79.38	80.76
（32×32，64×64，128×128，256×256）	78.09	80.32	88.76
（64×64，128×128，256×256，512×512）	**78.39**	**80.55**	86.81
（32×32，64×64，128×128，256×256，512×512）	78.32	80.45	**88.89**

注：表中粗体表示相应网络在不同候选区域尺度下的最高精度。

从表1中可以看出，特征提取网络的选择以及候选区域尺度的设置对遥感图像飞机目标检测精度均影响重大。当候选区域尺度为Faster R-CNN原始候选区域尺度（128×128，256×256，512×512）时，3种特征提取网络均取得最差的检测精度，这是由于原始的候选区域尺度设置针对的是自然图像目标，相比遥感图像目标尺度较大，不适用于遥感图像目标检测。表1中，ZF、VGG-16网络均在候选区域尺度为（64×64，128×128，256×256，512×512）时取得最佳的检测精度，分别为78.39%以及80.55%，而ResNet-50则是在候选区域尺度为（32×32，64×64，128×128，256×256）时取得最优的检测精度88.89%。即使对ZF以及VGG-16设置了相应较小尺度的候选区域，但是由于其网络特征提取能力相较于ResNet-50较弱，对于尺度较小的候选区域提取到的特征更加有限，造成对小目标的提取精度不高。

本文在进行RPN候选区域参数设置时，对ZF、VGG-16以及ResNet-50分别按照其取得最佳检测精度时候选区域的尺度进行RPN参数设置，而对3个改进后的网络，具体的参数设置见表2。

表2 改进后网络的候选区域尺度设置

Table 2 Anchor scale settings for proposed networks

特征图	分辨率等级	候选区域尺度
F2	高分辨率	32×32
F3	较高分辨率	64×64
F4	较低分辨率	128×128
F5	低分辨率	256×256

3.4 特征提取网络的对比

为验证本文方法的有效性，将3个改进后的网络ZF*、VGG-16*以及ResNet-50*分别与相应的改进前网络进行对比，以RSOD数据集为训练测试数据集，这6种网络的检测精度以及测试速率如表3所示，对应的PRC如图7。

表3 不同特征提取网络检测精度时间对比

Table 3 Comparison of detection performance of different feature extraction networks

特征提取网络	平均精度/%	测试速率/(s/张)
ZF	78.39	0.097
VGG-16	80.55	0.135
ResNet-50	88.89	0.206
ZF*	89.73	0.210
VGG-16*	90.42	0.249
ResNet-50*	**90.55**	0.290

