基于 YOLO v2 的无人机航拍图像定位研究

魏湧明，全吉成，侯宇青阳

中国人民解放军空军航空大学航空航天情报系，吉林 长春 130022

摘要 为了保证定位的速度和准确率，采用2016年在目标检测领域取得最佳检测效果的YOLO v2网络制作了以明显特征的地物作为目标区域的目标检测数据集。通过目标框维度聚类、分类网络预训练、多尺度检测训练及更改候选框的筛选规则等方法改进YOLO v2网络，使其更好地适应定位任务。能够将无人机实时获取的航拍图像定位到目标区域，并通过投影关系进行坐标转换得到无人机的经纬度。结果表明：该方法效果较为理想，在航拍图像的目标区域检测任务中检测网络的平均准确率提高到79.5%；在包含目标区域的航拍图像中，经模拟飞行的仿真实验验证，其网络定位准确率大于84%。

关键词 图像处理；卷积神经网络；YOLO v2；目标检测；图像定位

中图分类号 TP183 **文献标识码** A

doi：10.3788/LOP54.111002

Aerial Image Location of Unmanned Aerial Vehicle Based on YOLO v2

Wei Yongming, Quan Jicheng, Hou Yuqingyang

Department of Aerospace Intelligence, Aviation University Air Force, People's Liberation Army of China, Changchun, Jilin 130022, *China*

Abstract In order to ensure the speed and accuracy rate of location, the YOLO v2 network with the best detection effect in the field of object detection in 2016 is used to make the target detection data sets with the obvious features of surface features as the object area. Through the dimension clustering of object box, classified network pre-training, multi-scale detection training, change the candidate box filtering rules and other methods, the YOLO v2 network is improved, and it can better adapt to the location task. The network is able to locate the object area in the aerial image acquired from the unmanned aerial vehicle in real time. And the latitude and longitude of unmanned aerial vehicle are obtained by the projection relationship and coordinate transformation. The experimental results show that the proposed method can achieve better effect, and the average accuracy rate of the detection network increases to 79.5% in the object area detection task of the aerial image. It is verified by simulation experiment of simulated flight, the accuracy rate of the network location is over 84% in the aerial image that contains the object area.

Key words image processing; convolutional neural networks; YOLO v2; object detection; image location

OCIS codes 100.4996; 100.3008; 100.4999; 100.2000

1 引 言

近年来深度学习凭借卷积神经网络(CNN)在提取图像高层特征上的优势[1]，以图像分类任务的成果为基础，在目标检测领域取得了突破性进展。Girshick等[2]提出了区域卷积神经网络(R-CNN)，在VOC 2012数据集上，将目标检测的平均准确率(MAP)提升了30%，达到53.3%。Girshick[3]和Ren等[4]分别提出了快速区域卷积神经网络(Fast R-CNN)和超快区域卷积神经网络(Faster R-CNN)，在提高准确率的同时增

收稿日期：2017-05-22；**收到修改稿日期**：2017-06-16

基金项目：吉林省自然科学基金(20130101069JC)

作者简介：魏湧明(1993—)，男，硕士研究生，主要从事深度学习与图像定位方面的研究。E-mail：2297303678@qq.com

导师简介：全吉成(1960—)，男，博士，教授，博士生导师，主要从事三维可视化技术方面的研究。

E-mail：2297303678@qq.com(通信联系人)

加了检测速度,帧速率可以达到 5 frame/s。Redmon 等[5]提出的 YOLO 真正达到了可以检测视频的速度(45 frame/s)。YOLO 虽然提升了检测速度但牺牲了准确率,并且为今后的研究提供了一种将分类和定位整合到一起的新思路。在 YOLO 基础上,Liu 等[6]和 Redmon 等[7]相继提出的单发射击探测器(SSD)和 YOLO v2 对 MAP 和检测速度的提升取得了令人满意的效果,其中 YOLO v2 得到的效果更好。在 VOC 2007 数据集中进行目标检测过程中,当检测速度为 67 frame/s 时,MAP 达到 76.8%,在目标检测领域可以取得最佳的检测效果。

目标检测与航拍图像定位任务有很多相似之处,都需要对目标区域进行快速和精准的定位。从人类认知事物和判断位置的角度可知,在眼睛能够看到的场景中,人类可以快速发现并定位到目标物体,这也是目标检测过程中需要计算机完成的任务。同样,在飞行过程中,飞行员可以根据地面上熟悉的目标区域大致判断飞机所处位置,也是航拍图像定位需要教会计算机完成的任务。近年来,目标检测技术逐渐成熟,准确率和检测速度均得到显著提升,因此,采用检测效果最佳的 YOLO v2 网络,进行航拍图像定位的研究。

