测绘科学
Science of Surveying and Mapping
ISSN 1009-2307,CN 11-4415/P

《测绘科学》网络首发论文

题目：基于 Mask-RCNN 的建筑物目标检测算法
作者：李大军，何维龙，郭丙轩，李茂森，陈敏强
收稿日期：2018-11-15
网络首发日期：2019-06-24
引用格式：李大军，何维龙，郭丙轩，李茂森，陈敏强．基于 Mask-RCNN 的建筑物目标检测算法[J/OL]．测绘科学.
http://kns.cnki.net/kcms/detail/11.4415.p.20190621.1604.020.html

基于 Mask-RCNN 的建筑物目标检测算法

李大军[1]，何维龙[2,3]，郭丙轩[2]，李茂森[1]，陈敏强[1]

(1. 东华理工大学 测绘工程学院，南昌 330013；
2. 武汉大学 测绘遥感信息工程国家重点实验室，武汉 430079；
3. 甘肃林业职业技术学院，甘肃 天水，741020)
✉通信作者 何维龙 硕士研究生 E-mail：1205488897@qq.com

摘要 针对在航空影像中，城区 80%的人工目标物为建筑物和道路，建筑物是遥感影像中主要地物的类别，所以建筑物的检测会直接影响到地物提取的自动化水平这一问题。该文提出了一种基于 Mask-RCNN 的建筑物目标检测方法，是基于卷积神经网络思想，在深度学习框架下通过多线程迭代训练，将无人机影像作为训练样本，在卷积神经网络中得到目标特征再通过 区域建议网络（RPN）与 ROIAlign 操作将特征输入不同的全连接分支。最后得到具优化的权重参数的目标检测模型。在不同场景图像中，该模型可以检测出建筑物目标。实验结果达到了预期要求，提高了航空影像中建筑物检测的准确性。

关键词 建筑物目标检测；卷积神经网络；Mask-RCNN；ResNet101 网络；TensorFlow
中图分类号 P231 **文献标志码** A

Building target detection algorithm based on Mask-RCNN

LI Dajun[1], *HE Weilong*[2,3], *GUO Bingxuan*[2], *LI Maosen*[1], *CHEN Minqiang*[1]

(1. Faculty of Geomatics, East China University of Technology, Nanchang 330013, China;
2. State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China;
3. Gansu Forestry Polytechnic, Tianshui, Gansu 741020, China)

Abstract Aiming at the problem that in aerial imagery, 80% of the artificial targets in urban areas are buildings and roads, and buildings are the main types of features in remote sensing images, so the detection of buildings will directly affect the automation level of feature extraction. This paper proposed a building object detection method based on Mask-RCNN. It was based on the idea of convolutional neural network. Through the multi-thread iterative training in the deep learning framework, the unmanned aerial vehicle(UAV) image was used as the training sample. The target features were obtained in the neural network and then the features were input into different fully connected branches through the region proposal network(RPN) network and the ROIAlign operation. Finally, the target detection model with optimized weight parameters was obtained. In different scene images, the model could detect building targets. The experimental results met the expected requirements and improved the accuracy of building detection in aerial imagery.

Key words building target detection; convolutional neural network; Mask-RCNN; ResNet101 network; TensorFlow

0 引言

近年来,随着科学技术的发展，无人机测绘已成为测绘工作的主力军之一，传统的人机交互模式的无人机影像地物提取方法无法满足现今对快速化、智能化、自动化地提取地物的要求,正成为制约无

收稿日期：2018-11-15
基金项目：国家重点研发计划项目（2016YFB0502200）；国家自然科学基金项目（41127901）；测绘遥感信息工程国家重点实验室专项科研经费资助项目
作者简介：李大军（1965—），男，湖南澧县人，教授，博士，主要研究方向为 GIS 理论与应用。E-mail: djli@ecit.cn
网络首发时间：2019-06-24 11:32:51 网络首发地址：http://kns.cnki.net/kcms/detail/11.4415.p.20190621.1604.020.html

人机技术进一步快速发展的瓶颈之一，针对这个问题国内外学者做了大量的研究，主要集中在半自动提取和自动提取 2 个方面。在半自动提取方面，代表性的有 Freeman 编码和 Hough 变换相结合的直角型建筑物半自动提取的方法[1]、基于 Snake 和动态规划提取建筑物轮廓[2]和一种几何约束和影像分割相结合的快速半自动建筑物提取的方法[3]等，这类方法没有很好地实现高效的人机交互。只是对提取过程中的优化计算方法做了改进；对于全自动提取，主要有基于原始激光雷达点云数据使用规则化提取建筑物轮廓的方法[4]、利用多层次特征的建筑提取算法[5]，这类方法的不足是单纯依靠影像是很难分离出建筑物区域和非建筑物区域，并且影像中建筑物周围的道路边界和阴影部分会干扰建筑物的识别效果，在建筑物密集重叠区域识别效果也不能满足实际要求，故寻找准确地检测和提取建筑物的方法需要进一步努力。