注：表中粗体表示不同特征提取网络的最高检测精度；*表示改进后提取网络。

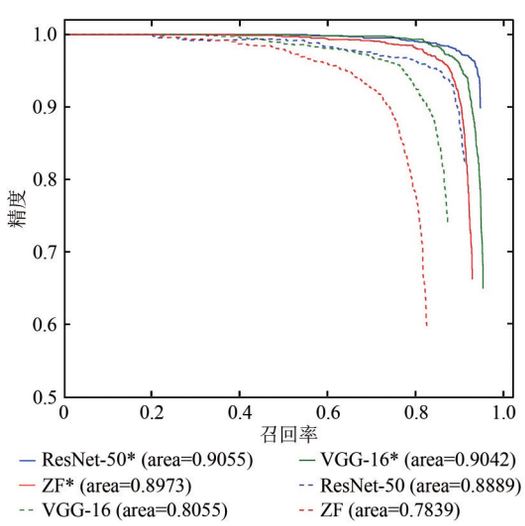

图7 各个网络对应的PRC

Fig. 7 The PRC for each method

从表3中可以看出，改进后的各个网络在检测精度上均有显著提高。其中，ZF*提高了11.34%，VGG-16*提高了9.87%，而ResNet-50*则提高了1.66%。相比于ZF*以及VGG-16*网络，ResNet-50*提高的幅度更小，这是由于ResNet-50本身的特征提取能力已经很强，通过对各个尺度候选区域的位置修正对于整体精度的提高不是那么明显。对于VGG-16以及ZF网络，其本身的特征提取能力稍弱，如图9中（a）、（b）图相比于（c）图出现了更多的漏检以及误检现象，并且这种漏检以及误检现象多发生于尺度偏小的多尺度目标，而改进后的ZF*、VGG-16*网络通过在多种尺度的浅层次特征图中加入深层次语义信息从而增强各个尺度特征图提取的特征。以VGG-16网络为例，尺度为128×128的候选区域在特征提取时对应F5中4×4区域的特征，而VGG-16*网络中128×128的候选区域对应着F5中4×4区域上采样2倍后的8×8区域加上F4中8×8区域的特征，其他尺度的候选区域以此类推。因此，这两种网络在提高各个尺度候选区域定位精度的同时大幅减少其漏检以及误检现象，从而大幅提高检测精度。同样的，从图7中可以看到，改进后的3种网络与两个坐标轴围成的面积均分别大于相应的改进前的网络，各个网络的precision值先是趋于平缓，当recall值增加到0.7左右，ZF的precision值出现大幅降低，随着recall值的进一步增加，性能相对较差的VGG-16的precision值大幅降低，而改进后的3种网络在保持着高recall值的同时具有较高的precision值，这也充分说明了本文方法对于提高目标检测精度的有效性。

图8为测试样本图，可以看到图中飞机目标尺度差异较大，从十几像素到上百像素不等。图9展示了各个网络对于图8的检测结果图，图9（a）、（b）、（c）、（d）、（e）、（f）分别对应着ZF、VGG-16、ResNet-50、ZF*、VGG-16*以及ResNet-50*的检测结果图。

图8 测试样本图

Fig. 8 Test image sample

从图9中可以看到，首先，ZF*、VGG-16*以及ResNet-50*相比于改进前的特征提取网络，对于

目标的定位精度更高，可以明显的看到，相比于图9（a）、图9（b）、图9（c），图9（d）、图9（e）、图9（f）中红色预测框与蓝色标签框更为贴合。除此之外，加入这种结构后的网络能够检测出原始特征提取网络遗漏的目标，如图9（d）相对于图9（a），图9（e）相对于图9（b），绿色漏检标签框的数量减少。为了进一步定量的说明本文方法对于目标定位精度的提高，本文通过设置更高的IoU阈值进行各个方法检测精度对比，对比结果如表4。