将图像定位的核心问题转化为目标检测问题,选定飞行实验区域,并以该区域中具有明显特征的地物作为目标区域,制作目标检测数据集。以 YOLO v2 网络为主体进一步提出了改进措施,通过目标框维度聚类、分类网络预训练、多尺度检测训练及更改候选框的筛选规则使其更好地适应定位任务,能够在无人机实时获取的航拍图像中定位到目标区域,然后通过投影关系进行坐标转换得到无人机的经纬度。同时加入了对定位到 2 个或 2 个以上目标区域情况的判断,这样目标之间的相对位置关系可以大幅度提高定位准确率。虽然引入倾斜投影形成倾斜图像增加了检测难度,但可以扩大航拍图像收容范围,使单张航拍图像中尽可能多的出现目标区域,进一步提高了定位准确性。

2 YOLO v2 原理

2.1 特征提取网络 Darknet-19

YOLO v2 参考了 YOLO 和 SSD 的网络结构,设计了一个新的分类网络 Darknet-19 作为网络的基础模型。在 YOLO v2 之前,大部分目标检测框架是以 VGG-16[8]作为特征提取网络,但 VGG-16 比较复杂,计算量大。而 YOLO 框架使用了类似 googlenet[9]的网络结构,计算量小于 VGG-16,但准确率略低于 VGG-16。因此,Redmon 设计了兼顾复杂度与准确率的 CNN 来提升网络的检测性能[5]。

最终得出的基础模型为 Darknet-19,包含了 19 个卷积层和 5 个最大值池化层。类似于 VGG-16,网络使用了大量的 3×3 卷积核,经过每一次池化操作(尺寸为 2×2,步长为 2)后,通道数翻倍。借鉴了 Network in network[10]的思想,使用全局平均池化进行网络预测,把 1×1 卷积核置于 3×3 卷积核之间,用来压缩特征并增加网络深度。在每一个卷积层后增加批量归一化操作和去除 dropout 操作,没有出现过拟合现象。Darknet-19 与 Alexnet[11]、VGG-16 的性能对比如表 1 所示,Darknet-19 在 Top-1 和 Top-5 中的准确率分别为 72.9%和 91.2%,高于 Alexnet 和 VGG-16;中央处理器(CPU)和图形处理器(GPU)耗时分别为 6.0 ms 和 0.66 s,时间略长于 Alexnet 和 VGG-16,综合对比可以看出,Darknet-19 性能更优异。性能测试实验中,GPU 为 Titan X,CPU 为 Intel i7-4790K (4 GHz)。

表 1 Darknet-19 与 Alexnet、VGG-16 的性能对比表

Table 1 Performance comparison of Darknet-19 and Alexnet, VGG-16

Model	Top-1/%	Top-5/%	GPU/ms	CPU/s
Alexnet	57	80.3	1.5	0.3
VGG-16	70.5	90	10.7	4.9
Darknet-19	72.9	91.2	6.0	0.66

2.2 YOLO v2 结构及改进

YOLO v2 检测网络以 Darknet-19 为基础模型进行特征提取,并对其网络结构进行相应修改。去掉 Darknet-19 网络的最后一个卷积层,增加 3 个尺寸为 3×3、通道数为 1024 的卷积层,并且在每一个卷积层后加入一个尺寸为 1×1 卷积层,输出维度即检测所需的数量。和 YOLO 相比,该结构移除了全连接层,整个网络均为卷积操作,很好地保留了空间信息,最终得到的每个特征点与原图中的每个 cell 一一对应。并且

借鉴了 Faster R-CNN 中的 anchor 思想：使用 k-means[12] 方法对数据集中的目标框进行维度聚类，确定 anchor 的大小和数量。YOLO v2 中的类别预测不再与每个 cell 绑定一起，而是由 anchor 同时预测类别和坐标。

由于移除了全连接层，模型只包含卷积层和池化层，因此可以随时改变输入尺寸。在训练时，每隔几轮便改变模型输入尺寸，以使模型对不同尺寸图像具有稳健性。每 10 个周期，模型随机选择一种新的输入图像尺寸继续训练。这种训练规则强迫模型适应不同的输入分辨率。模型对于小尺寸的输入处理速度更快，因此 YOLOv2 可以按照需求调节速度和准确率。在低分辨率情况下(288 pixel×288 pixel)，YOLOv2 可以在准确率和 Fast R-CNN 持平的情况下，处理速度达到 90 frame/s。在高分辨率情况下，YOLOv2 在 VOC2007 数据集上的 MAP 可以达到目前最佳效果为 78.6 ，如表 2 所示。

表 2 目标检测框架性能对比表

Table 2 Performance comparison of object detection box

Detection network	Fast R-CNN	Faster R-CNN VGG-16	Faster R-CNN ResNet	YOLO	SSD 300	SSD 500	YOLO v2 288 pixel×288 pixel	YOLO v2 544 pixel×544 pixel
MAP	70	73.2	76.4	63.4	74.3	76.8	69.0	78.6
FPS	0.5	7	5	45	46	19	91	40