目标检测作为图像处理和计算机视觉领域中的经典课题之一，广泛的应用于交通监控、图像检索、人机交互等方面。建筑物目标检测作为图像处理和计算机视觉的一个重要分支，主要研究方法有基于背景建模的方法和基于表观特征信息的方法 2 种[6]。大致过程一般分为 3 步：第 1 步，对图像中的各个区域进行预处理，一般通过颜色、边缘、纹理等特征判断可能是建筑物目标的区域，缩小计算的区域，降低计算的复杂度；第 2 步，采用尺度不变特征变换 [7]、方向梯度直方图[8]、局部二值模式[9]等特征提取方法验证预处理得到的区域是否为建筑物目标；第 3 步，将获得特征向量输入到分类器中进行分类，通常采用的分类器有支持向量机[10]、决策树[11]、迭代器 Adaboos[12]等。上述分类器几乎都需要根据人工设计特征对目标进行特征提取，人为主观因素的影响较大。根据现有研究，主流目标检测算法 R-CNN、Fast R-CNN、Faster R-CNN、Mask-RCNN，其优点为检测结果的精度和速度在不断提升，缺点是需要大量训练数据，无法实现端对端检测，即使倍增训练数据，准确度最多只有 2%～3%的提升，对目标检测框的定位优化能力有限，而且随着现实生活中的建筑物场景也日趋复杂，传统的目标检测方法已经存在瓶颈，无法满足复杂场景和建筑物目标密集情况下的的检测要求。

针对上述问题，本文将深度学习的目标检测算法应用于建筑物检测，同时对传统深度学习目标检测算法的训练数据和特征提取器做了调整。一方面，检测模型的训练主要任务是通过海量的数据来学习目标的特征属性，本文选取低成本、易获取等优点的无人机影像解决了数据源获取问题，而且选取了一定量的同一目标不同角度的倾斜影像，使得模型可以更好地学习目标属性特征，提升了模型的整体性能和泛化能力。另一方面，将最新 Mask-RCNN 算法和 ResNet101 特征提取网络两者结合，以无人机影像作为数据源，在主流的 TensorFlow 深度学习框架进行多线程迭代训练模型，最后通过与主流深度学习目标检测算法的检测结果对比分析。

1 Mask-RCNN 建筑物目标检测框架

1.1 检测框架的设计思路

本文方法流程主要有检测模型的训练和模型测试 2 个阶段，如图 1 所示。

检测模型训练阶段，通过训练样本对具有初始参数的卷积神经网络进行迭代训练，根据训练结果和 Tensorboard 查看训练过程，从而不断修改优化训练模型相关训练参数，最终得到目标检测模型。

测模型试阶段，将待检测样本输入目标检测模型得到检测结果。

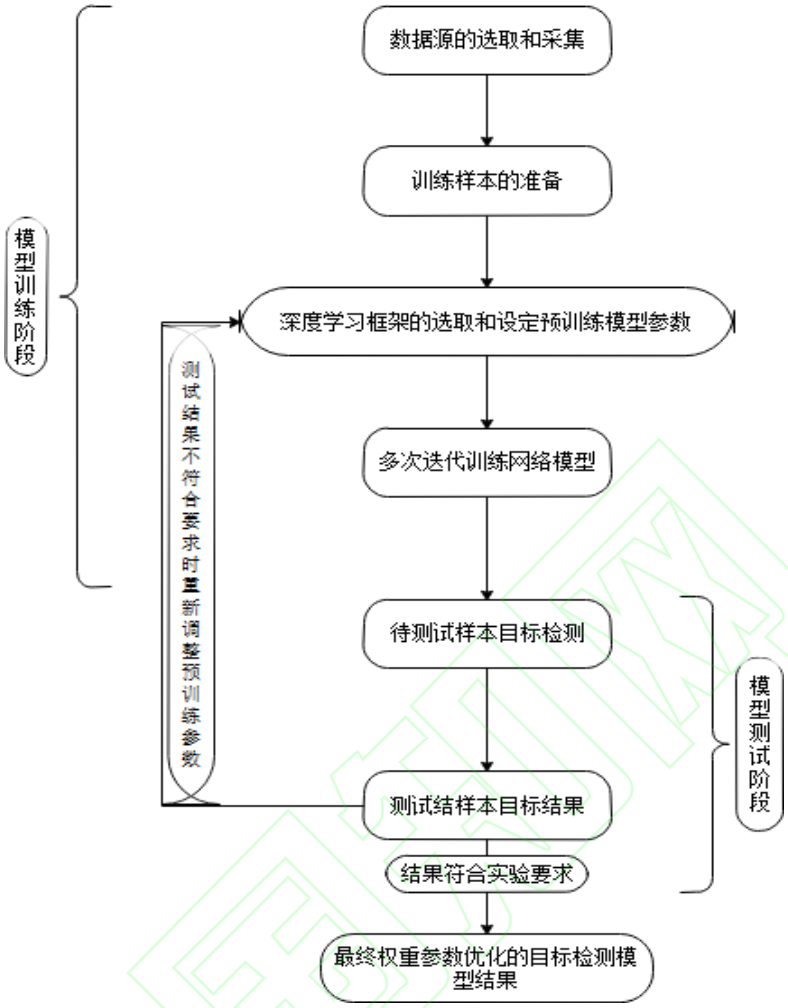


图 1 建筑物目标检测系统框架
Fig.1 Framework of Buildings Target Detection System

本文的目标检测方法整个训练流程包含以下 6 个步骤:

1）根据检测目标特征属性准备训练数据源。

2）对训练样本进行预处理，采用 LabelMe 标注工具进行数据标注和 Mask 掩膜制作

3）将其输入到一个预训练模型的神经网络中（ResNet101）获得对应训练样本图片的特征图，对这个特征图中的每一点设定预定个的 ROI（region of interest），从而获得多个候选 ROI。

4）接着，将这些候选的 ROI 送入区域建议网络（region proposal network，RPN）进行二值分类（前景或背景）和 BB（bounding box）回归，过滤掉一部分候选的 ROI，对这些剩下的 ROI 进行 ROIAlign 操作。

5）最后，对这些 ROI 进行分类（*N* 类别分类）、BB 回归和 MASK 生成。

6）重复步骤 4）～步骤 5），训练完所有样本得到最终优化调整的检测模型。

本文的目标检测方法整个模型检测流程包含以下 2 个步骤:

