



## 一种基于级联神经网络的飞机检测方法

王晓林, 苏松志, 刘晓颖, 蔡国榕, 李绍滋

引用本文:

王晓林, 苏松志, 刘晓颖, 等. 一种基于级联神经网络的飞机检测方法[J]. 智能系统学报, 2020, 15(4): 697–704.

WANG Xiaolin, SU Songzhi, LIU Xiaoying, et al. Cascade convolutional neural networks for airplane detection[J]. *CAAI Transactions on Intelligent Systems*, 2020, 15(4): 697–704.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.201908028>

## 您可能感兴趣的其他文章

### 基于注意力融合的图像描述生成方法

An image caption generation method based on attention fusion

智能系统学报. 2020, 15(4): 740–749 <https://dx.doi.org/10.11992/tis.201910039>

### 基于小样本学习的LCD产品缺陷自动检测方法

An automatic small sample learning-based detection method for LCD product defects

智能系统学报. 2020, 15(3): 560–567 <https://dx.doi.org/10.11992/tis.201904020>

### 基于生成式对抗网络的道路交通模糊图像增强

Enhancement of blurred road-traffic images based on generative adversarial network

智能系统学报. 2020, 15(3): 491–498 <https://dx.doi.org/10.11992/tis.201903041>

### 深度学习的双人交互行为识别与预测算法研究

Human interaction recognition and prediction algorithm based on deep learning

智能系统学报. 2020, 15(3): 484–490 <https://dx.doi.org/10.11992/tis.201812029>

### 一种高效的稀疏卷积神经网络加速器的设计与实现

Design and implementation of an efficient accelerator for sparse convolutional neural network

智能系统学报. 2020, 15(2): 323–333 <https://dx.doi.org/10.11992/tis.201902007>

### 基于生成对抗网络的机载遥感图像超分辨率重建

Super-resolution reconstruction of airborne remote sensing images based on the generative adversarial networks

智能系统学报. 2020, 15(1): 74–83 <https://dx.doi.org/10.11992/tis.202002002>

微信公众平台



关注微信公众号, 获取更多资讯信息

DOI: 10.11992/tis.201908028

# 一种基于级联神经网络的飞机检测方法

王晓林<sup>1</sup>, 苏松志<sup>1</sup>, 刘晓颖<sup>1</sup>, 蔡国榕<sup>2</sup>, 李绍滋<sup>1</sup>

(1. 厦门大学 智能科学与技术系, 福建 厦门 361005; 2. 集美大学 计算机工程学院, 福建 厦门 361005)

**摘要:** 由于旋转角度多样性、极端的尺度差异的影响, 遥感图像中的飞机检测目前仍存在挑战。为了解决旋转和尺度的问题, 目前的策略是将现有的自然场景下的目标检测算法 (如 Faster R-CNN、SSD 等) 直接迁移到遥感图像中。这些算法的主干网络复杂, 模型占用空间大, 难以应用到低功耗和嵌入式设备中。为了在准确率不降低的情况下提高检测速度, 本文提出了一个仅包含 9 层的卷积神经网络来解决飞机检测问题。该网络采用了由粗到细的策略, 通过级联两个网络的方式减少计算开销。为了评估方法的有效性, 我们建立了一个针对飞机检测的遥感数据集。实验结果表明, 该方法超越了 VGG16 等复杂的主干网络, 达到了接近主流检测方法的性能表现, 同时显著降低了参数量并使检测速度提高了 2 倍以上。

**关键词:** 飞机检测; 遥感图像; 级联; 深度学习; 卷积神经网络; 两阶段; 由粗到细; 嵌入式设备

**中图分类号:** TP391.4 **文献标志码:** A **文章编号:** 1673-4785(2020)04-0697-08

中文引用格式: 王晓林, 苏松志, 刘晓颖, 等. 一种基于级联神经网络的飞机检测方法 [J]. 智能系统学报, 2020, 15(4): 697-704.

英文引用格式: WANG Xiaolin, SU Songzhi, LIU Xiaoying, et al. Cascade convolutional neural networks for airplane detection[J]. CAAI transactions on intelligent systems, 2020, 15(4): 697-704.