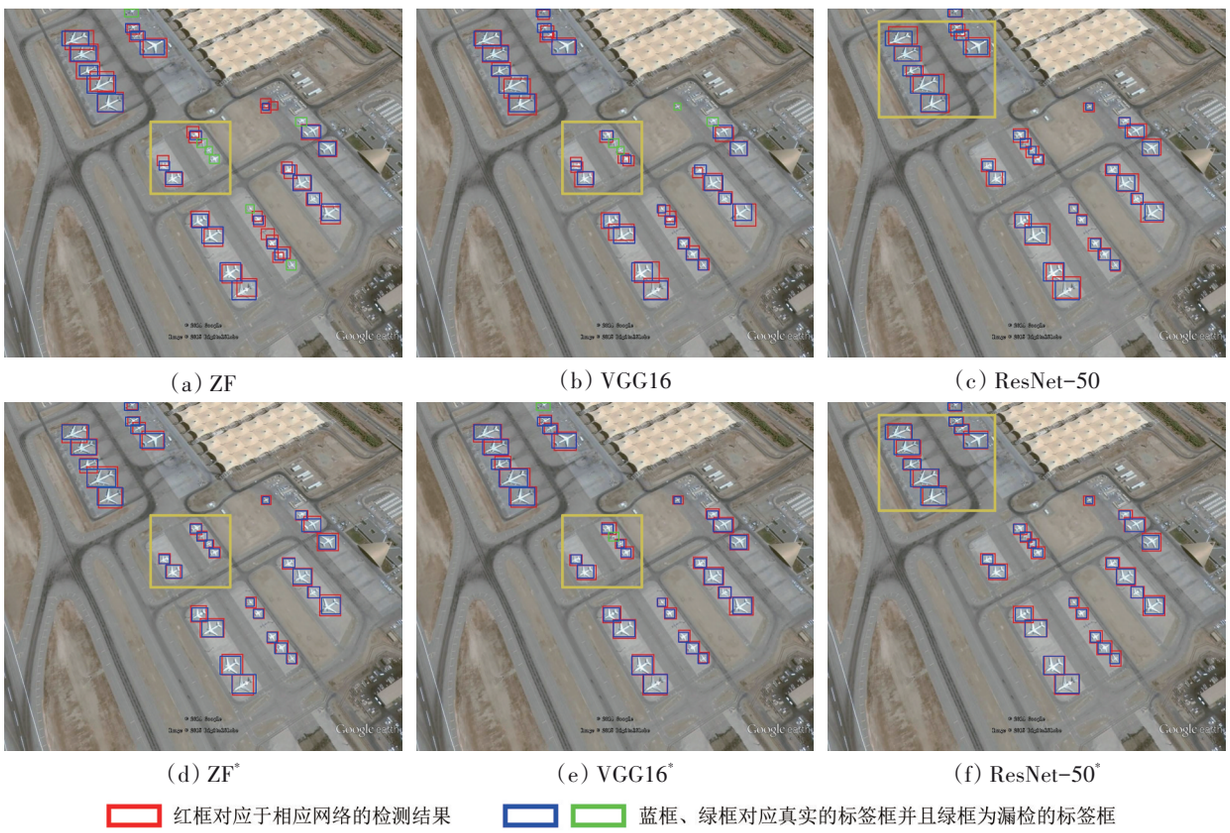

图9 各个网络检测结果示例图

Fig. 9 Detection results diagram of each method

表4 不同IoU阈值下各个网络检测精度对比

Table 4 Comparison of AP values of each method under different IoU

/%

特征提取网络	IoU		
	0.5	0.6	0.7
ZF	**78.39**	59.28	37.69
VGG-16	**80.55**	69.34	47.04
ResNet-50	**88.89**	77.48	45.61
ZF*	**89.73**	81.03	69.93
VGG-16*	**90.42**	89.02	76.65
ResNet-50*	**90.55**	90.21	80.12

注：表中粗体表示相应网络在不同IoU阈值下的最高精度值。

从表4中我们可以看到，随着IoU阈值的增加，各个网络的AP值都在降低，其中，改进前网络精度值降低的幅度更大，而改进后网络尤其是ResNet-50*即使在IoU阈值为0.7时依然取得了80.12%的AP值。这也充分说明本文方法对于提高目标定位精度的有效性。

3.5 多尺度目标检测泛化能力实验

为了充分验证本文方法的有效性，本文还将使用GF-2影像数据进行多尺度飞机目标检测的泛化能力实验。实验选取首都国际机场对应的影像区域，该影像为全色波段与多光谱的红、绿、蓝波段融合后的图像，分辨率为1 m，像素为4600×6500。将该区域以100像素的重叠进行裁剪，得到40幅1000像素×900像素的图像切片，使用ResNet-50*以及ResNet-50分别对这40幅图像进行检测，再将检测好的图像进行拼接，对于重叠处的多余检测框，使用NMS（Non-Maximum Suppression）进行相应的后处理。具体的定量检测结果如表5，相

应的检测结果图如图10，其中，左上侧黄框对应ResNet-50*的检测局部放大图。

表5 ResNet-50与ResNet-50*对于多尺度飞机目标检测精度对比

Table 5 Comparison of multi-scale aircraft detection performance of ResNet-50 and ResNet-50*

检测方法	正检	误检	漏检	precision/%	recall/%
ResNet-50	170	1	25	99.42	87.18
ResNet-50*	176	1	19	99.44	90.27

从图10可以看出，对于图像中多尺度飞机目标，ResNet-50*大多可以将其检测出来，从局部放大图可知，ResNet-50*生成的检测框与目标贴合的较好，定位精度较高。结合表5进行进一步的定量分析，可以看到，相比于ResNet-50，ResNet-50*的precision值略微提高，而recall值则增加了3.09%，这是由于ResNet-50*在高层级的语义特征中融入了高分辨率的低层级特征，在提高目标定位精度的同时，语义信息也更为充分，目标漏检的数量也随之减少。以上分析充分表明了，ResNet-50*不仅适用于多尺度飞机目标检测而且具有良好的泛化能力。

