光电技术及应用

DOI：10.16818/j.issn1001－5868.2021111306

一种基于FCOS神经网络的小建筑物目标检测方法

林思玉，王敬东*，顾东泽，姜宜君

（南京航空航天大学 自动化学院，南京 211106）

摘　要： 提出了一种基于FCOS神经网络的小建筑物目标检测算法，针对FCOS模型在特征提取阶段提取到的小建筑物目标特征较少问题，引入多尺度检测和可变形卷积方式，加强网络对小建筑物目标的特征提取能力，并通过改进后的SGE注意力机制降低特征图中的干扰噪声权重。改进后的网络可以提取到更多的小建筑物目标特征，对环境干扰噪声的鲁棒性更强。在自己搭建的数据集上进行了实验测试，结果表明，在相同环境下网络改进后建筑物的整体检测准确率提升了1.7%，其中对小建筑物目标提升了3.6%，减少了小建筑物目标漏检、误检的问题。

关键词： FCOS算法；可变形卷积；多尺度检测；空间注意力机制；建筑物检测

中图分类号： TP391　**文章编号：** 1001－5868(2022)02－0369－08

Small Buildings Detection Method Based on FCOS Neural Network

LIN Siyu，WANG Jingdong，GU Dongze，JIANG Yijun

（College of Automation Engineering，Nanjing University of Aeronautics and Astronautics，Nanjing 211106，CHN）

Abstract： A small building target detection algorithm is proposed based on FCOS neural network. Aiming at the problem of insufficient target features of small buildings extracted in the feature extraction stage of the FCOS algorithm，multi-scale convolution and deformable convolution were used to improve the ability of the network to extract features of small buildings. And through the improved SGE attention mechanism，the weight of interference noise in the feature map becomes lower. The improved network can extract more target feature information of small buildings and is more robust to noise. Experimental results obtained from the building dataset made by ourselves show that，under the same test environment，the network's overall detection accuracy（mAP）of buildings under normal，dense and occluded conditions is improved by 1.7%，and that of small buildings is improved by 3.6%，reducing the missing and error detection of small building targets.

Key words： FCOS algorithm；deformable convolution；multi-scale detection；spatial attention mechanism；building detection

0　引言

随着城市建设速度的日益加快，城区管理和规划的难度也日益增大，进行高精度的建筑物检测有助于实现对建筑物的动态监测和管理，对城市建设具有重要意义。目前基于人工设计特征的传统检测算法缺乏足够的泛化能力，受外界环境影响较大。因此近几年的研究逐渐开始利用深度学习方法来自动提取建筑物特征[1-3]。相比于传统算法，深度学习算法采用自主学习更新权重方式，可以提取更加深层次的特征，有效提高了目标检测的准确性。

国内外针对小建筑物目标检测的研究大体分为语义分割和目标检测两个方向，且均有一些卓有成效的算法被提出[4-10]。

这些算法通过特征融合、边缘辅助、扩大感受野

收稿日期：2021－11－13.

基金项目：国家自然科学基金项目(U1531110).

* **通信作者：**王敬东　**E-mail：** wjd1291@nuaa.edu.cn

等方法来增加建筑物目标的特征信息，有效改善了建筑物目标漏检和错检的情况。但是建筑物周围环境复杂，往往存在分布密集、尺度变化大，被树木遮挡等干扰[11]。小建筑物目标占像素点较少，受上述干扰影响更为严重，如果网络提取到的特征不够充分，则很容易被漏检和误检。

为了解决小建筑物目标难检测的问题，本文选取 FCOS 无框检测网络作为基础网络模型，对其中的特征提取网络进行改进，引入多尺度检测和可变形卷积方式，并加入改进后的 SGE 注意力机制，使得模型可以提取出更多小建筑物目标的特征信息，抑制环境噪声，提高检测准确度。

1 问题分析与改进方案

常用的一阶段、二阶段有框(anchor)系列检测算法虽然检测精度较高，但是存在参数多、计算量大、正负样本失衡等问题。对于建筑物这种形状尺寸差异较大的物体，选择合适尺寸比例的锚框比较困难。而且图像中小建筑物目标往往只占据几个或几十个像素点[13]，如果检测中锚框选取过大，即使已经被检测到也可能因为交并比(IOU)较小而在 NMS 阶段被抑制掉，导致漏检。

无框检测系列网络中的 FCOS 算法直接学习像素点到边界框的四个边距，同时将落在真值框中的像素点都作为正样本学习的方式，可以避免锚框选取困难的问题，让小建筑物目标产生更多的正样本，更适用于复杂情况下的小建筑物目标检测。本文选取 FCOS 无框检测算法作为基础算法，并针对小建筑物目标实际检测效果进行相应算法优化。