3 投影关系与坐标转换

考虑到网络需对倾斜航拍图像进行检测的情况，建立如图 1 所示的成像模型，相机坐标系(x_c,y_c,z_c)以相机的光学镜头中心 S 为原点，z 轴与成像平面垂直，向上为正方向，x 轴和 y 轴分别与成像平面的两条边平行。全局坐标系(X_g,Y_g,Z_g)，使用国际上采用的地心坐标系，以地球质心为坐标原点的 WGS-84 坐标系。

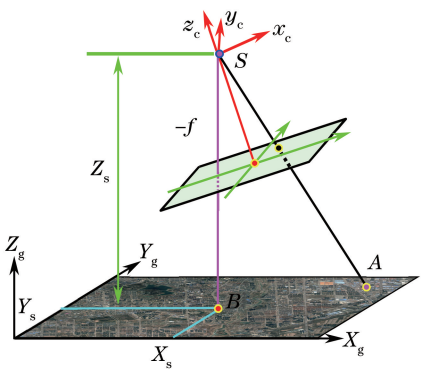

图 1 成像模型图

Fig. 1 Imaging model diagram

成像模型的外方位元素为 3 个角元素(α,ω,κ)和 3 个线元素(X_S,Y_S,Z_S)，分别用来描述相机的空间姿态和光学中心点 S 的空间位置坐标。全局坐标(X_g,Y_g,Z_g)与相机坐标(x_c,y_c,z_c)的转换关系[13]为

$$\begin{bmatrix} X_g \\ Y_g \\ Z_g \end{bmatrix} = \begin{bmatrix} \cos\alpha\cos\kappa - \sin\alpha\sin\omega\sin\kappa & -\cos\alpha\sin\kappa - \sin\alpha\sin\omega\cos\kappa & -\sin\alpha\cos\omega \\ \cos\omega\sin\kappa & \cos\omega\cos\kappa & -\sin\omega \\ \sin\alpha\cos\kappa + \cos\alpha\sin\omega\sin\kappa & -\sin\alpha\sin\kappa + \cos\alpha\sin\omega\cos\kappa & \cos\alpha\cos\omega \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} + \begin{bmatrix} X_S \\ Y_S \\ Z_S \end{bmatrix}, \quad (1)$$

式中 X_S、Y_S 分别为待求解项，Z_S 为已知项代表飞机的飞行高度。

点 A 在相机坐标系中的坐标表示为(x_A,y_A,$-f$)，其中 x_A、y_A 代表点 A 在航拍图像中的位置，f 代表焦距，均为已知项，在全局坐标系中的坐标为(X_A,Y_A,0)，其中 X_A 和 Y_A 代表目标区域的坐标为已知项。点 B 在全局坐标系中的坐标为(X_S,Y_S,0)，根据点 S、A、B 之间位置关系和坐标转换关系为

$$\begin{bmatrix} X_A \\ Y_A \\ Z_A \end{bmatrix} = \boldsymbol{M} \begin{bmatrix} x_A \\ y_A \\ -f \end{bmatrix} + \begin{bmatrix} X_S \\ Y_S \\ Z_S \end{bmatrix}, \quad (2)$$

式中 $\boldsymbol{M}$ 为(1)式中由 3 个角元素构成的 3×3 的坐标旋转矩阵。

综合(1)式和(2)式可以解算出 X_S 和 Y_S,即通过投影关系和坐标转换的计算,可以由目标区域中心的坐标得到机载相机的坐标,从而得到无人机质心的坐标。

4 改进方法

虽然 YOLO v2 已经取得了最佳的检测效果,但其并不完全适合图像定位任务,因此,针对具体问题对 YOLO v2 进行的相应改进。如图 2 所示,在 YOLO v2 网络的基础上主要进行如下改进:

1) 对自制数据集的目标框进行维度聚类,确定 anchor 的参数。YOLO v2 的 anchor 是由 VOC 2007 和 VOC 2012 数据集聚类确定的,数据集中类别丰富,所确定的 anchor 参数具有普遍性,但不适合特定的检测任务,因此,需要在自制的航拍图像检测数据集中重新进行聚类操作。

2) 在分类网络训练的过程中使用分辨率不同的自制数据集对网络进行微调。与 YOLO v2 相同,首先使用 Imagenet 数据集进行预训练,而不同点是使用自制分辨率不同的图像分类数据集,可以取得更好的微调效果。

3) 训练过程中,每隔 10 轮改变模型的输入尺寸,使模型对不同尺度的图像具有稳健性。输入数据为自制航拍图像检测数据集。

4) 修改候选框的筛选规则,将非极大值抑制(NMS)操作更改为最大值操作。YOLO v2 中候选框的筛选规则是 NMS 操作,但在本文图像定位问题上可以直接进行最大值操作,以改善检测效果。