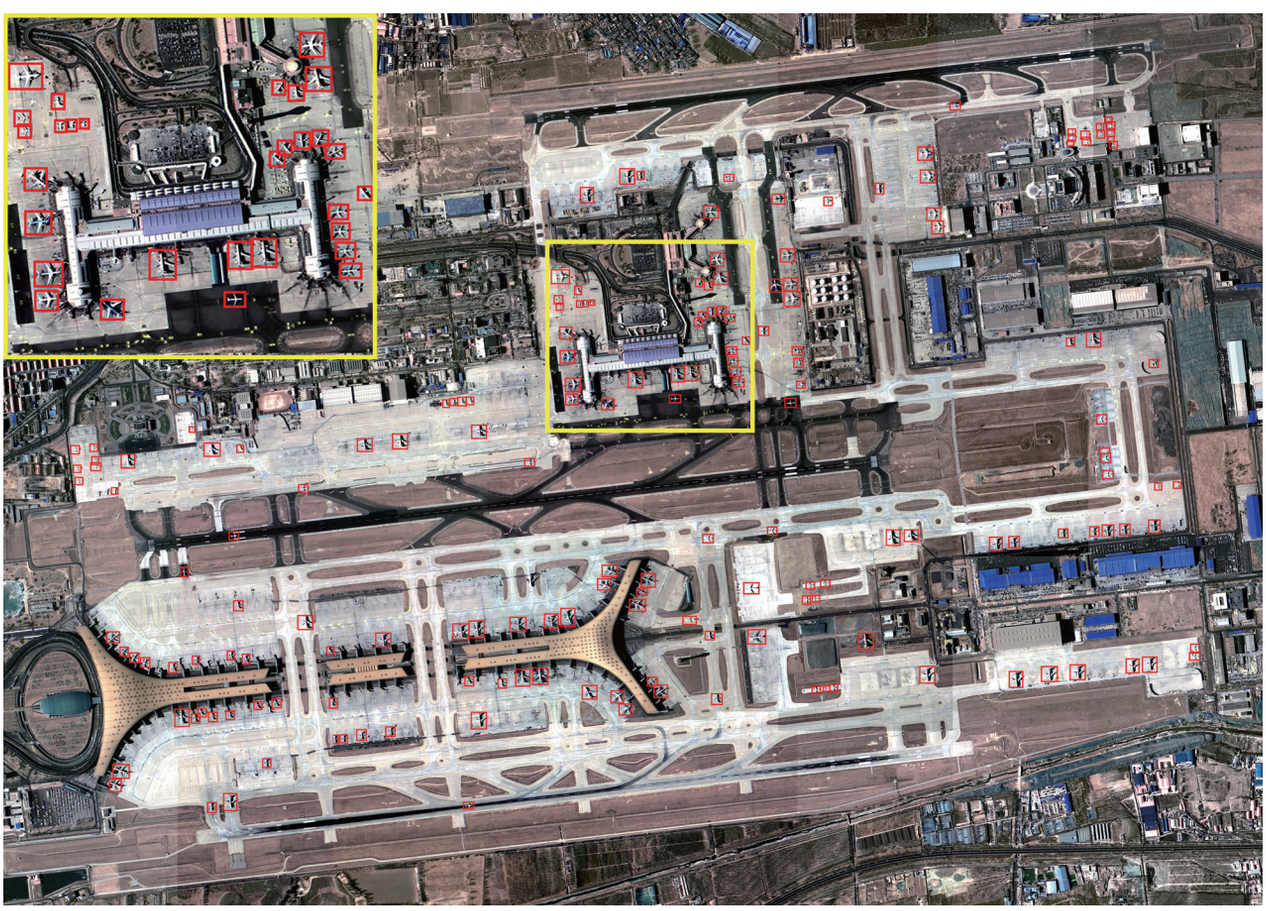

图10 ResNet-50*网络对GF-2首都国际机场图像的检测结果

Fig. 10 Detection results of ResNet-50* on Beijing Capital International Airport GF-2 imagery

4 结 论

本文针对目前目标检测方法使用单一尺度的特征图进行多尺度飞机目标检测造成检测精度不佳的问题，提出使用多尺度的特征图进行多尺度飞机目标检测的方法。该方法基于改进的Faster R-CNN，通过在其特征提取网络中加入多层级融合结构，充分利用不同网络层级的特征，生成的多尺度特征图既具有低层级精确的位置信息又具有高层级的语义特征，从而在提高多尺度飞机目标检测精度的同时，提高其定位精度。然后，对其RPN候选区域尺度进行修正，使其适应于遥感图像飞机目标检测。实验结果表明：（1）加入多层级融合结构的网络可以对多尺度飞机目标生成