FCOS 算法[14]采用类似于 FCN 的语义分割方式对图像进行逐像素处理预测，如图 1 所示。主要由特征提取网络、特征融合网络(FPN)和头部检测网络三部分组成。其中特征提取网络一般采用

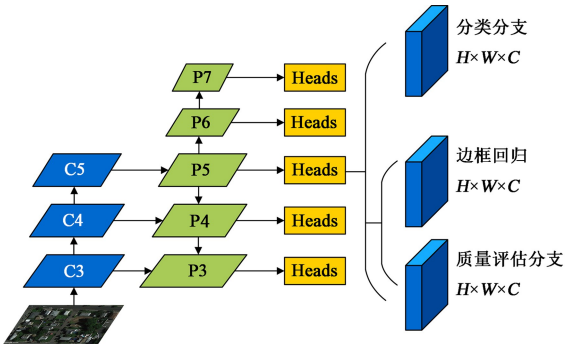


图 1 FCOS 网络特征结构

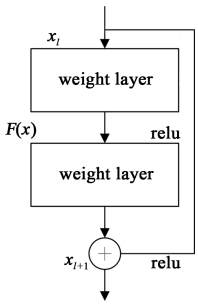


图 2 残差基本单元结构图

ResNet101 网络，ResNet 网络创新性地设计了一种残差结构，残差基本单元结构如图 2 所示，网络卷积层输出等于输入 x_l 和 x_l 经过两个卷积层后的输出 $F(x_l,w_l)$ 之和。有效地解决了由于网络层数加深引起的梯度爆炸或消失的问题。

具体工作过程为，首先输入图片经由特征提取网络提取特征，获得 C1～C5 层特征，通过下采样操作后，尺度分别为原始图片大小的 1/4，1/8，1/16，1/32，下采样倍数越多，空间位置信息越少，高级语义信息越丰富。为了更好地利用各层特征，将特征提取网络中的 C3～C5 层特征图送入特征融合网络(FPN)进行特征融合，获得 P3～P5 层特征图，P5 层特征图再通过卷积操作获得 P6 和 P7 层特征图。最后在头部检测网络中采用分层预测的方式预测建筑物类别和位置坐标。FCOS 改变了通过计算预设 anchor 框与真实框的 IOU 来选取正样本的方式，直接将落在真实框内的像素点作为正样本，并学习像素点到真实框四个边界的距离，回归一个四维向量，如式(1)所示，示意图如图 3 所示。

$$t=(l,t,r,b) \tag{1}$$

其中，l,t,r,b 分别表示采样点到边界框的四条边的边距。为了筛选质量较高的预测框，FCOS 算法提出了 Center-ness 分支，计算公式为

$$\text{Centerness}=\sqrt{\frac{\min(l,r)}{\max(l,r)}\times\frac{\min(t,b)}{\max(t,b)}} \tag{2}$$

采样点越接近真实框边界位置，预测的框质量越差。通过 Center-ness 分支可以在预测阶段，降低低质量

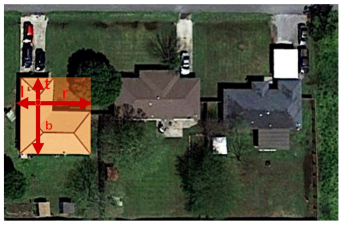


图 3 FCOS 采样示意图

预测框的权重，便于后续 NMS 筛选阶段保留高质量的候选框。

采用 FCOS 算法训练后的测试结果如图 4 所示，图中用黄色圈圈出的地方为漏检的建筑物，可以看出图片中的中、大型建筑物检测效果良好，但小建筑物目标存在很多漏检现象。为了更清晰地看到网络特征提取效果，本文对图 4(a)特征提取网络输出的 C2～C5 层和特征融合网络输出的 P3～P5 层的特征图进行可视化，如图 5 所示。图中颜色越偏亮黄色，表明此处学到的特征越多。从图中可以看到，随着网络层数的加深，小建筑物目标的空间和语义特征信息损失严重，有一些甚至在网络深层消失，同时道路、树木等干扰噪声也被当作建筑物特征学习，对小建筑物目标检测产生干扰。



(a) 漏检建筑物 1　(b) 漏检建筑物 2　(c) 漏检建筑物 3

图 4　FCOS 检测结果

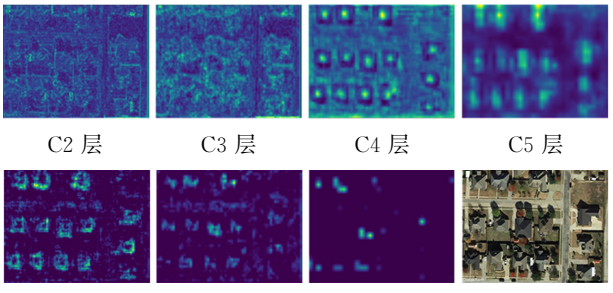


图 5　特征图可视化结果

分析认为产生上述问题的原因主要有三个，一是特征提取不够充分；即使 FCOS 算法中采用特征融合网络对不同层的特征进行了特征融合，补充浅层网络的语义信息，但是提取到的小建筑物目标特征不够充足，有的小建筑物目标经过多层卷积后语义信息可能产生丢失，无法通过这种特征融合的方式传递到浅层网络中；二是提取的特征中含有噪声干扰；三是有限的小建筑物目标特征没有得到有效利用。因此需要从根源上也就是特征提取网络中学习到更多小建筑物目标特征信息。