## Cascade convolutional neural networks for airplane detection

WANG Xiaolin<sup>1</sup>, SU Songzhi<sup>1</sup>, LIU Xiaoying<sup>1</sup>, CAI Guorong<sup>2</sup>, LI Shaozi<sup>1</sup>

(1. Intelligent Science & Technology Department, Xiamen University, Xiamen 361005, China; 2. Computer Engineering College, Jimei University, Xiamen 361005, China)

**Abstract:** Detecting airplanes from remote sensing images remains a challenging task, since the images of airplanes always have the characteristics of multiple rotation angles and severe scale change. In order to solve these problems, the most commonly used strategies are to transfer the existing mainstream object detection algorithms based on natural scenario into the remote sensing images directly, such as Faster R-CNN or SSD. However, the backbones of such networks are generally heavy and occupying large space, which are difficult to be applied to low-power consumption devices or front-end embedded systems. To this end, we designed a simple convolutional neural network architecture with only 9 convolutional layers for airplane detection. Our method adopted a coarse-to-fine strategy by cascading a two-stage network, which further reducing the computation cost of detection. Finally, we built a remote sensing dataset for airplane detection to verify our proposed method. The experimental results show that compared with heavy backbone networks such as VGG16, the performance of our method is close to popular methods, but with much less parameters and more than 2 times higher detection speed.

**Keywords:** airplane detection; remote sensing images; cascade; deep learning; convolutional neural network; two-stage; coarse-to-fine; embedded device

收稿日期: 2019-08-24.

基金项目: 国家自然科学基金项目 (61806172, 41971424); 厦门市海洋与渔业局海洋科技成果转化与产业化示范项目 (18CZB033HJ11).

通信作者: 苏松志. E-mail: ssz@xmu.edu.cn.

遥感图像是人类获取空间信息的一种重要方式, 在军事和民用领域都有着很大的潜在应用价值, 如机场管理、空中交通管制等。飞机检测是遥感图像领域的一类经典问题, 已经被研究了许

多年。但是由于这些图像中的飞机具有旋转角度多样性和尺度多样性等特点,直到今天,该问题仍未被很好地解决。

近年来,深度学习的热潮使得深度卷积神经网络在图像分类、目标检测、目标识别等计算机视觉任务上取得了非常好的效果。主流的基于深度学习的检测方法(Faster R-CNN<sup>[1]</sup>、YOLO<sup>[2-4]</sup>、SSD<sup>[5]</sup>、RetinaNet<sup>[6]</sup>等)已经在自然场景的图像数据集(MS COCO<sup>[7]</sup>、PASCAL VOC<sup>[8]</sup>)上达到了相当惊人的性能表现。一种简单的策略是将这些检测算法直接迁移到遥感数据集上。尽管这种策略能够取得不错的效果,但是仍然存在2个缺点:1)主流检测网络是针对多类别检测任务设计的,这使得它们的主干网络必须足够复杂才能应对特征多样性,而针对单一类别检测任务,网络中的大量参数都是冗余的;2)遥感图像和自然场景图像之间的差异被忽略了。在自然场景下,大部分的消费级摄像头会受限角度问题而只能拍摄到飞机的侧面图像,而在遥感图像中,飞机只存在尺度和旋转角度的差异,我们可以清楚地观察到飞机的结构特征,这使得对飞机的检测变得更加容易。

基于以上两点,提出了一种简单但有效的飞机检测方法:首先,通过第1阶段产生若干个候选窗口;然后,在第2阶段进行更精准的分类和边框回归。本文方法属于两阶段检测器,而使用的网络结构相比于Faster R-CNN更加简单,故命名为RCNN-Lite。相比于主流的检测方法,RCNN-Lite在保持较高检测精度的同时将速度提升了2~3倍。同时,RCNN-Lite的参数量仅有9.4万个,占用存储空间约为2.1 MB,可以轻易地部署到各种嵌入式设备中。

为验证RCNN-Lite的有效性和可靠性,构建了一个新的数据集——AMOYAirplane。数据集包括9 372个飞机实例。除此之外,还为1 583个飞机实例提供了5关键点标注。

## 1 相关工作

目标检测任务可以看作由3个部分组成:候选区域提取、图像分类和边框回归。早期,滑动窗口是候选区域提取的主流策略,并结合HOG<sup>[9]</sup>、SIFT<sup>[10]</sup>等手工设计的特征进行特征提取,最后由SVM<sup>[11]</sup>等算法给出分类结果。这种策略虽然非常直观易懂,但缺点也十分明显:为了不漏检,滑动窗口需要尽可能地穷举所有位置,时间复杂度高。

为了解决滑动窗口的计算复杂问题,出现了以Selective Search<sup>[12]</sup>为代表的基于图像纹理、颜色、边缘等信息的候选区域提取方法。这种方法在保持高召回率的同时极大地减少了候选区域的数量。深度学习兴起之后,Girshick等<sup>[13]</sup>将Selective Search和卷积神经网络结合起来,提出了RCNN,大幅提升了目标检测的精度,开启了目标检测的深度学习时代。Fast RCNN<sup>[14]</sup>通过在最后一层特征图上提取区域特征的策略,极大地降低了算法的时间复杂度。此时,Selective Search方法成了算法流程中最耗时的部分。Faster RCNN使用了锚框的策略,提出了基于卷积神经网络的区域候选网络(region proposal networks, RPN),并使得检测算法成为了端到端的框架。