1）利用测试样本对建筑物目标检测模型进行测试，得到新样本的检测结果。

2）测试结果不符合要求时返回重新调整模型与训练参数，重新进行模型训练；测试结果符合要求时得最终权重参数优化的目标检测模型结果。

1.2　Mask-RCNN 建筑物检测的核心过程

Mask-RCNN 算法的核心过程如图 2 所示。

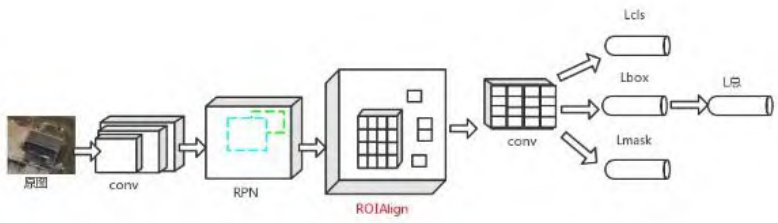


图 2 Mask-RCNN 算法的核心流程图
Fig.2 Core Flow Chart of Mask-RCNN Algorithm

Mask-RCNN 算法步骤：

1）首先，输入一幅需要处理的图片，然后进行对应的预处理操作，或者预处理后的图片；然后，将其输入到一个预训练好的神经网络中获得对应的 feature map。

2）对这个 feature map 中的每一点设定预定个的 ROI，从而获得多个候选 ROI。

3）将这些候选的 ROI 送入 RPN 网络进行二值分类（前景或背景）和 BB 回归，过滤掉一部分候选的 ROI。

4）对这些剩下的 ROI 进行 ROIAlign 操作（即先将原图和 feature map 的 pixel 对应起来，然后将 feature map 和固定的 feature 对应起来）。

5）最后，对这些 ROI 进行分类（*N* 类别分类）、BB 回归和 MASK 生成（在每一个 ROI 里面进行 FCN（fully convolutional networks）操作）。

1.2.1 RPN

RPN 基本原理与搜索检测方法（selective search）相似，将影像输入网络，输出不同的矩形候选区域，但网络效率更高。RPN 网络是基于卷积神经网络结构，不同的是它的输出是包含了二类 softMax 函数和边框回归的多任务模型。具体实现过程：首先用一个 $n\times n$ 的滑窗（n = 3，即 3×3 的滑窗）在 conv 5-3 的卷积特征图上生成一个维度为 256（对应于 ZF[13]网络）或 512（对应于 VGG 网络）维长度的全连接特征，然后利用产生的 256 维或 512 维的特征生成 2 个全连接层分支：

1）回归层（reg layer）。作用是预测候选区域的中心锚点对应的候选区域的坐标 x，y 和宽高 w，h。

2）分类层（cls layer）。作用是判定该候选区域属于前景还是背景。

RPN 网络的损失函数的计算，它是由 softMax loss 和 regression loss 两者按一定比重组成的。softMax loss 是通过映射（anchors）对应的背景标定结果和预测结果计算得到，regression loss 的计算需要以下 3 组信息计算得到。

1）预测框。RPN 网络预测出的候选区域中心位置坐标 x，y 和宽高 w，h。

2）锚点参考盒（reference boxes）。9 个锚点对应 9 个不同尺寸和宽高比的参考盒，9 个参考盒都有对应的中心点位置坐标 x_a，y_a 和宽高 w_a，h_a。

3）正确标注（ground truth）。标定的框对应的中心点位置坐标 x^*，y^* 和宽高 w^*，h^*。

综上，计算 regression loss 和总 Loss 公式如下：

$$t_x = \frac{(x - x_a)}{w_a} \quad t_y = \frac{(y - y_a)}{h_a} \tag{1}$$

$$t_w = \log(w/w_a) \quad t_h = \log(h/h_a) \tag{2}$$

$$t_x^* = \frac{(x^* - x_a)}{w_a} \quad t_y^* = \frac{(y^* - y_a)}{h_a} \tag{3}$$

$$t_w^* = \log(w^* / w_a) \quad t_h^* = \log(h^* - h_a) \tag{4}$$

$$L(\{p_i\},\{t_i\}) = \frac{1}{N_{reg}}\sum_i L_{cls}(p_i, p_i^*) + \lambda\frac{1}{N_{reg}}\sum_i L_{cls} p_i^* L_{reg}(t_i, t_i^*) \tag{5}$$

1.2.2 ROIAlign 操作

采用 ROIAlign 来优化 RoIPooling 带来的空间位置错位问题，采用新的插值方法，先通过双线性插值到 14×14，再池化到 7×7，前者使用了两次量化操作，而后者并没有采用量化操作，使用了线性插值算法。ROIPooling 区域提案是对原图进行兴趣区域提取，再用 ROIPooling 找到卷积产生的特征图对应位置提取特征小块，通过计算原图中的位置除以在卷积过程中的步长乘积，再取整，这样找到对应区域的特征小块就因为取整导致对不齐。因此 ROIAlign 不对计算结果取整，而是通过双线性插值确定原图兴趣区域中每个点的特征值，再进行池化等操作来提升精度，为池化过程直接采样带来的偏差对齐问题提供了新的解决方法。

图 3 中虚线框为 5×5 的特征图，实线框为映射到特征图上的 ROI 区域，要对该 ROI 区域做 2×2 的池化操作。首先把该 ROI 区域划分 4 个 2×2 的区域，然后在每个小区域中选择 4 个采样点和距离该采样点最近的 4 个特征点的像素值图中黑色小方格的 4 个角点 1、2、3、4，通过双线性插值的方法得到每个采样点的像素值；最后计算每个小区的，生成 ROI 区域的 2×2 大小的特征图。