为改善上述问题，本文将从以下方面对特征提取网络进行改进。第一，为扩大网络感受野，增加上下文信息，引入多尺度检测方式，使用 Res2Net 网络替换原始的 ResNet 网络，以此增加小建筑物目标特征。第二，为减少网络中学到的噪声特征信息，本文使用可变形卷积替换 Res2Net 基本构建块中的 3×3 卷积核，使网络更加适应建筑物形变和尺寸变化。第三，针对特征无法有效利用的问题，添加 SGE 空间注意力机制，增加小建筑物目标有用特征信息权重，抑制干扰噪声。

2　特征提取网络改进

2.1　多尺度检测方式

对于建筑物这种尺度和形状变化较大的目标，ResNet 网络中的单一 3×3 卷积核缺乏多尺度特征，感受野比较局限，不利于模型对全局信息的学习。对于小建筑物这种本身在图像中占像素点很少的目标，如果无法提供足够的特征信息，容易导致漏检误检。因此需要在网络中引入多尺度检测，扩大网络的感受野，补充不同尺度下的特征信息，辅助小建筑物目标检测。

Res2Net 网络结构由 Gao 等[15]提出，将原始 ResNet 基本构建块中的 3×3 卷积核改进为一种分组级联结构，计算公式如下：

$$y_i=\begin{cases}x_i, & i=1\\ K_i(x_i), & i=2\\ K_i(x_i+y_{i-1}), & 2<i\leqslant s\end{cases} \tag{3}$$

对原结构中 1×1 卷积输出的特征图按通道均分为 s 组子集，每组分别表示为 x_i，$i\in(1,2,\cdots,s)$，通道数为输入特征图通道数的 $1/s$，x_1 不做操作，x_2-x_s 每个分支内都有一个 3×3 的卷积核 $K_i()$，输出为 y_i。结构内使用类残差方式连接这些特征分支中的 3×3 卷积核 $K_i()$，多个尺寸的卷积层并行对特征图进行特征提取，再将提取到不同尺度特征的特征图拼接到一起，这样网络就同时包含了多种感受野模块，增大了感受野范围，可以充分提取图中各种尺度建筑物的特征，获得更多的建筑物特征信息。

因此本文将 FCOS 特征提取网络 ResNet 中基本构件单元中的 3×3 卷积替换为 Res2Net 的类残差结构，增加建筑物的上下文信息，补充更多有助于建筑物检测的特征。本文在实验中采用的基本结构块如图 6 所示。理论上 s 值设置越大，可以获得的粒度特征越细，经过实验对比，本网络中 s 取值为 4 效果较好。通过 Res2Net 这种类残差结构处理后，

原本只有 3×3 的感受野信息增加为 3×3,5×5,7×7 三种多尺度特征信息,它们包含了不同尺度下不同空间位置的上下文信息,最后输入到一个 1×1 的卷积层中来更好地融合前面不同尺度的信息。

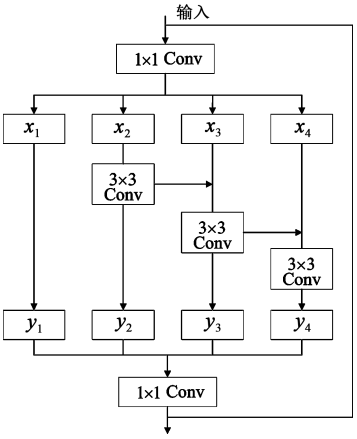


图 6 Res2Net 基本结构块

2.2 可变形卷积

引入 Res2Net 网络后可以补充建筑物周围的上下文特征信息,增加小建筑物目标特征,但是由于卷积单元采样尺寸和形状的固定性,导致同一层特征图中不同尺寸和形变的物体采用相同大小的感受野,不利于小目标建筑物检测:一方面固定位置的采样会导致网络提取特征比较局限单一,另一方面也会增强小建筑物目标周围的噪声信息,影响检测效果。因此有必要对不同尺度的物体采用自适应形状、尺度的感受野大小。故此本文在 Res2Net 的基础上引入了可变形卷积。