RCNN系列方法采用了“候选区域提取-每个区域单独分类”的思想,因此被称为“两阶段”检测算法。由于在每个阶段中都会有分类和回归,因此两阶段算法在检测精度上非常优秀,但在检测速度上无法达到实时的要求。Redmon等<sup>[2]</sup>将检测问题视作一个回归任务,提出了YOLO,该方法抛弃了候选区域提取,直接得出检测结果,又被称为“一阶段”算法。一阶段检测算法在速度上非常快,但是检测精度不足。Lin等<sup>[6]</sup>认为问题在于一阶段算法在训练时的正负样本比例极度不平衡,简单的负样本在总样本中的比例非常大,提出了Focal Loss并应用到RetinaNet中改善训练,实现了速度与精度的平衡。

遥感图像目标检测与识别一直是遥感图像领域的重点任务。进入21世纪以来,随着大量高空分辨率、短重访周期的成像卫星的出现,遥感图像的数据量大大增加,极大地推动了该领域的发展。遥感图像与自然场景图像相比,有着图像分辨率高、覆盖区域广、目标多为人造刚体等特征。因此,早期在对遥感图像进行目标检测时,通常需要针对遥感图像的特点引入相应的先验信息。以飞机检测为例,飞机在遥感图像中属于小目标,不易检测,但机场和跑道是遥感图像里比较明显的区域。An等<sup>[15]</sup>采用了两阶段检测的思想,通过先检测机场区域来减少候选区域的数量,再对飞机进行更精细的检测。而Li等<sup>[16]</sup>则针对飞机是具有轴对称结构的刚体的特点,在使用视觉显著性计算生成候选区域的基础上使用对称性检测来确定飞机的位置。深度学习出现之后,由于卷积神经网络具有强大的提取特征能力,将主流的深度学习目标检测算法迁移到遥感图像检测任务中并进行适当的微调成为了主要策略。



如 Hsieh 等<sup>[17]</sup>在 RPN 的基础上提出了具有空间布局结构的 LPN, 有效地减少了候选区域生成的数量并提高了生成质量; Liu 等<sup>[18]</sup>利用多角度的先验框提出了 DRBox 方法, 在检测目标的同时还估计了目标的旋转角度; Yang 等<sup>[19]</sup>提出了一种将全卷积网络 (fully convolutional network, FCN) 和马尔可夫随机场结合起来的飞机检测框架。

## 2 数据集

尽管对飞机检测已经研究了很多年, 但是目前大多数的遥感图像数据集都是针对通用目标的, 如 NWPU-VHR10<sup>[20-22]</sup>包含了 10 类目标, 而 DOTA<sup>[23]</sup>包含了 15 类目标, 这些数据集中仅有一小部分可以用于飞机检测任务。以 DOTA 为例,

该数据集中包含了 2 806 张图像, 但只有 197 张图像包含飞机实例, 其余图像在飞机检测任务中不起作用。

我们建立了一个名为 AMOYAirplane 的新的飞机遥感图像数据集。数据集包括了 96 个主要机场的场景, 共包含 9 372 个飞机实例, 每个实例都是用矩形框进行标注。每个标注由左上、右下两个顶点的坐标值构成。数据集既包含了低分辨率图像, 又包含了高分辨率图像, 共 1 639 张图片, 其中 1 248 张用于训练, 391 张用于测试。除此之外, 还为其中 1 583 个飞机实例提供了五关键点标注, 这可以用于对飞机的进一步研究, 如旋转角度估计任务。数据集的详细对比内容见表 1。图 1 给出了一些 AMOYAirplane 数据集的示例。

表 1 数据集对比

Table 1 Comparison between datasets

数据集	图像分辨率	标注格式	类别	图片数量	实例数量	关键点
DOTA	高	带方向的矩形框	15	197	8 300	×
NWPU-VHR10	较高	矩形框	10	90	757	×
AMOYAirplane	高、低均含有	矩形框	1	1 639	9 372	√



图 1 AMOYAirplane 数据集示例  
Fig. 1 Samples in AMOYAirplane dataset

## 3 检测方法

### 3.1 模型总体介绍

本文提出了一种级联的飞机检测方法。该方法由两阶段组成, 每个阶段都是一个非常简单的卷积神经网络, 如图 2 所示。第 1 阶段通过产生若干个候选区域来粗略地定位, 然后用非极大值抑制 (NMS) 来去除有较高重叠度的窗口。剩下的窗口会在第 2 阶段中进行更精细的分类和边框回归, 然后再次进行非极大值抑制并输出最终结果。和 MTCNN<sup>[24]</sup>类似, 本文把这 2 个网络分别叫作 P-Net 和 R-Net。2 个网络的具体结构如图 3 所示。