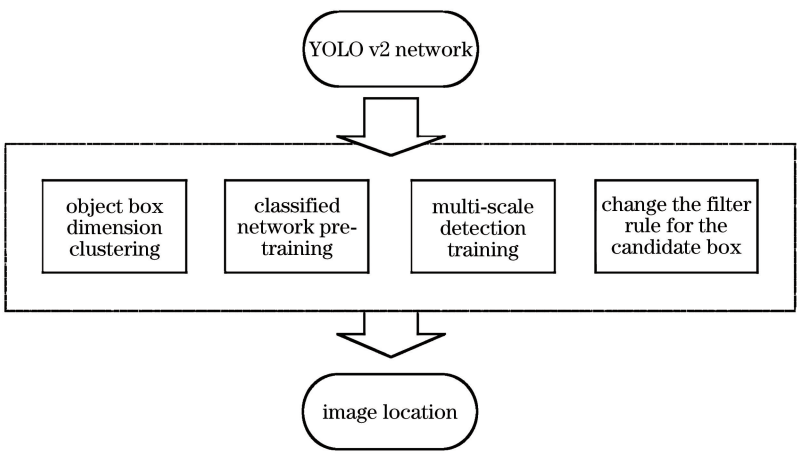

图 2 改进方法示意图

Fig. 2 Schematic of the improved method

4.1 目标框维度聚类

YOLO v2 借鉴了 Faster-RCNN 的思想,引入 anchor。anchor 是一组固定尺寸和宽高比的初始候选框,anchor 设计的好坏影响着目标检测的速度和目标框位置的精度。但因 Faster-RCNN 中 anchor 的个数和宽高维度是人工设定的,所以 Redmon 等[5]提出了维度聚类的方法,通过 k-means 对数据集中手工标记的目标框进行聚类,找到目标框的统计规律,以聚类个数 k 作为 anchor 个数,以 k 个聚类中心 box 的宽高维度作为 anchor 的维度。voc 和 coco 数据集的聚类结果为 5,因此确定 YOLO v2 网络中 anchor 的个数为 5。采用相同的方法,对自制数据集中目标区域对应的目标框进行聚类分析,得到适合检测数据集的最优 anchor 个数与宽高维度。

J 函数表示每个样本点到其聚类中心距离的平方和,k-means 的目的是要将 J 函数调整到最小。使用爬山法来选取 k 值,当聚类个数增加到一定值时,随着聚类个数的增加目标函数的变化很小,这个拐点可以认为是最优聚类个数。用 k-means 算法对数据集中目标框的宽高进行聚类,聚类过程中目标函数变化如图 3 所示,当 $k>4$ 时,目标函数变化很小,因此取 4 为最优聚类个数。当 $k=4$ 时,目标框的聚类分布如图 4 所示,4 种不同颜色区域对应着 4 种不同类别的目标框,则 anchor 的个数为 4;宽高维度分别为 4 个颜色区域

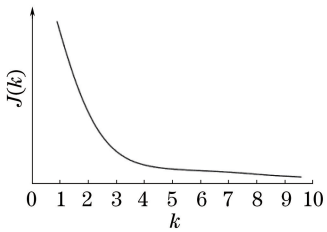

图 3 目标函数变化曲线

Fig. 3 Objective function change curve

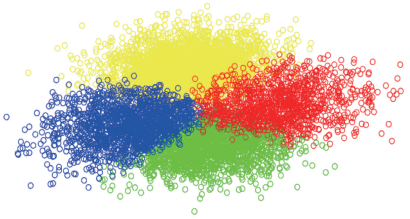

图 4 目标框聚类分布图

Fig. 4 Cluster of objective box

的聚类中心点对应的目标框宽高，在配置文件中更改 anchors 参数分别为(11.68,10.38)、(18.11,17.23)、(9.59,16.83)和(15.72,20.75)，分别对应绿、红、蓝、黄区域的聚类中心。

4.2 分类网络预训练

分类网络的预训练是目标检测中的重要环节，分类网络提取特征的能力和速度直接影响着目标检测的效果。主流的检测框架都会选择在 ImageNet[11] 上预训练的分类网络进行特征提取，但由于全连接层的限制，输入数据会被统一调整为固定大小。针对预训练阶段的不足，进行如下改进：

1) 采用 ImageNet 数据集对 Darknet-19 进行预训练。

2) 采用低分辨率(224 pixel×224 pixel)的航拍图像分类数据集，微调 Darknet-19，使网络适应航拍图像特征。

3) 修改 Darknet-19 分辨率为 448 pixel×448 pixel，在高分辨率的航拍图像分类数据集上训练 10 轮，让网络调整每一层的权重去适应高分辨率输入。