可变形卷积[16-17]在原来的卷积核基础上,对每个元素额外学习方向参数,使得卷积核不再是固定的尺寸形状,可以根据目标形状自适应地调整卷积核学习范围,增强模型的泛化能力。可变形卷积具体工作过程示意图如图 7 所示。图 7(a)中红色格子区域为标准 3×3 卷积的规则采样格点 R,采样格点范围为

$$\{(-1,-1),(-1,0),\cdots,(1,-1),(1,0),(1,1)\} \tag{4}$$

输出得到的特征矩阵为

$$y(p_0)=\sum_{p_n\in R} w(p_n)\cdot x(p_0+p_n) \tag{5}$$

其中,$y(p_0)$为每个采样位置的输出特征值,$w(p_n)$为卷积权重,$x(p_0+p_n)$为输入特征值,p_n 为感受野区域内的所有采样位置。

图 7(b)为可变形卷积的示意图。首先卷积提取图片中的特征获得特征图,然后通过一个额外的卷积层来学习 x,y 两个方向的偏移量,维度为 $2N$,偏移量为 Δp_n,表示为

$$\{\Delta p_n \mid n=1,\cdots,N\},N=|R| \tag{6}$$

获得偏移量后,卷积核的大小和位置可以根据建筑物的形状和尺寸自动进行动态调整,可变形卷积核得到的输出矩阵公式为

$$y(p_0)=\sum_{p_n\in R} w(p_n)\cdot x(p_0+p_n+\Delta p_n) \tag{7}$$

其中,$x(p_0+p_n+\Delta p_n)$为采样偏移位置的输入特征值。

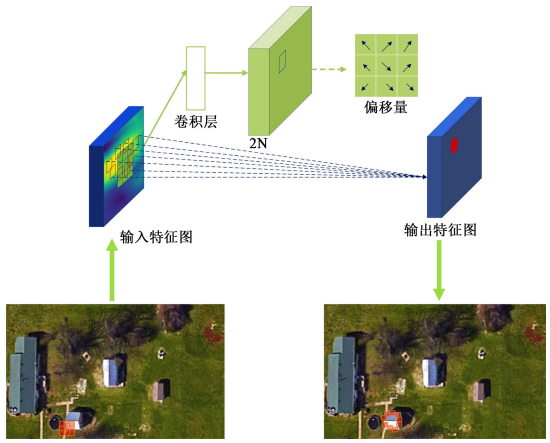


图 7 可变形卷积工作示意图

将 Res2Net 基本构建块里的 3×3 卷积替换为可变形卷积,改进后的网络结构图如图 8 所示。

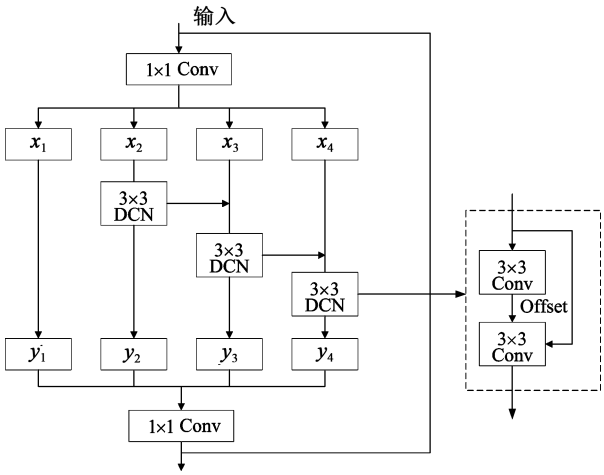


图 8 添加可变形卷积后结构图

2.3 SGE 注意力机制

虽然使用可变形卷积后可以提取到更加丰富的适合于建筑物检测的特征,减少学习环境噪声,但是在检测网络中仍然会学习到噪声特征,且噪声特征与有用特征权重相同[18],导致有用特征利用率不

高。为进一步抑制噪声特征，本文在网络中引入空间注意力机制并进行改进，根据特征对检测的有用程度调整各部分权重因子，使得网络更关注于有用特征[19]，抑制无用噪声特征，提高有用特征利用率。

SGE 注意力机制[20]先将总特征图按通道分组为很多组子特征，每一组子特征图代表各自不同的语义信息。算法默认对子特征图进行全局池化的结果代表子特征图的总体语义信息。对每组子特征图计算子特征局部与全局的空间关系相似性，用来指导子特征的空间分布，加强有用区域的语义特征，抑制无关噪声。具体流程如图 9 所示。