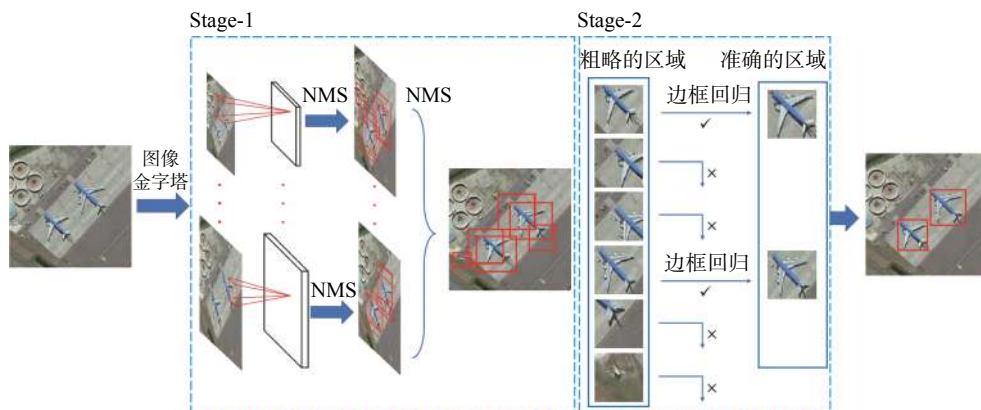


图 2 RCNN-Lite 模型结构  
Fig. 2 Model structure of RCNN-Lite

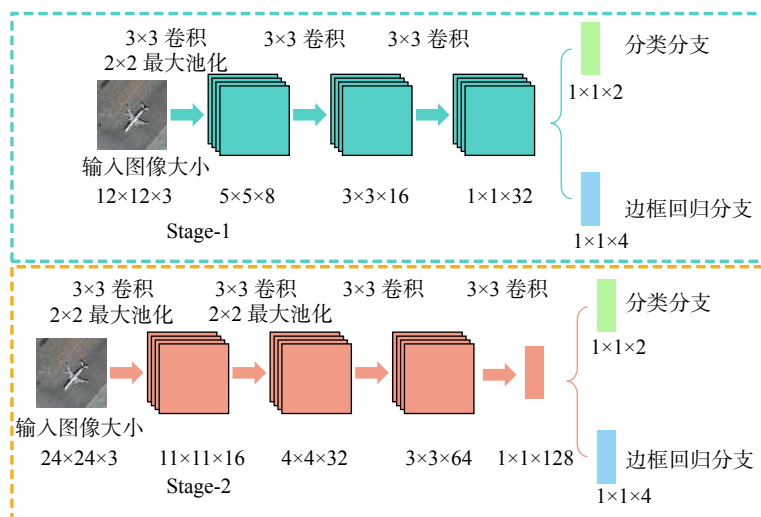


图 3 网络结构

Fig. 3 Network structure

### 3.2 网络具体结构

**P-Net** 该网络负责对目标进行粗略定位。为了使网络能够应对不同尺寸的输入图像, P-Net 采用了全卷积神经网络的设计, 并且由于只包含 4 个卷积层, 网络的参数量也非常小。

给定一张图片, 推理时首先需要对图像建立图像金字塔, 金字塔的每一层都会作为 P-Net 的输入并得到相应的特征图。由图 3 可知, 一个 12 像素×12 像素的区域经过 3 层卷积后尺寸变为 1×1。因此, 一张输入为  $H \times W$  的图片经过 3 层卷积后得到大小为  $\left\lfloor \frac{H-10}{2} \right\rfloor$  像素  $\times$   $\left\lfloor \frac{W-10}{2} \right\rfloor$  像素的特征图, 特征图上的每一点对应输入图像上的一个 12 像素×12 像素的区域。该特征图进一步送入两个并行的 1×1 的卷积得到分类和边框回归结果。分类分支得到的特征图上的每一点代表了对应区域包含飞机的置信度, 回归分支的每一点代表了对应区域的偏移量。根据输入图像的大小和缩放尺度可以计算得到所有候选框, 去除置信度较低的区域并利用偏移量对候选框的位置进行调整, 然后通过非极大值抑制 (non-maximum suppression, NMS) 减少冗余。由于 P-Net 的输入图像是测试图像经过多尺度变换得到的, 输入图像上的 12 像素×12 像素区域会对应测试图像上不同大小的区域, 从而保证 P-Net 可以产生不同大小的候选区域。

最后, 对每一层输出的候选框取并集然后再次进行 NMS 来抑制重复的检测。由于 P-Net 所生成的锚框数量非常大 (一张 800 像素×800 像素的图像会产生大约 32.2 万个锚框, 而采用 ResNet+FPN 结构的 Faster R-CNN 产生的锚框数量大约