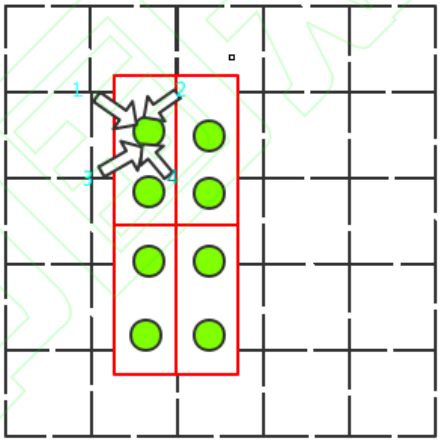


图 3 ROIAlign 原理
Fig.3 ROIAlign Principle

1.2.3 特征金字塔网络

将新提出的特征金字塔网络（feature pyramid networks，FPN）引入算法中，实现对不同尺度下的特征有效利用，如图 4 所示，FPN 采用了自上而下的侧向连接将不同尺度的特征连接融合（上采样后相加）起来，再进行 3×3 的卷积以消除混叠现象，而后在不同尺度的特征上进行预测，不断重复这个过程，直到得到最佳的分辨率。FPN 优点在与它可以在不增加计算量的同时，提升对多尺度下小物体的精准快速检测能力。

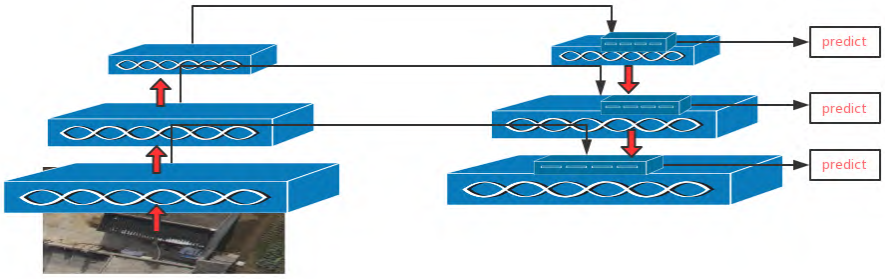


图 4 FPN 特征金字塔
Fig.4 FPN Feature Pyramid

1.2.4 多任务 Loss

$$L = L_{cls} + L_{box} + L_{\text{Mask}} \tag{6}$$

式中：L_{cls}、L_{box} 和 Faster R-CNN 中定义的相同。

对于每一个 ROI，Mask 分支有 $Km \times m$ 维度的输出，其对 K 个大小为 $m \times m$ 的 Mask 进行编码，每一个 Mask 有 K 个类别。使用了像素级 Sigmod 激活函数，并且将 L_{Mask} 定义为平均二值交叉熵损失函数（the average binary cross-entropy loss）[13]。对应一个属于正确类别中的第 k 类的 ROI，L_{Mask} 仅仅在第 k 个 Mask 上面有定义（其它的 K-1 个 Mask 输出对整个 Loss 没有贡献）。本文定义的 L_{Mask} 允许网络为每一类生成一个 Mask，不会与不同类产生竞争，同时依赖于分类分支所预测的类别标签来选择输出的 Mask。可以将分类和 Mask 生成分解开来。而 FCN 语义分割通常是用一个像素级 Sigmod 激活函数和一个多项交叉熵损失值，此时 Mask 之间存在竞争关系；本文的算法是一个像素级 Sigmod 激活函数和一个二元损失值，不同的 Mask 之间没有竞争关系，L_{Mask} (Cls_k) = Sigmoid (Cls_k)，通过逐像素的 Sigmoid 激活函数计算得到。经验表明，通过对每个类别对应一个 Mask 可以有效避免类间竞争（其他目标类别不贡献 Loss 值 ），这可以提高实例分割的效果[14]。

如 5 图所示，首先得到预测分类为 k 的 Mask 特征，然后把原图中边界框包围的 Mask 区域映射成 $m \times m$ 大小的 Mask 区域特征，最后计算该 $m \times m$ 区域的平均二值交叉损失熵。

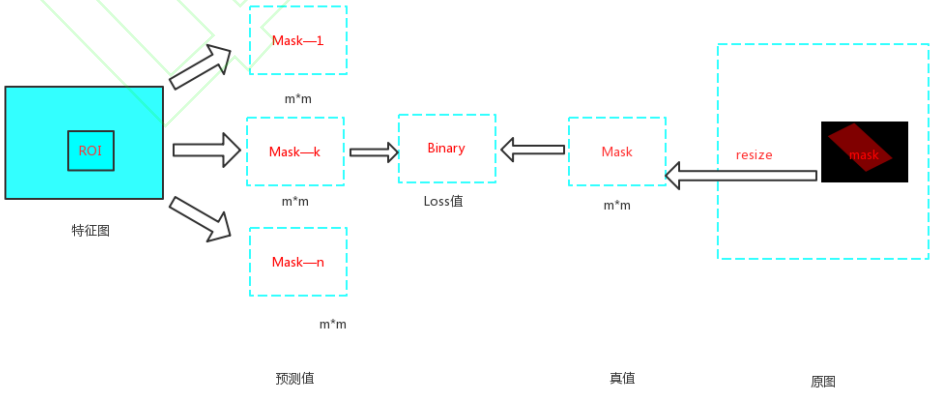


图 5 多任务 Loss 原理
Fig.5 Multitasking Loss Principle

2 传统 Mask-RCNN 目标检测算法的改进

Mask-RCNN 目标检测选用了经典的目标检测算法 Faster-RCNN 和经典的语义分割算法 FCN[15]。Faster-RCNN 可以既快又准地完成目标检测的功能；FCN 可以精准地完成语义分割的功能。不但可以获得建筑物目标检测的高准确率，又可以获得待检测影像中目标像素级分割的 Mask 掩膜，可以地获取目标轮廓。其次，算法思路简单，只在原始 Faster-RCNN 算法的基础上面增加了 FCN 来产生对应的 Mask[17]分支。简化可以看作 RPN + ROIAlign + Fast-RCNN + FCN[16]。使得算法更容易理解和调整，简化了检测过程。最后，Mask-RCNN 算法非常的灵活，可以用来完成多种任务，包括[18]目标分类、目标检测、语义分割、实例分割、人体姿态识别等多个任务。提升了检测算法的泛化能力，为今后不断优化改进提供了更多可能。