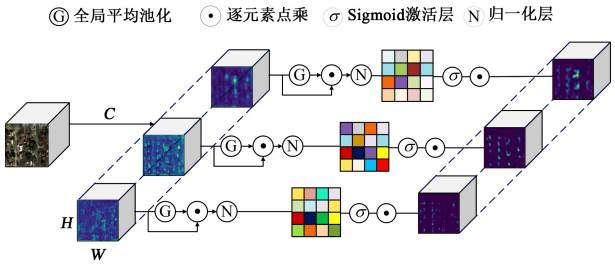


图 9　SGE 注意力原理示意图

首先特征提取网络获取大小为 $H\times W\times C$ 的特征图，按通道数将特征图划分为 G 组，对每个子特征组进行空间注意力操作。子特征组在每个空间位置上都有一个特征向量 x_i，$x_i\in R^{C/G}$，$1<i<m$，m 为特征向量维度。为了获取子特征组的全局特征，对子特征组进行全局平均池化，近似作为这一组的全局特征 g：

$$g=\frac{1}{m}\sum_{i=1}^{m}x_i \tag{8}$$

子特征组的某个空间位置的重要程度 c_i 是通过对组内每个空间位置的特征向量与全局特征向量进行点积的结果来判断的。公式为 $c_i=g\cdot x_i$，子特征图的每个空间位置上的特征向量 x_i 与总体语义特征 g 的点积结果越大，说明此空间位置上的语义信息越重要，贡献度越大，应当增大权重值。而点积结果很小的空间位置则说明可能是一些无关噪声，对于整体无关紧要，应当减小权重值。

为了避免因不同样本之间系数差距过大而影响权重合理分配，需要对权重系数 c_i 进行归一化，然后送入 sigmoid 激活层得到激活输出 α_i，再将原特征向量 x_i 与激活输出 α_i 进行逐元素点乘，获得增强后的组特征向量 $\hat{x}_i$：

$$\hat{x}_i=x_i\cdot\sigma(\alpha_i) \tag{9}$$

为使网络更关注于有用特征，本文将 SGE 注意力机制加入 Res2Net 输出的多尺度特征图后面，以此调整特征图各空间位置语义信息权重，抑制无关噪声。

SGE 注意力机制采用全局平均池化结果来表示全局特征，全局平均池化是对整个特征图求取平均值，通过这种方法可以获取特征图的综合特征，特征鲁棒性更强。但是由于小建筑物目标的像素点较少，噪声特征对于小建筑物目标的影响较大，全局平均池化可能会模糊小建筑物目标部分特征，不利于小建筑物目标检测。相对来说，最大池化可以保留更多的细节特征。因此本文对 SGE 注意力机制进行改进，同时综合全局最大池化信息和全局平均池化信息来获得子特征图的总体语义信息，具体做法为在计算特征图整体特征时，分别对特征图做全局平均池化和全局最大池化，然后计算两种池化方式得到结果的平均值，这样获得的全局特征可以综合特征图的综合特征和细节特征优点，使得网络更加关注对小建筑物目标有用的特征，有利于小建筑物目标检测。改进后的网络结构示意图如图 10 所示。

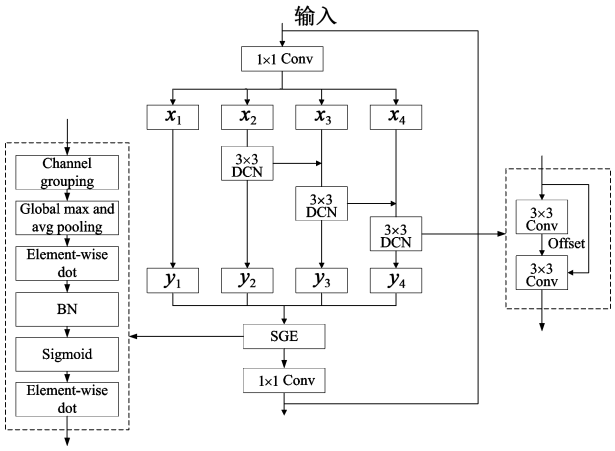


图 10　添加改进 SGE 注意力后的特征提取网络结构示意图

3　实验及分析

3.1　实验环境

软件基于 Ubuntu 系统下的 Pytorch1.6.0 和 mmdetection2.3.0 目标检测框架编写，电脑硬件配置为：Intel i7-7200H @ 2.7 Hz 8 CPU 和 GTX1080Ti GPU。

3.2　数据集介绍

由于缺乏统一公开的数据集评判标准，且网上的大部分相关建筑物数据集形状尺寸单一，本文自己搭建了一组建筑物检测数据集。图片来源为谷歌卫星地图，由于美国区域卫星图较为清晰，所以从地图上选取了 1 200 张美国各地区的航拍建筑物图

像，尺寸为 1 024×768 像素。所挑选图像包含密集、含遮挡及正常情况下建筑物图片，其中选取 120 张作为测试集，剩下用于进行模型训练和验证。数据集标注采用 Labelme 软件，标注格式为 json 文件，然后统一转换为 COCO 数据集格式。