为 15.9 万, RetinaNet 大约产生 12 万个), 为了减少计算量, 在进行第一次 NMS 的时候, 只选择置信度最高的 1.2 万个候选区域作为输入, 在输出时只保留置信度最高的 6 000 个候选区域。而在第 2 次 NMS 时, 选择置信度最高的 1 万个候选区域作为输入, 只保留置信度最高的 2 000 的候选区域。

**R-Net** 上一个阶段输出的所有候选区域将会作为 R-Net 的输入, 同样经过分类和边框回归 2 个分支输出结果, 得到置信度和偏移量, 筛除置信度较低的边框并根据偏移量调整边框, 最终通过 NMS 输出检测结果。

由于 R-Net 含有全连接层, 其输入图像大小必须是固定的, 因此必须在输入前将窗口全部调整至 24 像素×24 像素。本文采用了各向异性、各向同性、RoI Pooling<sup>[14]</sup>、RoI Align<sup>[25]</sup> 等 4 种方法进行比较。各向异性方法将候选区域的宽和高分别缩放至 24 像素, 各向同性方法则首先用 0 将候选区域填充成正方形, 然后再缩放至 24 像素×24 像素, 缩放前后候选框区域的宽高比保持不变。RoI Pooling 是 Fast R-CNN 中提出的一种池化方法, 可以将任意大小的区域池化成固定尺寸。RoI Align 利用双线性插值法对 RoI Pooling 进行了改进, 使得池化过程中的位置信息损失更少。实验结果表明 RoI Align 表现最好。

### 3.3 P-Net 和 R-Net 的训练策略

在每个阶段的训练中都要考虑分类和边框回归 2 个任务。

**分类** 对于飞机检测, 该任务是一个二分类问题, 因此本文使用交叉熵损失函数, 对每个样本  $x_i$ , 损失为

$$L_i^{\text{det}} = -(y_i^{\text{det}} \log p_i + (1 - y_i^{\text{det}})(1 - \log p_i))$$

式中:  $p_i$  代表  $x_i$  是飞机的概率;  $y_i^{\text{det}} \in \{0, 1\}$  代表了样本的真实标记。

**边框回归** 在大多数情况下,直接使用目标边框的坐标值进行训练会使得网络难以收敛。R-CNN中使用了边框回归的方法,通过学习预测框和真实框之间的偏移量减少定位误差,同时不容易受到目标尺度的影响。本文方法也采用了这种策略。给定正样本  $x_i$ ,我们选择与  $x_i$  的交并比(intersection over union, IoU)最大的真实边框作为该样本的学习目标,并用四元向量代表  $t_i^{\text{box}}$  真实偏差,  $\hat{t}_i^{\text{box}}$  代表网络预测的偏差,则有,

$$t_i^{\text{box}} = (dx_i, dy_i, dw_i, dh_i)^T$$

$$dx_i = (x_i - \hat{x}_i) / \hat{w}_i$$

$$dy_i = (y_i - \hat{y}_i) / \hat{h}_i$$

$$dw_i = \ln(w_i / \hat{w}_i)$$

$$dh_i = (h_i - \hat{h}_i) / \hat{h}_i$$

式中:  $x_i$ 、 $y_i$ 、 $w_i$ 、 $h_i$  分别是真实边框中心的横、纵坐标和边框的宽、高;  $\hat{x}_i$ 、 $\hat{y}_i$ 、 $\hat{w}_i$ 、 $\hat{h}_i$  则是网络预测框中心点的横、纵坐标以及矩形框的宽、高。该任务的损失函数为

$$L_i^{\text{box}} = \|t_i^{\text{box}} - \hat{t}_i^{\text{box}}\|^2$$

**多任务训练** 每个阶段的总损失函数为

$$\text{Loss} = \frac{1}{N} \sum_{i=1}^N \sum_j \lambda_j L_i^j$$

式中:  $j \in \{\text{box}, \text{det}\}$ ;  $N$  是样本数目;  $\lambda_j$  是平衡系数,使得两种损失函数大约处在同一量级上。

**在线困难样本挖掘 (online hard example mining, OHEM)** Zhang 等<sup>[24]</sup>在 MTCNN 中提出一种 OHEM 方法:将同一批样本按损失值的大小进行排序,只将损失最大的 70% 样本进行回传计算梯度。这样做的原因是样本的损失越小,对参数更新的贡献就越小。而考虑到损失最大的样本有可能是异常点的情况(可能是数据集的标签错误、边框坐标错误或者其他情况),因此对该方法进行了修改:不仅抛弃损失最小的 30% 样本,还抛弃了损失最高的 2% 的样本,仅用剩下的样本损失进行反向传播。