在预训练阶段进行微调并提高网络的分辨率，使得分类网络由分类算法切换为检测算法时，能够更好地适应航拍图像特征和多尺度检测的任务。

4.3 多尺度检测训练

由于 YOLO v2 只包含卷积层和池化层，因此可以随时改变输入图像的尺寸。通过借鉴多尺度输入的方法对检测网络进行训练，在训练过程中，每隔 10 轮改变模型的输入尺寸，从而模型对不同尺寸的图像具有稳健性。由于模型的下采样因子为 32，因此要求输入图像的尺寸均为 32 的倍数，尺寸的计算式为

$$S=32(7+a), \tag{3}$$

式中 S 为输入图像的尺寸，a 为在 0～12 中随机产生的自然数。

这种多尺度的训练规则强迫模型去适应不同的输入分辨率。相比于固定分辨率的模型，多尺度检测训练对于低分辨率的输入图像检测速度更快，对于高分辨率的输入图像检测准确率更高。

4.4 更改候选框的筛选规则

训练过程中，YOLO v2 的每一个候选框都会计算出自己的置信度，计算式为

$$\Pr(O_{\text{object}})*R_{\text{pred}}^{\text{truth}}, \tag{4}$$

式中 $\Pr(O_{\text{object}})$为网格中是否存在目标，$R_{\text{pred}}^{\text{truth}}$为预测候选框与真实目标框之间的重叠率。如果有人工标记的目标框落在候选框所在的网格里，则 $\Pr(O_{\text{object}})$取 1，否则取 0。

每个候选框的置信度与网格预测的类别信息相乘，会得到每个候选框对应的综合得分为

$$\Pr(C_i\,|\,O_{\text{object}})*\Pr(O_{\text{object}})*R_{\text{pred}}^{\text{truth}}=\Pr(C_i)*R_{\text{pred}}^{\text{truth}}, \tag{5}$$

式中 $\Pr(C_i\,|\,O_{\text{object}})$为每个网格预测的类别概率得分。

得到每个候选框的综合得分后，设置阈值，滤掉得分低的候选框，对保留的候选框进行非极大值抑制 NMS 处理，这时会得到若干组极大值的检测结果，但从实际问题的角度考虑，在航拍图像中一个目标区域最多出现一次。因此，将 NMS 操作更改为最大值操作，即在若干组大于阈值的综合得分中取最大值，唯一确定候选框的位置及其预测类别表示为

$$(i,p_{\text{pred}})=\max_{i,p_{\text{pred}}}[\Pr(C_i)*R_{\text{pred}}^{\text{truth}}],\Pr(C_i)*R_{\text{pred}}^{\text{truth}}\geqslant T_{\text{threshold}}, \tag{6}$$

式中 p_{pred}为目标框的预测位置，i 为对应的预测类别，$T_{\text{threshold}}$为筛选阈值。

通过最大值操作，在航拍图像中一类目标区域最多只会检测出一个目标框，可以有效地避免与目标区域

相似的区域对网络检测造成的误识别,从而提高目标检测准确率。

5 对比实验及仿真

5.1 实验数据

以吉林省长春市市区为中心的矩形区域作为研究对象。以谷歌地球上 2013 年 4 月、2015 年 10 月与 2016 年 11 月的吉林省长春市地区的卫星遥感图像作为图像源。图像瓦片级数为 15 级,空间分辨率为 9.55 m。本文的数据集分为两类:1) 预训练过程中需要使用的分类数据,按照分辨率高低分为两组,分别为 224 pixel×224 pixel 与 448 pixel×448 pixel;2) 检测网络训练需要的检测数据,按照分辨率的不同共分为 13 个小组。将研究区域分成 256 个大小相同的正方形区域,作为分类数据集的 256 类,在研究区域中标记了 64 个分布均匀的具有明显特征的区域,作为检测数据集的目标区域[14]。通过旋转、加入噪声、调整色调等方法来扩大样本数[15]。最终,得到的分类数据集样本总数为 53040,其中样本高、低分辨率的比例约为 3∶1,检测数据样本总数为 38200,不同分辨率的样本数量几乎相同,并且其中正射图像与倾斜图像的比例为 1∶1,以实现倾斜航拍图像的目标区域识别与定位。

5.2 实验配置与训练结果

实验配置如下:显卡为英伟达 GTX 1070,CPU 为 Intel Core i7-6700,主频为 3.40 GHz,内存为 32G,操作系统为 ubuntu 14.04,框架为 caffe。

网络参数如下:learning_rate 为 0.0001;policy 为 steps;batch 为 64;steps 分别取为 100,20000,35000;max_batches 为 50000;scales 为 10,0.1,0.1;momentu 为 0.9;decay 为 0.0005。

网络训练各参数收敛散点如图 4 所示,横坐标代表迭代次数,范围处于 0~2 万次,当网络迭代次数超过 2 万次时,各参数变化已基本稳定。散点图的纵坐标代表目标检测网络训练过程中重要的 4 个参数分别为:类别准确率、平均重叠率、召回率和损失值。从图 5 可以看出,随着迭代次数的增加,类别准确率和召回率逐渐接近于 1,平均重叠率稳定在 0.83,损失值下降到约 0.1。从各参数的收敛情况来看,网络的训练结果比较理想。