为了更好地拟合模型，增强泛化能力，在训练前对训练数据进行了数据增强操作。一方面进行了普通的亮度、翻转、加噪声等增强，另一方面针对小建筑物目标数量过少的问题，使用 Mosaic 算法将数据集中的图片每次随机选取四张，然后进行翻转、缩放等操作，最后将处理后的四张图片拼接在一起，拼接后的图片示意图如图 11 所示。



图 11 Mosaic 算法拼接后示意图

这样操作不仅可以增加大量的小建筑物目标，还可以模拟建筑物更换环境后的情况，增强模型的表达能力。由上述传统数据增强算法和 Mosaic 算法进行数据增强后，训练和验证集图片扩充为 5 100 张，训练集和验证集按照 8∶1 划分。

分析自己搭建的数据集发现，小型建筑物和中大型建筑物的面积分布大致与图像检测公开数据集中 COCO 数据集对小中大目标的定义相同。因此本文直接采用 COCO 数据集的评判标准计算检测准确率。定义面积 32×32 以下的目标为小目标，面积 32×32～96×96 区间的目标为中等目标，面积大于 96×96 的目标为大目标。准确度采用 AP 值(平均准确度均值)来衡量，对建筑物目标在 0.5～0.95 阈值区间以 0.05 为步长分别计算精确率值，最后求平均值。按照 COCO 数据集评价标准，统计出自建数据集中各尺寸建筑物数量分布，如图 12 所示。

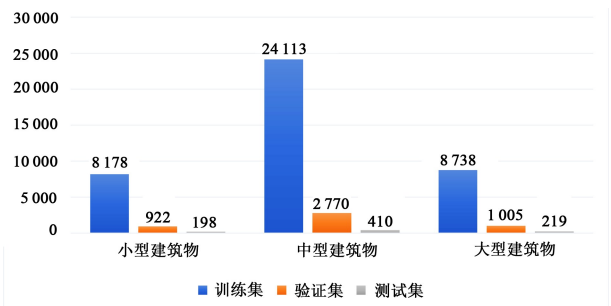
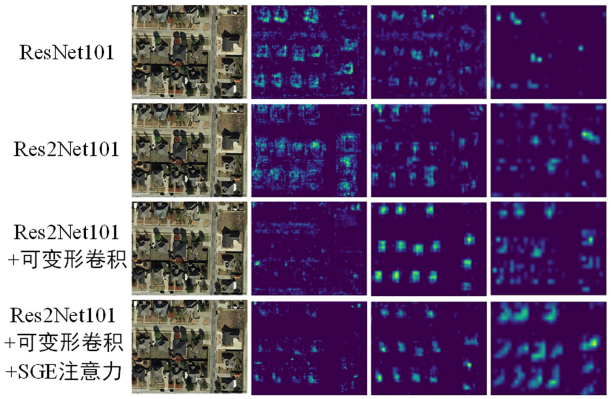


图 12 数据集中建筑物尺寸分布情况统计

3.3 实验结果及对比分析

对改进前后网络的特征图进行可视化，结果如图 13 所示。FCOS 网络是在 FPN 层分层进行目标检测和分类，因此本文可视化了 FPN 中具有代表性的 P3～P5 层，用来对比网络的改进效果。



(a) 原图 (b) P3 层 (c) P4 层 (d) P5 层

图 13 网络改进前后可视化结果对比图

从图 13 可以看出，将 ResNet 网络替换为 Res2Net 网络后，建筑物的特征信息更加丰富，小建筑物目标特征信息增多，有助于对小建筑物目标的检测。但是特征图中含有较多的噪声信息，加入可变形卷积后，网络的噪声信息减少，同时特征也更加具有针对性，但是浅层仍有噪声信息，且过多学到了道路信息。添加改进后的 SGE 注意力机制，让网络更加关注对建筑物检测有用的信息后，各层特征图噪声基本都被抑制掉，同时也含有了更加丰富的特征信息，验证了网络改进的有效性。

实验改进前后对测试集测试效果如表 1 所示，与目前的几种常用目标检测网络测试结果对比如表 2 所示。其中 IOU 代表预测框和图像标注时的真值框的交并比，AP 为整体建筑物在 0.5～0.95 的 IOU 阈值区间每隔 0.05 步长阈值求取的平均精度均值，AP_s，AP_m，AP_l 分别为小型建筑物、中型建筑物、大型建筑物在 0.5～0.95 的 IOU 阈值区间每隔 0.05 步长阈值求取的平均精度均值。表中 DCN 表示可变形卷积，SGE 表示原始 SGE 注意力机制，SGE* 表示改进后的 SGE 注意力机制。

从表 1 中可以看出，改进后的网络对建筑物整体的检测准确率提升了 1.7%，小型建筑物检测准确率提升了 3.6%。这里由于测试集中所包含图片大部分为复杂场景比如建筑物分布密集、树木大面