深度学习的目标检测算法在发展过程中也有很多制约和不足[7]。一方面深度学习模型的训练需要大量的数据样本来实现，而传统的数据源大多为遥感卫星影像、监控影像和网上公开的深度学习数据集[11]等，特定检测目标的数据获取十分困难，成本大。传统航空遥感平台及传感器的限制，普通的航空摄影测量[19]手段在获取小面积、大比例尺数据方面存在成本高、性价比差等问题。另一方面，不同的检测目标在影像中的分布规律，形状大小，排列方式和纹理信息都千差万别，而深度学习主要是对检测目标的纹理特征进行学习训练，传统的数据多为遥感影像以及图片，不能获取同一目标在多角度多方位下的全面丰富的纹理特征，不能很好解决判别面偏向，影响模型的性能[16]。最后，随着卷积网络的不断发展，网络的深度越来越深，特征学习能力也不断提升，但是当网络层数达到一定的数目以后，网络的性能就会饱和，再增加网络的性能就会开始退化[20]，训练精度和测试精度都在下降。

针对上述问题，本文在继承 Mask-RCNN 算法的优势性能的同时针对建筑物检测目标的特殊性和模型性能的考虑，做了以下调整与改进。

2.1 数据源的选取

近年来地理空间信息技术取得了飞速的发展，尤其是灵活机动、具有快速响应能力的轻小型航空，更是在最近几年迅速成长，成为航空遥感领域一个引人注目的亮点。本文实验数据源采用低成本和机动灵活等诸多优点的低空无人机遥感在小区域内快速获取高质量遥感影像[11]。低空无人机遥感系统，作为卫星遥感与普通航空摄影不可缺少的补充，特有优势在：

1）无人机拍摄区域可以根据需要进行规划，节约成本。

2）无人机航行时的高度较低，不受云层影响，且受天气影响较小。

3）可以快速获取分辨率很高的影像，并且可以调整航高、重叠度获得不同尺度下的影像。

4）起飞条件要求低，无需专门跑道。

2.2 训练样本增强

建筑物作为本文主要的检测目标，建筑物在不同区域的无人机影像中的分布规律多样，建筑物形状多变，多为不规则的多边形，房屋的排列方向也不同，为了防止模型在训练过程中出现过拟合现象，本文的数据增强方式除了传统的图像旋转平移、随机修剪、色彩抖动、平移变换、尺度变换、对比度变换、噪声扰动方法以外，充分结合无人机影像数据源的特点增加了新的数据增强方式，方法如下：

1）同一建筑物目标的特征增强。为了使得训练模型更好地学习建筑物丰富的特征、解决判别面偏向问题，提升模型整体性能，本文选取了无人机倾斜影像作为数据源的一部分，对同一建筑物目标可以从不同角度获取纹理特征（图 6），模型更好的提取与学习建筑物目标特征。

（a）垂直角度 （b）上倾斜 45° （c）左倾斜 45° （d）右倾斜 45°

图 6 无人机不同角度倾斜影像
Fig.6 Different Angles Tilt Images of Unmanned Aerial Vehicle(UAV)

2）采取类别不平衡数据的增广（label shuffling）的类别平衡策略。

3）选取不同地区不同季节不同光照条件下的影像数据，如图 7 所示。

（a）城区 （b）郊区 （c）冬季郊区

图 7 不同区域和季节无人机影像
Fig.7 Different Area and Season Images of UAV

4）图片随机顺序生成。在数据增强过程中同一建筑图片会连续出现，会导致模型在训练过程中对同一建筑的特征进行连续学习，出现过拟合现象，因此本文将数据样本顺序打乱后在进行随机排序，从而在训练过程中达到提升模型性能的作用。

2.3 特征提取器的选取

深度残差网络是 2015 年提出的深度卷积网络，在 ImageNet 中斩获图像分类、检测、定位三项的冠军[21]。它很好地解决了网络深度与性能退化的问题，本文综合建筑物目标特征的特殊性和其他特征提取器的适用性，使用特征表达能力更强的残差网络，在硬件基础上选取了 ResNet101 作为特征提取器，该网络就是在传统 AlexNex 网络的基础上通过增加网络深度大幅度提高了网络性能，有更好的学习和表达能力。传统的卷积神经网络的的特征提取器有很多种，针对不同检测目标的特征属性可以选择不同的特征提取器，但由于本文的主要检测目标是建筑物，无人机影像中的建筑物分布规律，形状大小，排列方式和纹理信息都千差万别，而深度学习主要是对检测目标的纹理特征进行学习训练。为了更好的处理在合适自身数据量的训练样本的同时又可以使得训练模型充分的学习到建筑物目标特征的问题。本文通过将 ResNet + FPN 网络结构来提取目标的特征，不但在传统 AlexNex 网络的基础上通过增加网络深度大幅度提高了网络性能；而且使得网络结构更加简明，模块化；网络中的手动调节的超参数少，方便训练。