## 4 实验结果及分析

### 4.1 实现细节

在训练过程中, P-Net 的输入样本来自于在训练集图像上的随机裁剪,所有样本会按照与真实边框的 IoU 被分成 3 类: IoU 大于 0.65 的是正样本, 低于 0.3 的是负样本, 在 0.4~0.65 的则被视为

部分样本,其余的则不参与训练。3 种样本的比例控制在 1:1:3 左右,在计算损失函数时,分类任务只使用正样本和负样本,边框回归任务则只使用正样本和部分样本。P-Net 训练完成后,在训练集上进行检测,得到的所有候选区域按照同样的方法划分成正、负和部分样本。在使用 P-Net 生成 R-Net 训练样本时,为了保证样本数量足够,取消了 P-Net 的第 2 次 NMS。

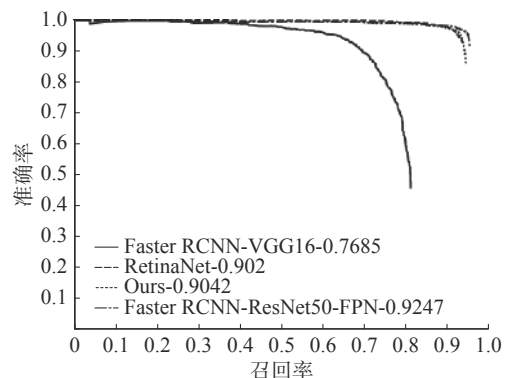
在所有实验中,全部采用随机梯度下降法进行参数更新, P-Net 和 R-Net 均迭代 7 万次,采用了热身策略更新学习率:在最初的 1 000 次迭代中,学习率线性增加到 0.005,随后才开始正式训练。在迭代至 4 万和 6 万次时,学习率衰减十分之一。

### 4.2 实验结果

基于深度学习的目标检测方法可大致分为两类:两阶段检测方法和一阶段检测方法。本文分别选择了两种方法的代表: Faster R-CNN<sup>[1]</sup> 和 RetinaNet<sup>[6]</sup>。Faster R-CNN<sup>[1]</sup> 是目前非常流行的两阶段检测算法,最初该算法使用 VGG16 作为主干网络,目前主流的主干网络通常采用 ResNet+FPN 结构。而 RetinaNet<sup>[6]</sup> 则是一阶段算法中速度和精度都比较均衡的检测器。在所有的实验中, Faster R-CNN 和 RetinaNet 中所有的超参数均使用默认参数。

#### 4.2.1 测试集结果

我们用准确率-召回率曲线和 ROC 曲线来验证本文的方法,如图 4 所示,详细比较见表 2。可以看到,在 mAP 上,本文方法超过了一阶段检测算法 RetinaNet 和以 VGG16 为主干网络的 Faster R-CNN 算法,仅比以 ResNet-50+FPN 为主干网络的 Faster R-CNN 略低 2% 左右。但从模型占用存储空间来看, Faster R-CNN 和 RetinaNet 都占用了数百兆的空间,这使得它们很难直接部署到一些嵌入式设备中,而 RCNN-Lite 仅需 2.1 MB,占用空间极小。



(a) 准确率召回率曲线



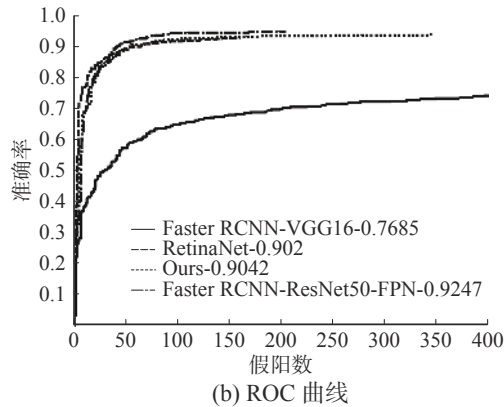


图4 测试集上的准确率-召回率曲线和 ROC 曲线

Fig. 4 PR curve and ROC curve on test set

表2 检测精度对比

Table 2 Accuracy comparison between different methods

方法	主干网络	mAP/%	存储空间/MB
Faster R-CNN	VGG16	76.85	261
	ResNet-50+FPN	92.47	160
RetinaNet	ResNet-50+FPN	90.20	146
RCNN-Lite	—	90.42	2.1

本文同时测试了模型检测速度。作为对照, Faster R-CNN 和 RetinaNet 均使用 ResNet-50+FPN 作为主干网络。测试集图片主要包括了3种不同的分辨率:  $256 \times 256$ 、 $1\,000 \times 900$  和  $4\,800 \times 2\,800$ 。表3给出了模型在不同分辨率输入下的检测速度。为公平起见,所有模型运行都在一张 Titan X (Pascal) 显卡上。可以看出,本文的方法在各种分辨率的输入上都是检测速度最快的,并且领先其他2种方法2~3倍。