积遮挡、小建筑物目标分布杂乱等，总体准确率提升幅度较小。由表 2 中可知，改进后的 FCOS 算法相比于常见的几种网络模型，建筑物总体检测准确率和小型建筑物检测准确率更高，检测效果更好。

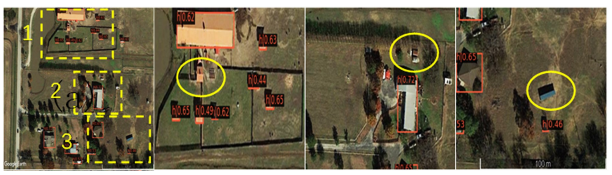
表 1　网络改进前后测试结果对比

网络结构	AP	AP_s	AP_m	AP_l
Resnet101	0.639	0.351	0.652	0.832
Res2net	0.643	0.353	0.669	0.822
Res2net＋DCN	0.644	0.369	0.664	0.818
Res2net＋DCN＋SGE	0.645	0.375	0.669	0.820
Res2net＋DCN＋SGE*	0.656	0.387	0.683	0.822

表 2　与其他网络模型对比测试结果

网络结构	AP	AP_s	AP_m	AP_l
Faster-RCNN	0.632	0.345	0.667	0.812
Mask-RCNN	0.644	0.357	0.679	0.825
RetinaNet	0.637	0.324	0.684	0.796
FoveaBox	0.635	0.331	0.682	0.792
本文网络	0.656	0.387	0.683	0.822

网络改进前测试结果如图 14 所示。图 14(a)，(b)，(c)分别代表含小建筑物目标多、含树木遮挡和建筑物分布密集三种情况。最左列为原图，为了使图片对比效果更加清晰，将检测中标记为 1，2，3 的有问题区域进行局部放大，分别如图中右侧 3 列所示。图 15 为改进后的网络检测结果，图片构成与图 14 相同。对比图 14 和图 15 可以看出网络改进后，



(a)　含小建筑物目标多

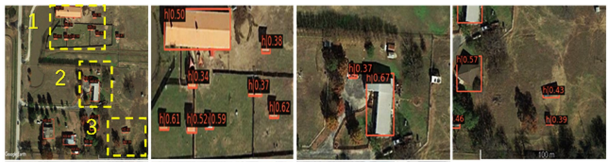


(b)　含树木遮挡



(c)　建筑物分布密集

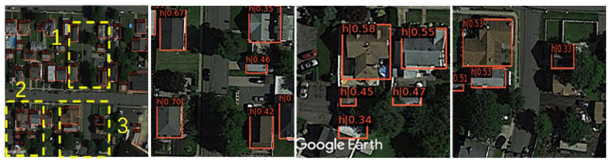
图 14　改进前检测结果



(a)　含小建筑物目标多



(b)　含树木遮挡



(c)　建筑物分布密集

图 15　改进后检测结果

图 14(a)，(b)框出的 1，2，3 区域中漏检的建筑物可以准确检测出来，图 14(c)框出的 1，2，3 区域中被误识别为建筑物的道路，在图 15(c)中没有被错误检测为建筑物，说明网络改进效果良好。

4　总结

本文以在特征提取阶段增加小建筑物目标特征作为出发点，改善小建筑物目标在检测中容易出现的漏检、误检等问题。考虑到无框检测网络中的 FCOS 算法在检测小建筑物目标上的优势，本文主要在 FCOS 网络的基础上进行改进。使用 Res2Net 网络替换原有的 ResNet 网络，增加不同尺度上下文信息，扩大网络感受野，增加小建筑物目标特征信息。同时为了更加适应不同尺寸形状的建筑物，减少学习周围噪声信息，关注有用特征，本文在网络中引入了可变形卷积和改进后的空间注意力机制。改进后网络对小建筑物目标特征提取能力变强，受周围环境噪声影响减弱，在含遮挡和分布密集的情况下，表现效果也很好。最后在自己搭建的数据集上进行实验对比测试，建筑物总体检测准确率达到 0.656，提升了 1.7%，小建筑物目标检测准确率达到了 0.387，提升了 3.6%。

本文通过改进特征提取网络增强对小建筑物目标的特征提取能力，但是网络中其他部分，如特征融合阶段和损失函数等的设计也会影响到小建筑物目标的检测效果。后续工作可以从这两部分入手改进，融合不同特征层对小建筑物目标检测有用的特

征信息,增加对小建筑物目标的关注度,进一步加强目标特征,提升检测准确率。

参考文献:

[1] Liu J, Wang S, Hou X, et al. A deep residual learning serial segmentation network for extracting buildings from remote sensing imagery[J]. Inter. J. of Remote Sensing, 2020, 41(14): 5573-5587.