3 实验与结果分析

3.1 深度学习框架与预训模型的选取

深度学习的学习框架有很多 TensorFlow, Theano, Caffe, Keras, MXNet, Scikit-learn 等，而本文选取了最为热门的 Google 在 2015 年 Google Research Blog 宣布推出新一代人工智能学习系统——TensorFlow，它是一个开源的异构分布式系统上的大规模机器学习框架，移植性好，支持多种深度学习模型。同时，TensorFlow 的速度相比前代 DistBelief 有了很大的提升，在一些跑分测试中，TensorFlow 的得分是第一代系统的两倍[20]。深度学习的模型训练需要大量时间和海量训练样本来保证，本文考虑自身条件和硬件水平，在开始的研究和探索阶段，本文采用网上公开的官方 COCO2014 数据集进行了预训练得到了该数据集的训练模型。COCO 数据集中的有近 9 000 张图片，数据集包含了自然图片以及生活中常见的目标图片，背景比较复杂，目标数量比较多，大量图片样本中有建筑物目标，因此本文采用迁移学习方法将官方 COCO2014 数据集训练得到的权重模型作为本文建筑物检测算法模型的预训练模型，在此预训练模型的基础上通过自己建立的训练样本集再进行样本训练，以

此通过迁移学习的方式不但可以减少训练人力物力成本、提升训练效率，而且能有效地提升检测模型的整体检测精度和模型性能。

3.2 训练平台、初始参数的选取与数据集的准备

本文所有实验均在同一台服务器上，采用 GPU 编程实现。实验中使用 windows 下 GTX 1080 TI 显卡（12 GB 显存）。实验参数设置如下：初始学习率 0.000 01，迭代 2 000 次后为 0.000 001，momentum 系数为 0.9，权重衰减系数为 0.000 5 正则化为 0.001 6。

为了使训练模型的检测性能更好，通过数据的增强和预处理之后，训练数据集选取了垂直和 4 个倾斜（倾斜 45°）的不同方向选取,城区大多为正射影像，郊区采用不同季节和不同区域的部分正射和倾斜影像，见表 1。

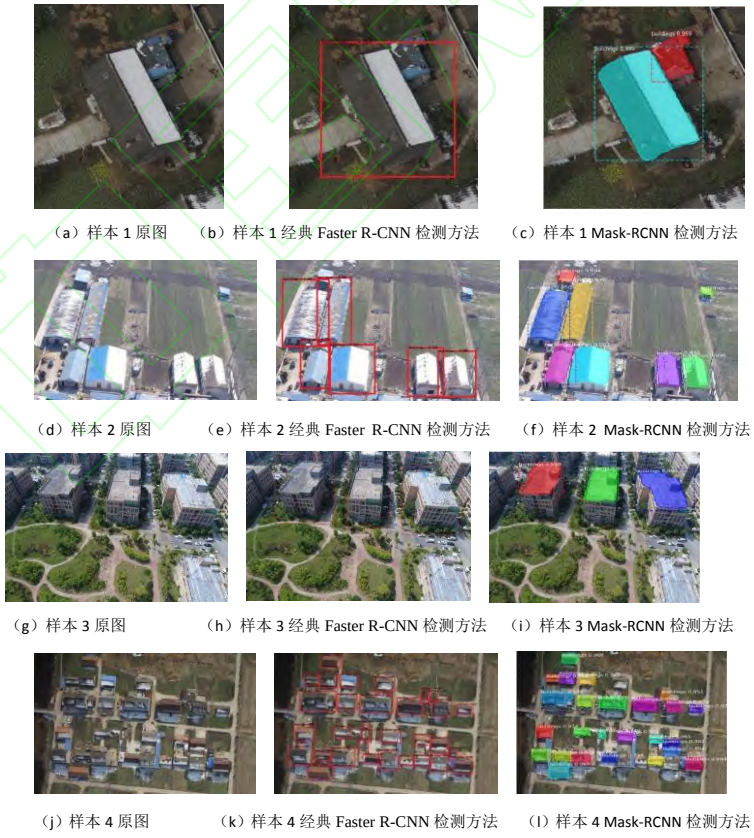
表 1 房屋样本训练数据集

Tab.1 Training Data Set of House Sample Training Data Set

拍摄方向	数据总数/张	区域	分辨率/dpi
垂直方向	5 000	城区	
上倾斜 45°	500	城区	
下倾斜 45°	500	城区	640×640
左倾斜 45°	500	城区	
右倾斜 45°	500	城区	
垂直方向	3 000	郊区	

3.3 实验结果及分析

此次实验中，为了更好说明实验的泛化性和检测精度，检测数据是在模型训练前在训练集随机抽取和其他地区无人机影像截图，抽取的验证集数量是训练集的五分之一，验证集包括了城区、郊区、单一建筑物、不同角度建筑物以及密集度较高的建筑物图片来展示实验结果，如图 8 所示。



（a）样本 1 原图　（b）样本 1 经典 Faster R-CNN 检测方法　（c）样本 1 Mask-RCNN 检测方法

（d）样本 2 原图　（e）样本 2 经典 Faster R-CNN 检测方法　（f）样本 2 Mask-RCNN 检测方法

（g）样本 3 原图　（h）样本 3 经典 Faster R-CNN 检测方法　（i）样本 3 Mask-RCNN 检测方法

（j）样本 4 原图　（k）样本 4 经典 Faster R-CNN 检测方法　（l）样本 4 Mask-RCNN 检测方法

图 8 不同拍摄视角与不同目标数量建筑物目标检测结果

Fig.8 Test Results of Different Shooting Perspectives and Non-target Number of Building Target