与之尺度相符的检测框，在提高飞机目标检测精度的同时降低目标漏检的情况；（2）通过对RPN候选区域尺度的修正，提高了遥感图像飞机目标检测精度；（3）改进后的网络具有良好的泛化能力，适用于遥感图像多尺度飞机目标检测。然而，本文方法在提高目标检测精度的同时对于目标检测速率也造成了一定的影响，因此，后续的研究将着重于网络模型的优化，以期在较小的时间代价下得到最高的检测精度。

参考文献（References）

Cai D, Chen Y M and Wei W. 2014. Study on aircraft recognition in multi-spectral remote sensing image based on skeleton characteristics analysis. Bulletin of Surveying and Mapping, (2): 50-54, 71 (蔡栋, 陈焱明, 魏巍. 2014. 基于骨架特征的多光谱遥感影像飞机目标识别方法研究. 测绘通报, (2): 50-54, 71) [DOI: 10.13474/j.cnki.11-2246.2014.0052]

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

Girshick R, Donahue J, Darrell T and Malik J. 2013. Rich feature hierarchies for accurate object detection and semantic segmentation. arXiv:1311.2524

He K M, Zhang X Y, Ren S Q and Sun J. 2014. Spatial pyramid pooling in deep convolutional networks for visual recognition//Computer Vision - ECCV 2014. Switzerland: Springer, 8681: 346-361 [DOI: 10.1007/978-3-319-10578-9_23]

He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]

Krizhevsky A, Sutskever I and Hinton G E. 2017. ImageNet classification with deep convolutional neural networks. Communications of the ACM, 60(6): 84-90 [DOI: 10.1145/3065386]

Li W, Xiang S M, Wang H B and Pan C H. 2011. Robust airplane detection in satellite images//2011 18th IEEE International Conference on Image Processing. Brussels, Belgium: IEEE: 2821-2824 [DOI: 10.1109/ICIP.2011.6116259]

Li Y B, Zhang S Y, Zhao J F and Tan W A. 2019. Aircraft detection in remote sensing images based on deep convolutional neural network. IOP Conference Series: Earth and Environmental Science, 252(5): 052122 [DOI: 10.1088/1755-1315/252/5/052122]

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

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]

Qiu J B, Li S J and Wang W. 2011. A new approach to detect aircrafts in remote sensing images based on corner and edge information fusion. Microelectronics and Computer, 28(9): 214-216 (仇建斌, 李士进, 王玮. 2011. 角点与边缘信息相结合的遥感图像飞机检测新方法. 微电子学与计算机, 28(9): 214-216) [DOI: 10.19304/j.cnki.issn1000-7180.2011.09.056]

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]

Ren Y, Zhu C R and Xiao S P. 2018. Deformable faster R-CNN with aggregating multi-layer features for partially occluded object detection in optical remote sensing images. Remote Sensing, 10(9): 1470 [DOI: 10.3390/rs10091470]

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

Wang H Z, Gong Y C, Wang Y, Wang L F and Pan C H. 2017. DeepPlane: a unified deep model for aircraft detection and recognition in remote sensing images. Journal of Applied Remote Sensing, 11 (4): 042606 [DOI: 10.1117/1.JRS.11.042606]

Wang Y, Yang Y, Wang B S, Wang T, Bo X H and Wang C Y. 2019. Building segmentation in high-resolution remote sensing image through deep neural network and conditional random fields. Journal of Remote Sensing, 23(6): 1194-1208 (王宇, 杨艺, 王宝山, 王田, 卜旭辉, 王传云. 2019. 深度神经网络条件随机场高分辨率遥感图像建筑物分割. 遥感学报, 23(6): 1194-1208) [DOI: 10.11834/jrs.20198141]

Zeiler M D and Fergus R. 2014. Visualizing and understanding convolutional networks//Fleet D, Pajdla T, Schiele B and Tuytelaars T, eds. Computer Vision - ECCV 2014. Switzerland: Springer: 818-833 [DOI: 10.1007/978-3-319-10590-1_53]