表3 检测速度对比

Table 3 Speed comparison between different methods

方法	输入分辨率/ms		
	$256 \times 256$	$1\,000 \times 900$	$4\,800 \times 2\,800$
Faster R-CNN	55	102	819
RetinaNet	29	81	877
RCNN-Lite	17	40	314

#### 4.2.2 P-Net 与 RPN 的对比

RPN 是 Faster R-CNN 中使用的候选区域生成的方法。表4以召回率为指标,比较了 P-Net 和 RPN 提取候选区域的能力。可以看出,在生成候选区域数量较少时, P-Net 和 RPN 的效果基本一致。而生成候选区域数量较多时, P-Net 领先了1.33%。可以认为这是由于 P-Net 使用了更多的锚框(约32.2万个),且可以产生更多种尺度不同

的候选区域,因此能更有效地定位到实例所在的区域。

表4 P-Net 和 RPN 的对比

Table 4 Comparison between P-Net and RPN

方法	生成候选区域数量	召回率/%
RPN	300	92.75
	2\,000	95.12
P-Net	300	92.77
	2\,000	96.45

#### 4.2.3 OHEM 的有效性

为了验证 OHEM 方法的有效性,比较了不使用 OHEM、使用 MTCNN 中的 OHEM 方法以及使用本文中的 OHEM 这3种情况下的模型检测效果,结果如表5所示。可以看出, OHEM 可以使网络在训练期间更关注不容易被正确分类的样本,而本文提出的改进可以让模型在训练过程中忽略某些异常点,进一步提高性能。

表5 不同 OHEM 的效果

Table 5 Comparison between different OHEM methods

方法	mAP/%
无OHEM	88.45
MTCNN	90.26
本文方法	90.42

#### 4.2.4 不同 NMS 的影响

在后处理阶段,通常需要采用非极大值抑制的方法去除冗余窗口,表6给出了采用常规的 NMS 和 Soft-NMS 对检测结果的影响。其中,高斯加权使用的方差  $\sigma=0.5$ 。可以看出,采用 Soft-NMS 相比常规 NMS 可以获得一定的提升,而线性加权和高斯加权两种不同方式带来的差异非常小,可以忽略。

表6 不同 NMS 的效果

Table 6 Comparison between different NMS methods

方法	加权方式	mAP/%
NMS	—	89.9
Soft-NMS	线性加权	90.39
	高斯加权	90.42

#### 4.2.5 损失函数的影响

表7比较了在分类和回归两类任务中,不同损失函数的影响。在分类任务中,使用交叉熵损失和 Focal Loss 的效果基本相同,这是由于 Focal Loss 主要是为了解决训练期间正负样本数量极度

不均衡而提出的,而在 P-Net 和 R-Net 的训练样本生成过程中,我们已经将正负样本的比例控制在 1:3 左右,因此 Focal Loss 没有带来性能提升。而在边框回归任务中,使用 L2 损失得到了最佳性能。

表 7 损失函数的影响

Table 7 Comparison between different loss function

任务	方法	mAP/%
分类	Cross Entropy	90.42
	Focal Loss	90.40
回归	L2	90.42
	L1	88.6
	smooth L1	90.16

## 5 结束语

飞机检测任务是当前遥感图像领域的一个重要研究方向。针对目前的检测算法依赖复杂主干网络的问题,本文提出了一种级联两个简单神经网络来解决飞机检测任务的模型,在检测精度上接近了目前主流方法的水平,而本方法的模型参数量远少于主流方法,在检测速度上也取得了一定的领先。此外,还制作了一个新的遥感图像数据集,该数据集针对飞机检测任务标注了超过 9 000 个飞机实例,能够较好地验证检测算法的有效性。