本文实验数据选取了不同的视角和不同密集程度的建筑物图片，如图 8 中所示，矩形框表示检测目标位置，矩形框上的数字表示属于建筑物的概率大小，二值 Mask 表示建筑物的大致轮廓。本文为提升本文检测方法准确度，将模型中的矩形框概率的阈值设置为 0.75。（检测概率小于 0.75 时删除目标矩形框，大于 0.75 保留）一方面减少了网络中确定建筑物目标边界框的计算量，提升计算速度。另一方面防止阈值过高引起过拟合现象。经过测试，通过实验发现把 NMS（non maximum suppression）在 RPN 网络的预测阶段在 proposal layer 使用的非极大值抑制的阈值设定为 0.75 时实验结果较好，本文方法被标注的目标概率均高于 0.884。通过图 8 和表 2 可以看出， Faster R-CNN 检测算法在有建筑物有遮挡情况下和小型建筑物密集区域漏检较多，定位也存在一定的偏差，整体检测精确度不高。本文 Mask-RCNN 检测方法对建筑物密集情况下的小建筑物目标检测结果比 Faster R-CNN 检测效果要好，不但可以获取检测建筑物目标的定位框，而且得到了建筑物的轮廓二值 Mask，为以后进一步改进获取建筑物轮廓边缘提供了可能，整体检测精确度更高。但是通过分析发现，测试样本和训练样本的的相关性会影响到测试结果，上述两种方法在密集的建筑物区域的小型建筑物目标和不完整的建筑物目标均存在不同程度的漏检状况。传统的目标检测是从待检测数据中准确完整的识别和检测出所有目标，因此本文的检测实验也采用了检出率（true positive rate，TPR）、误检率（false positive rate，FPR）、漏检率（loss positive rate，LPR）作为算法性能的评价指标。

检出率表示一个样本集中，被分类器检测的正样本数目（n_P）占目标实际总数（n）的比值：

$$\mathrm{TPR} = \frac{n_P}{n} \tag{7}$$

误检率表示一个样本集中，被分类器错误判定为目标的数目（n_{FP}）占所有被分类器检测的正样本数目（n_P）的比值：

$$\mathrm{FPR} = \frac{n_{FP}}{n_P} \tag{8}$$

漏检率表示一个检测样中，未被分类器检测到的目标数量（n_{L}）占目标实际总数（n）的比值：

$$\mathrm{LPR} = \frac{n_L}{n} \tag{9}$$

从表 3 可以看出神经网络算法在对目标分类识别方面精度都比较好，但是标定框精度上 Mask R-CNN 算法优于 Faster R-CNN 方法，虽然算法的中增加了一个掩膜分支，模型每秒传输帧数虽然有降低，但是相比之下并没有受到很大的影响，本文算法的总体效果较好。在表 4 中的 A 测试集是大量郊区和少量城区的 90° 正射影像，B 测试集主要为城区的 45° 倾斜影像，通过大量实验数据的 AP（average precision）值和 mAP （mean average precision）值对比，AP 值越高，准确度越高，本文的方法准确度要高于 Faster R-CNN 方法 0.06%~0.12%，表明本文的方法检测准确度更高，mAP 是不同视角下建筑物目标检测结果的平均值，代表了算法的检测性能，可以看出本文中的检测算法性能更好。其中 AP 是平均准确率，是一个反应全局的性能指标，在统计学中，准确率召回率（precision recall，PR）曲线下的面积也就是 AP 的值，mAP 是不同种类 AP 值的平均值。综合来看，本文所得到的建筑物检测模型检测效果达到预期要求。

表2 测试结果对比

Tab.2 Test Result Comparison

算法	实际目标数目/个	检测数目/个	漏检数目/个	误检数目/个	正确判断数目/个	错误判断数目/个	TPR/(%)	FPR/(%)	LPR/(%)
Faster R-CNN	44	36	8	4	38	4	81	0.090	18
Mask-RCNN	44	39	5	2	39	2	89	0.045	11

表3 测试结果的精度与时间对比

Tab. 3 Comparison of Test Result Accuracy and Time

算法	分类精度/(%)	标定框精度/(%)	平均运行时间/FPS
Faster R-CNN	98.62	81.98	4.21
Mask-RCNN	99.03	86.25	4.30

注：FPS 是是指画面每秒传输帧数。

表4 算法整体性能对比

Tab. 4 Overall Performance Comparison of the Algorithm

测试集		拍摄方向	数据量/张Mask-RCNNAP 值	Faster R-CNN AP 值	Mask-RCNN mAP 值	Fast R-CNN mAP 值
A 测试集	垂直方向	300	0.988	0.857	0.942	0.880
	垂直方向	200	0.919	0.857	0.957	0.801
B 测试集	不同方向倾斜 45°	200	0.932	0.841	0.913	0.822

4 结束语

本文将建筑物检测与深度学习相结合，以 Mask-RCNN 目标检测方法为基础，选取无人机倾斜影像为数据源，在 TensorFlow 深度学习框架下，将 ResNET101 网络与 Mask-RCNN 算法相结合进行多线程迭代训练模型，最后得到具有训练优化的权重参数的建筑物检测模型。这种方法有效避免了传统目标检测的缺陷，不需要人工设定目标特征值，无人机倾斜影像容易获取成本较低，可以获取同一建筑物目标在不同方位角度的特征，克服了从不同角度下检测不准确的问题，提升了模型学习能力。实验证明了本文所采用的算法和框架的可靠性，实验结果达到预期模型效果，提升了建筑物密集层叠区域小型建筑物的检测和整体的检测精度，不但可以获取检测建筑物目标的定位框，而且获取了建筑物的轮廓的二值 Mask，为以后进一步改进获取建筑物轮廓边缘提供了可能，相比传统的航空影像的建筑物检测方法更加快速化、智能化、自动化。但是在建筑物密集层叠区域建筑物不完整情况下检测效果有待提升。