Zhang H Q, Liu X Y, Yang S and Li Y. 2017. Retrieval of remote sensing images based on semisupervised deep learning. Journal of Remote Sensing, 21(3): 406-414 (张洪群, 刘雪莹, 杨森, 李宇. 2017. 深度学习的半监督遥感图像检索. 遥感学报, 21(3): 406-414) [DOI: 10.11834/jrs.20176105]

Zhang K, Hei B Q, Zhou Z and Li S Y. 2018. CNN with coefficient of variation-based dimensionality reduction for hyperspectral remote sensing images classification. Journal of Remote Sensing, 22(1): 87-96 (张康, 黑保琴, 周壮, 李盛阳. 2018. 变异系数降维的CNN高光谱遥感图像分类. 遥感学报, 22(1): 87-96) [DOI: 10.11834/jrs.20187075]

Zhao A, Fu K, Sun H, Sun X, Li F, Zhang D B and Wang H Q. 2017. An effective method based on ACF for aircraft detection in remote sensing images. IEEE Geoscience and Remote Sensing Letters, 14(5): 744-748 [DOI: 10.1109/LGRS.2017.2677954]

Multiscale aircraft detection in optical remote sensing imagery based on advanced Faster R−CNN

SHA Miaomiao[1,2], LI Yu[1], LI An[1]

1. Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China;
2. University of Chinese Academy of Sciences, School of Electronic, Electrical and Communication Engineering, Beijing 100049, China

Abstract: Aircraft detection from optical imagery is a significant application in remote sensing. Traditional methods based on corner points or shape of the aircraft can only generate shallow features with limited representative ability. These methods are insufficient for detecting aircraft in remote sensing imagery under complex and diverse circumstances. Current methods based on CNNs, especially Faster R-CNN, have improved the detection performance greatly with its magnificent feature extraction ability. However, detecting aircraft on a single-scale feature map is unsuitable for multiscale aircraft in remote sensing imagery. After several pooling operations on a single-scale feature map, the feature map loses its precise details and small target that corresponds to a smaller area in the feature map. Thus, aircraft detection may result in low target positioning accuracy and target missing.

An advanced Faster R-CNN is presented by constructing a multiscale feature extraction network using multistage fusion structure to detect aircraft with multiple scales. The promoted network produces features of higher resolution by upsampling deep feature maps. These features are then enhanced with shallow features at the same scale. After this modification, we end up with four feature maps F2, F3, F4, and F5, which have different scales. The structure combines the high-level semantic information with the low-level detailed information. Thus, the generated multiscale feature maps have high positioning accuracy and good distinguishability. In addition, because the original RPN anchors are extremely large to cover the range of aircraft sizes in remote sensing imagery, we select suitable RPN anchor parameters for aircraft detection, i.e., anchor size of 32^2 for the larger-scale feature map F2, 64^2 for the large-scale F3, 128^2 is set for the F4, and 256^2 for the small-scale F5. With these settings, the RPN can generate proposals, which can cover the aircraft of multiple scales. Finally, these proposals are assigned to their corresponding feature map, and we use the classification and regression network to obtain our final detection results.

The experiment was carried out on RSOD dataset, in which only the aircraft dataset was used for training, validation, and testing. Comparison of detection performance with different anchor scales showed that anchor scales greatly affect detection accuracy, and our selection of anchor scales is suitable for the dataset. Three feature extraction networks (ZF, VGG-16, and ResNet-50) were modified based on Faster R-CNN using multistage fusion structure. The experiment showed that the modification can effectively improve the model's ability of detecting multiscale aircraft. Compared with models without the modification, AP increased by 11.34%, 9.87%, and 1.66% for the three networks. The qualitative and quantitative results also showed that this modification can generate adaptive detection box. The experiment results on Beijing Capital International Airport GF-2 imagery showed that this method performs well in different remote sensing imagery, in which most airplanes in the airport were detected successfully.

We can draw the following conclusions: (1) the proposed method is suitable for multiscale aircraft detection, and it can generate detection box consistent with the scale of multiscale aircraft targets while reducing missing targets; (2) correction of the RPN candidate region scale improves the accuracy of aircraft detection in remote sensing imagery; (3) the method has good generalization ability.

Key words: remote sensing image, object detection, Faster R-CNN, multiple stages fusion structure, multi-scale

Supported by National Natural Science Foundation of China (No. 61501460); Open Fund of Guangdong Province Modern Audio-visual Engineering Technology Research Center