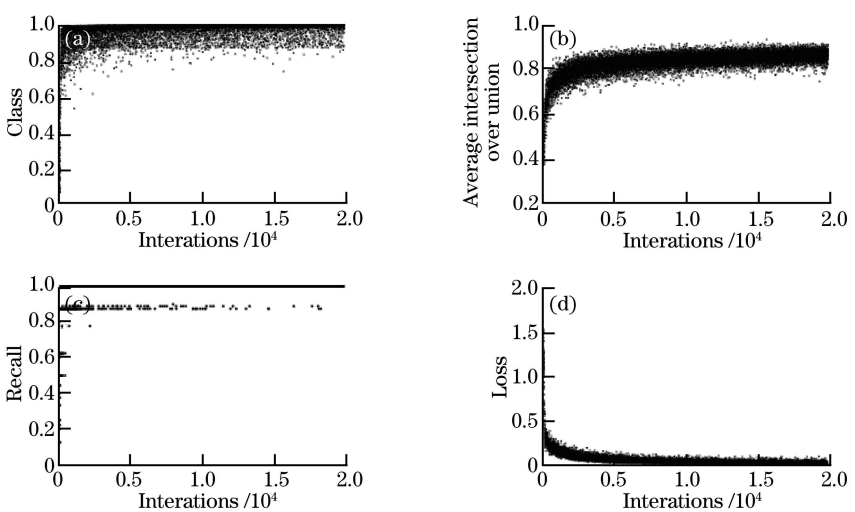

图 5 网络训练参数收敛散点图

Fig. 5 Network training parameters of the convergence scatter plot

5.3 性能对比

5.3.1 候选框生成方案对比

采用目标框维度聚类的方法,在 4.1 节中对自制数据集中目标区域对应的目标框进行聚类分析,得到适合检测数据集的最优候选框生成方案。本文方法与 Faster rcnn、YOLO v2 的候选框生成方案进行对比,结果如表 3 所示,维度聚类的方法可以在候选框数量较小,占用较少的计算资源的情况下,保证较高的平均重叠率。

表 3 候选框生成方案对比表

Table 3 Comparison of candidate box generation schemes

Candidate box generation scheme	Number of anchors	Average overlap rate
Faster rcnn	7	0.77
YOLOv2	5	0.79
Dimension clustering	4	0.83

5.3.2 分类网络预训练方法对比

分类网络预训练的方法共分为 3 个步骤，为了比较各步骤对于分类网络提取特征能力的影响，分别以 4.2节中 3 种不同阶段的预训练网络作为特征提取器，控制其他变量不变，采用相同的方法对检测网络进行训练。将 3 种网络的性能进行对比，来确定预训练方法的有效性。不同预训练方法的效果对比如图 6 所示，通过多分辨率微调后的分类网络，在检测任务中效果较好，MAP 值达到了 79.5。结果表明，多分辨率微调的预训练方法可以大幅提高检测过程中分类网络的特征提取能力。

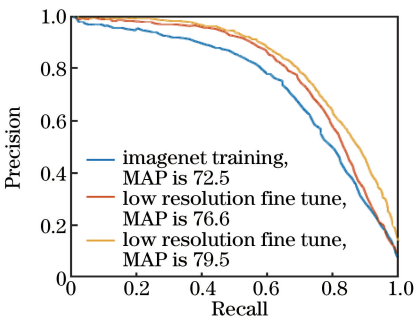

图 6 不同预训练方法效果对比图

Fig. 6 Comparison of different pre-training methods

5.3.3 多尺度网络与单一尺度网络的性能对比

经过多尺度训练的网络可以表现出对不同尺度的检测数据集很强的适应性，通过对不同尺度的检测数据集进行检测，得到多尺度网络与单一尺度网络的性能对比如表 4 所示，其中单一尺度网络的输入大小为 416 pixel×416 pixel。由表 4 可知，随着检测数据集尺度的增大，两种网络的检测效果均有所提高。多尺度网络和单一尺度网络相比，检测数据集尺度越小，检测速度越快，检测数据集尺度越大，MAP 值越高。当检测数据集尺度为 608 pixel×608 pixel 时，多尺度网络检测的 MAP 值已经超过了 80%，改进效果明显。

表 4 多尺度与单一尺度网络性能对比表

Table 4 Comparison of multi-scale and single-scale network performance

Detection of	Multi-scale network		Single-scale network	
dataset /(pixel×pixel)	Detection time /s	MAP	Detection time /s	MAP
224×224	0.01	71.1	0.013	70.3
320×320	0.012	74.8	0.014	74.2
416×416	0.015	77.5	0.015	77.8
512×512	0.018	79.4	0.016	78.4
608×608	0.029	80.9	0.018	78.9