[2] Femin A, Biju K S. Accurate detection of buildings from satellite images using CNN [C]// 2nd Inter. Conf. on Electrical, Commun. and Computer Engin., 2020.

[3] 顾 炼,许诗起,竺乐庆. 基于FlowS-Unet的遥感图像建筑物变化检测[J]. 自动化学报, 2020, 46(6): 1291-1300.
Gu Lian, Xu Shiqi, Zhu Leqing. Remote sensing image building change detection based on FlowS-Unet [J]. Acta Automatica Sinica, 2020, 46(6): 1291-1300.

[4] 伍广明,陈 奇,Shibasaki Ryosuke,等. 基于U型卷积神经网络的航空影像建筑物检测[J]. 测绘学报, 2018, 47(6): 864-872.
Wu Guangming, Chen Qi, Shibasaki Ryosuke, et al. Aerial image building detection based on U-shaped convolutional neural network[J]. J. of Surveying and Mapping, 2018, 47(6): 864-872.

[5] 田青林,秦 凯,陈 俊,等. 基于注意力金字塔网络的航空影像建筑物变化检测[J]. 光学学报, 2020, 40(21): 47-56.
Tian Qinglin, Qin Kai, Chen Jun, et al. Aerial imagery building change detection based on attention pyramid network [J]. Acta Optica Sinica, 2020, 40(21): 47-56.

[6] 王明常,朱春宇,陈学业,等. 基于FPN Res-Unet的高分辨率遥感影像建筑物变化检测[J]. 吉林大学学报(地球科学版), 2021, 51(1): 296-306.
Wang Mingchang, Zhu Chunyu, Chen Xueye, et al. High-resolution remote sensing image building change detection based on FPN Res-Unet[J]. J. of Jilin University (Earth Science Edition), 2021, 51(1): 296-306.

[7] Su S, Nawata T, Fuse T. Building change detection from bitemporal aerial images using deep learning[C]// ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2020.

[8] 李东子,范大昭,苏亚龙. 结合Faster R-CNN模型的遥感影像建筑物检测[J]. 测绘科学技术学报, 2018, 35(4): 389-394.
Li Dongzi, Fan Dazhao, Su Yalong. Remote sensing image building detection combined with Faster R-CNN model[J]. J. of Surveying and Mapping Science and Technol., 2018, 35(4): 389-394.

[9] Zhang L, Wu J, Fan Y, et al. An efficient building extraction method from high spatial resolution remote sensing images based on improved mask R-CNN[J]. Sensors (Switzerland), 2020, 20(5): 1465.

[10] Cao Z, Wu M, Yan R, et al. Detection of small changed regions in remote sensing imagery using convolutional neural network[C]// IOP Conf. Series Earth and Environmental Science, 2020, 502: 012017.

[11] Yang X, Yang J, Yan J, et al. SCRDet: Towards more robust detection for small, cluttered and rotated objects[C]// 2019 IEEE/CVF Inter. Conf. on Computer Vision (ICCV), 2019.

[12] Fang P, Shi Y. Small object detection using context information fusion in faster R-CNN[C]// 2018 IEEE 4th Inter. Conf. on Computer and Communications (ICCC), 2018.

[13] Zhang G, Wei J. An improved small target detection method based on Yolo V3[C]// 2021 Inter. Conf. on Electronics, Circuits and Information Engineering (ECIE), 2021.

[14] Tian Z, Shen C, Chen H, et al. FCOS: Fully convolutional one-stage object detection [C]// 2019 IEEE/CVF Inter. Conf. on Computer Vision (ICCV), 2020.

[15] Gao S, Cheng M, Zhao K, et al. Res2Net: A new multi-scale backbone architecture [J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2021, 43(2): 652-662.

[16] Dai J, Qi H, Xiong Y, et al. Deformable convolutional networks[C]//2017 IEEE Inter. Conf. on Computer Vision (ICCV), 2017: 22-29.

[17] Zhu X, Hu H, Lin S, et al. Deformable convNets v2: More deformable, better results[C]// 2019 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR), 2019.

[18] Xu Q, Lin R, Yue H, et al. Research on small target detection in driving scenarios based on improved Yolo network[J]. IEEE Access, 2020, 8: 27574-27583.

[19] Das P, Chand S. Attention BuildNet for building extraction from aerial imagery[C]// 2021 Inter. Conf. on Computing, Communication, and Intelligent Systems (ICCCIS), 2021.

[20] Li X, Hu X, Yang J. Spatial group-wise enhance: Improving semantic feature learning in convolutional networks [J]. arXiv e-prints, 2019: 1905-9646.

作者简介:

林思玉(1998—),女,硕士研究生,主要研究方向为数字图像处理;

王敬东(1966—),男,副教授,硕士生导师,主要研究方向为数字图像处理。