参考文献

[1] 秦永， 宋伟东． 基于 Freeman 编码的遥感影像直角型房屋半自动提取方法研究[J]． 测绘科学， 2009， 34（6）：203-205．（QIN Yong， SONG Weidong． Semi-automatic extraction of right-angle houses from remote sensing imagery based on Freeman code chain[J]． Science of Surveying and Mapping， 2009， 34（6）：203-205．）

[2] 杨贵宝， 李瑞俊， 高霞． 基于 Snake 和动态规划优化的屋顶轮廓提取算法[J]． 内蒙古大学学报（自然版）， 2015，46（6）：664-671．（YANG Guibao， LI Ruijun， GAO Xia． An algorithm of roof contour extraction based on Snake and dynamic programming optimization[J]． Journal of Inner Mongolia University （Natural Science Edition）， 2015， 46（6）：664-671．）

[3] 张煜， 张祖勋， 张剑清． 几何约束与影像分割相结合的快速半自动建筑物提取[J]． 武汉测绘科技大学学报，2000，25（3）：238-242．（ZHANG Yu， ZHANG Zuxun， ZHANG Jianqing． House semi-automatic extraction based on integration of geometrical constraints and image segmentation[J]．Journal of Wuhan Technical University of Surveying and Mapping， 2000，25（3）：238-242． ）

[4] SAMPATH A, SHAN J. Building boundary tracing and regularization from airborne lidar point clouds[J]. Photogrammetric Engineering & Remote Sensing, 2007, 73(7):805-812.

[5] 吕凤华， 舒宁， 龚龑，等． 利用多特征进行航空影像建筑物提取[J]． 武汉大学学报（信息科学版）， 2017， 42（5）：656-660．（LYU Fenghua， SHU Ning， GONG Yan， et al． Regular building extraction from high resolution image based on multilevel-features[J]． Geomatics and Information Science of Wuhan University， 2017， 42（5）：656-660．）

[6] 尹宏鹏，陈波，柴毅，等．基于视觉的目标检测与跟踪综述[J]．自动化学报，2016，42（10）：1466-1489． （YIN Hongpeng， CHEN Bo， CHAI Yi， et al． Vision-based object detection and tracking： a review[J]． Acta Automatica Sinica，2016，42（10）：1466-1489．）

[7] LUO J,GWUN O.A comparison of SIFT,PCA SIFT and SURF[J]. International Journal of Image Processing,2013,3(4):143-152.
[8] WANG X,HAN T X,YAN S. An HOG-LBP human detector with partial occlusion handling[C]//IEEE,International Conference on Computer Vision. [S.l.]:IEEE, 2010:32-39.
[9] WERGHI N,BERRETTI S, BIMBO A D. The mesh LBP: a framework for extracting local binary patterns form discrete manifolds[J]. IEEE Transactions on Image Processing: a Publication of the IEEE Signal Processing Society,2015,24(1):220-235.
[10] MASSA A,BONI A,DONELLI M. A classification approach based on SVM for electromagnetic subsurface sensing[J]. IEEE Transaction on Geoscience & Remote Sensing,2005,43(9):2084-2093
[11] TSANCARATOS P, LLIA I. Landslide susceptibility mapping using a modified detection tree classifier in the Xanthi Perfection,Greece[J]. Landslides,2016,13(2):305-320.
[12] HE K M,ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[J/OL].2015:770-778[2018-11-15]. http://www.cs.sjtu.edu.cn/~shengbin/course/cg/Papers%20for%20Selection/Deep%20Residual%20Learning%20for%20Image%20Recognition.pdf.
[13] HE K M, GKIOXARI G, DOLLAR P, et al. Mask R-CNN[J/OL]. [2018-11-15]. https://arxiv.org/pdf/1703.06870.pdf.
[14] BAI Baolin, LIU Yusong, ZHANG Chunyan. Accident elements detection based on improved DPM[C]//2017 32nd Youth Academic Annual Conference of Chinese Association of Automation (YAC). [S.l.]: [s.n.], 2017:1094-1097.
[15] HE K M,ZHANG X Y, REN S Q,et al. Spatial Pyramid pooling in deep convolutional neutral networks for visual recognition[J].IEEE Transactions on Pattern Analysis & Machine Intelligence ,2015,37(9):1904-1916.
[16] HAO L C, ZHANG Y, CAO Z M. Building extraction from stereo aerial images based on multi-layer line grouping with height constraint[C]// Geoscience and Remote Sensing Symposium. [S.l.]: IEEE, 2016.
[17] WANG X L, SHRIVASTAVA A, GUPTA A. A-Faster-RCNN: hard positive generation via adversary for object detection[J/OL]. [2018-11-15]. https://arxiv.org/pdf/1704.03414.pdf.
[18] DING C W, LIAO S Y, WANG Y Z, et al. CIRCNN: accelerating and compressing deep neutral networks using block-circulant weight matrices[J/OL]. [2018-11-15]. https://arxiv.org/pdf/1708.08917.pdf.
[19] ZHAO Z Q, BIAN H, HU D, et al. Pedetrian detection base on Faster-RCNN and batch normalization[C]//International Conference on Intelligent Computing. Cham: Springer,2017:735-746.
[20] REN S, GIRSHICK R, GIRSHICK R, et al. Faster-RCNN: towards real-time object with region proposal networks[J].IEEE Transactions on Pattern Analysis & Machine Intelligence. 2017,39 (6):1137-1149.
[21] HE K M, ZHANG X Y, REN S Q, et al. Spatial pyramid pooling in deep convolutional neutral networks for visual recognition[C]//European Conference on Computer Vision. Cham: Springer,2014 :346-361.