5.3.4 更改候选框筛选规则

使用最大值操作 MAX 来代替 NMS。采用不同筛选规则 MAX 和 NMS 训练的网络检测效果对比如图 7 所示，与 NMS 操作相比，使用 MAX 规则筛选的网络 MAP 值提升了 5，因此航拍图像中一类目标区域最多只会检测出一个目标框，尽可能地避免了与目标区域相似的区域对网络检测造成的误识别，从而提高了目标检测准确率。

5.3.5 确定最佳阈值及检测验证集

在检测过程中，通过设置阈值来滤掉综合得分低的候选框，得到不同的检测效果。通过改进的方法进行训练的检测网络中设置不同阈值进行对比，如表 5 所示，其中 Rps/Img 表示为平均每张样本上目标框的个

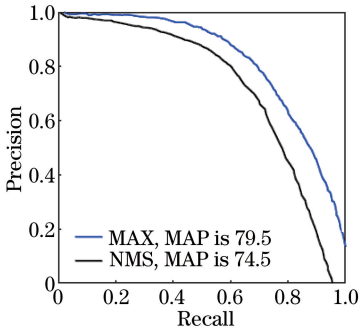

图 7 不同筛选规则效果对比图

Fig. 7 Comparison of different screeningrule effects

数。阈值偏高时精确度升高,但召回率降低;而阈值偏低时召回率升高,但精确率降低。当阈值约为 0.025 时,可以得到兼顾召回率和精确度的检测效果,其召回率为 69.91%,精确度为 80.45%。

表 5 不同阈值的检测效果对比表

Table 5 Comparison of the detectioneffects of different thresholds

Threshold	Rps /Img	Recall /%	Precision /%
0.001	7.75	100	12.07
0.005	3.13	86.78	50.27
0.025	1.56	69.91	80.45
0.45	1.17	50.29	93.73
0.8	0.12	12.03	100

因此,当使用改进的方法完成网络训练后,将阈值设置为 0.025,用验证集的样本对训练效果进行验证。部分正射结果如图 8 所示,在正射的航拍图像中可以精确地定位目标区域的位置,部分倾斜图像检测图如图 9 所示,且与图 8 中的结果一一对应。对于同一目标区域的航拍图像,因航拍角度不同会产生较大变形,而网络可以很好地将变形较大的目标检测出来。部分多目标图像检测结果如图 10 所示,当一张航拍图像中存在 2 个或 2 个以上的目标区域时,能够同时检测到多个目标区域[16]。根据第 3 节的坐标转换和投影关系,可以通过目标框标记航拍图像中的位置确定飞机的经纬度,当一张航拍图像中含有多个目标框时,需要综合判断;当多个目标框所确定的飞机经纬度同时在误差范围内时,输出定位结果,反之去掉置信度较低的目标框,继续判断,这样可以在多目标检测时,大幅提高飞机的定位准确度。

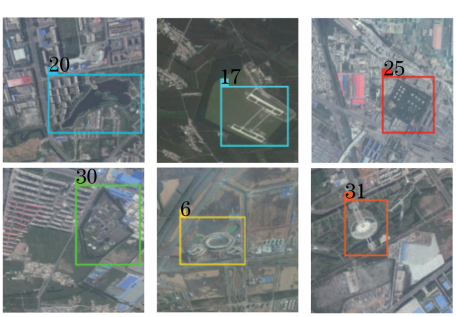

图 8 正射图像的检测结果

Fig. 8 Detection results of orthographic image

图 9 倾斜图像的检测结果

Fig. 9 Detection results of sloping image

5.4 仿真验证

在高分辨率地景的 Skyline 环境下进行验证实验。模拟无人机进行侦察飞行,设置无人机的相关参数,获取连续的航拍图像,将其作为待检测数据输入按改进方法训练好的 YOLO v2 网络,以此来判断航拍图像是否含有目标区域及其位置和类别,从而通过投影和坐标转换关系,近乎实时地计算出无人机的经纬度,然后与无人机在 Skyline 环境中的实际位置进行比较,以验证图像定位的准确性。

在研究区域上空选取 4 条不同的航线作为无人机在长春范围内的飞行轨迹,分别对这 4 条航线上所获

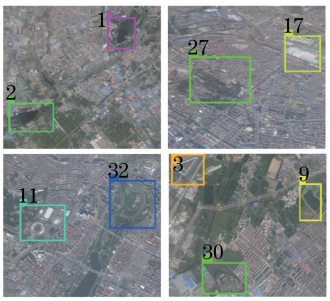

图 10 多目标图像检测结果

Fig. 10 Detection results of multi-target image

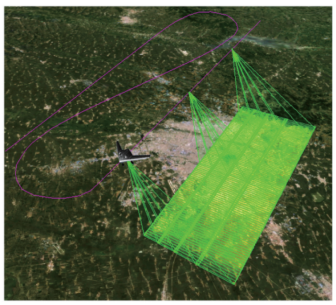

图 11 1 号航线示意图

Fig. 11 Diagram of No.1 route

取的航拍图像序列进行识别,图 11 为 1 号航线示意图。相机以 4 种不同的倾斜角度作为一个周期进行连续航拍,以对应的 4 张航拍图像作为相机拍照的一组序列。不同航线中图像定位的效果对比如表 6 所示,在研究范围内,图像序列中包含目标区域的比例平均在 40%以上;在包含目标区域的航拍图像中,其定位准确率达到 84%以上;当航拍图像中含有多个目标区域时,其定位准确率可以提高 6%,能够较好的完成图像定位任务。