## 参考文献:

- [1] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [C]//Proceedings of the 28th International Conference on Neural Information Processing System. Montreal, Canada, 2015: 91–99.
- [2] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA, 2015: 779–788.
- [3] REDMON J, FARHADI A. YOLO9000: better, faster, strong-er[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA, 2017: 6517–6525.
- [4] REDMON J, FARHADI A. YOLOv3: an incremental improvement[J]. arXiv preprint arXiv: 1804.02767, 2018.
- [5] LIU W, ANGUELOY D, ERHAN D, et al. SSD: single shot multibox detector[C]//Proceedings of the 14th European Conference on Computer Vision. Amsterdam, The Netherlands, 2015: 21–37.
- [6] 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. Venice, Italy, 2017: 2999–3007.
- [7] LIN T Y, MAIRE M, MAIRE M, et al. Microsoft COCO: common objects in context[C]// Proceedings of the 14th European Conference on Computer Vision. Zurich, Switzerland, 2014: 740–755.
- [8] EVERINGHAM M, GOOL L J V. The PASCAL visual object classes challenge[J]. *International journal of computer vision*, 2010, 88(2): 303–338.
- [9] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. San Diego, USA, 2005: 886–893.
- [10] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. *International journal of computer vision*, 2004, 60(2): 91–110.
- [11] CORTES C, VAPNIK V. Support-vector networks[J]. *Machine learning*, 1995, 20(3): 273–297.
- [12] UIJLINGS J, SANDE K, GEVERS T, et al. Selective search for object recognition[J]. *International journal of computer vision*, 2013, 104(2): 154–171.
- [13] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA, 2014: 580–587.
- [14] GIRSHICK R. Fast R-CNN[C]//Proceedings of the IEEE International Conference on Computer Vision. Santiago, Chile, 2015: 1440–1448.
- [15] AN Z, SHI Z, TENG X, et al. An automated airplane detection system for large panchromatic image with high spatial resolution[J]. *Optik*, 2014, 125(12): 2768–2775.
- [16] LI W, XIANG S, WANG H, et al. Robust airplane detection in satellite images[C]//Proceedings of International Conference on Image Processing. Brussels, Belgium, 2011: 2821–2824.
- [17] HSIEH M R, LIN Y L, HSU W H. Drone-based object counting by spatially regularized regional proposal[C]// Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy, 2017: 4165–4173.
- [18] LIU L, PAN Z, LEI B. Learning a rotation invariant detector with rotatable bounding box[J]. arXiv preprint arXiv: 1711.09405, 2017.
- [19] YANG Y, ZHUANG Y, BI F, et al. M-FCN: effective fully convolutional network-based airplane detection Framework[J]. *IEEE geoscience and remoting sensing letters*, 2017, 14(8): 1293–1297.



- [20] CHENG G, HAN J, ZHOU P, et al. Multi-class geospatial object detection and geographic image classification based on collection of part detectors[J]. *ISPRS journal of photogrammetry and remote sensing*, 2014, 98(1): 119–132.
- [21] CHENG G, HAN J. A Survey on object detection in optical remote sensing images[J]. *ISPRS Journal of photogrammetry and remote sensing*, 2016, 117: 11–28.
- [22] CHENG G, ZHOU P, HAN J. Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images[J]. *IEEE transactions on geoscience and remote sensing*, 2016, 54(12): 7405–7415.
- [23] XIA G S, BAI X, DING J, et al. DOTA: A large-scale dataset for object detection in aerial images[C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Salt Lake City, USA, 2018: 3974–3983.
- [24] ZHANG K, ZHANG Z, LI Z, et al. Joint face detection and alignment using multitask cascaded convolutional networks[J]. *IEEE signal processing letters*, 2016, 23(10): 1499–1503.
- [25] HE K, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]//*Proceedings of the IEEE International Confer-*

*ence on Computer Vision*. Venice, Italy, 2017: 2980–2988.

#### 作者简介:



王晓林, 硕士研究生, 主要研究方向为目标检测。



苏松志, 副教授, 主要研究方向为计算机视觉、机器学习、人脸识别与行人检测。发表学术论文 30 余篇。



李绍滋, 教授, 博士生导师, 主要研究方向为计算机视觉、机器学习和数据挖掘。先后主持或参加多项国家 863 项目、国家自然科学基金项目、教育部博士点基金项目、省科技重点项目等多个项目的研究。发表学术论文 300 余篇。

## 模式识别专委会年度会议暨常委/委员增选会议

为了推动模式识别、机器智能等相关领域的发展, 加强国内学者、校企间的交流与合作, 扩大专委会在国内外的影响力, 中国人工智能学会模式识别专委会(CAAI-PR)将于2020年10月16日在南京国际展览中心召开专委会年度工作会议。会议期间将面向科研院所、高校和企业发展会员及增选委员(包括正式委员及通讯委员), 同时还将从现任委员中增选常务委员。常务委员/委员人选需由个人申请并在工作会议中做个人陈述, 之后由专家评议并投票选出。委员增选人要求为副高职称或以上(中级职称特别优秀也可考虑)。落选委员将自动成为通讯委员。委员、常务委员增选结果将现场公布。在本次年会上, 本专委会将与自动化学会模式识别与机器智能专委会(CAA-PRMI)进行合作, 委员/常委互认, 委员/常委增选联合进行, 并举办联合论坛。

会议时间: 2020年10月16日。

会议地址: 南京国际展览中心。

参会人员: CAAI-PR 现任委员和通讯委员、个人申请者。

参会报名截止时间: 2020年10月10日。

更多详情请见中国人工智能学会微信公众号。