表 6 不同航线的图像定位效果对比

Table 6 Comparison of image positioning effects of different routes

Track number	Number of sequences	Proportions of object area /%	Number of correct object area	Number of wrong object area	Accuracy rate of location/%
1	276	45	136	25	84.5
2	355	38	144	31	82.3
3	210	54	130	17	88.4
4	395	46	202	34	85.6

6 结 论

以 YOLO v2 网络为基础,通过目标框维度聚类、分类网络预训练、多尺度检测训练及更改候选框的筛选规则等方法以改进检测效果,在各组实验中与改进前的方法进行对比并且得到了验证,网络在航拍图像的目标区域检测任务中 MAP 提高到 79.5%;且在 Skyline 环境下进行的模拟飞行仿真实验,图像序列中包含目标区域的比例平均约为 40%;在包含目标区域的航拍图像中,其定位准确率达到 84%以上,验证了将图像定位问题转化为目标检测这一思路的可行性。但是还存在研究范围偏小,目标区域偏少的问题和数据样本的制作工作量大的不足,下一步将拓展研究范围,并用有效的方法,尽可能的简化数据制作的相关工作,继续进行以目标检测网络为基础的无人机图像定位研究。

参 考 文 献

[1] Le Cun Y, Bengio Y, Hinton G. Deep learning[J]. Nature, 2015, 521(7553):436-444.

[2] 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, 2014: 580-587.

[3] Girshick R. FastR-CNN[C] Proceedings of the IEEE International Conference on Computer Vision, 2015: 1440-1448.

[4] Ren S Q, He K M, Girshick R, *et al*. Faster R-CNN: towards real-time object detection with region proposal networks [C]. Advances in Neural Information Processing Systems, 2015: 91-99.

[5] 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, 2016: 779-788.

[6] Liu W, Anguelov D, Erhan D, *et al*. SSD: single shot multibox detector [C]. European Conference on Computer Vision, 2016: 21-37.

[7] Redmon J, Farhadi A. YOLO9000: Better, Faster, Stronger[J]. arXiv preprint arXiv: 1612. 08242, 2016: 1-9.

[8] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [J]. International

Conference on Learning Representations, 2015: 1-14.

[9] Szegedy C, Liu W, Jia Y, *et al*. Going deeper with convolutions[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 1-9.

[10] Lin M, Chen Q, Yan S C. Network in network[J]. arXiv preprint arXiv: 1312. 4400, 2014: 1-10.

[11] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]. Advances in Neural Information Processing Systems, 2012: 1097-1105.

[12] Zhang Sujie, Zhao Huaici. Algorithm research of optimal cluster number and initial cluster center[J]. Application Research of Computers, 2017, 34(6): 1617-1620.
张素洁, 赵怀慈. 最优聚类个数和初始聚类中心点选取算法研究[J]. 计算机应用研究, 2017, 34(6): 1617-1620.

[13] Li Lichun. Terrain reconstruction based on unmanned aerial vehicle sequence imaging and its application in navigation [D]. Changsha: National University of Defense Technology, 2009.
李立春. 基于无人机序列成像的地形重建及其在导航中的应用研究[D]. 长沙: 国防科学技术大学, 2009.

[14] Lü Xinran, Chen Jie, Zhang Libao, *et al*. Region of interest detection based on salient features clustering for remote sensing images[J]. Acta Optica Sinica, 2015, 35(s1): s110001.
吕欣然, 陈洁, 张立保,等. 基于显著特征聚类的遥感图像感兴趣区域检测[J]. 光学学报, 2015, 35(s1): s110001.

[15] Liu Dawei, Han Ling, Han Xiaoyong. High spatial resolution remote sensing image classification based on depth learning[J]. Acta Optica Sinica, 2016, 36(4): 0428001.
刘大伟, 韩玲, 韩晓勇. 基于深度学习的高分辨率遥感影像分类研究[J]. 光学学报, 2016, 36(4): 0428001.

[16] Shou Chengxun, He Yuntao, Sun Qingke. Point cloud registration based on convolution neural network[J]. Laser & Optoelectronics Progress, 2017, 54(3): 031001.
舒程珣, 何云涛, 孙庆科. 基于卷积神经网络的点云配准方法[J]. 激光与光电子学进展, 2017, 54(3): 031001.