

用于遥感图像的轻量化实时目标检测模型

李宇环^{1*}, 王洁¹, 鲁力¹, 聂莹²

¹空军工程大学防空反导学院, 陕西 西安 710051;

²陕西三原 93861 部队, 陕西 咸阳 713800

摘要 无人机搜索和识别目标依赖于目标检测算法的快速性和准确性。针对经典目标检测算法的网络结构复杂、计算机性能要求高和目标检测速度慢等问题,提出一种基于改进轻量级检测模型(Tiny YOLO-V3)的实时检测方法。首先提出一种新的网络结构作为主干网络,将最大通道数压缩至 128,进一步减小模型的时间复杂度和空间复杂度;其次使用单检测头并结合上下文信息增强对不同尺寸目标的检测能力,在保持检测精度的同时能够提高检测速度。最后采用武汉大学的遥感数据集进行实验。实验结果表明,改进后的模型在检测速度上有明显提升,同时精度提高 0.22。

关键词 机器视觉; 目标检测; 无人机; 卷积神经网络; 实时检测; 轻量化

中图分类号 TP391.4

文献标志码 A

doi: 10.3788/LOP202158.1615007

Lightweight Real-Time Target Detection Model for Remote Sensing Images

Li Yuhuan^{1*}, Wang Jie¹, Lu Li¹, Nie Ying²

¹Air Defense and Missile Academy, Air Force Engineering University, Xi'an, Shaanxi 710051, China;

²Unit 93861 of Sanyuan, Shaanxi, Xianyang, Shaanxi 713800, China

Abstract Search and recognition of targets by using unmanned aerial vehicle depend on the speed and accuracy of target detection algorithms. Aiming at the complex network structure of classic target detection algorithms, high computer performance requirements and slow target detection speed, a real-time detection method based on an improved lightweight detection model (Tiny YOLO-V3) is proposed. First, a new network structure is proposed as the backbone network, compressing the maximum number of channels to 128, further reducing the time complexity and space complexity of the model. Secondly, the single detection head combined with context information is used to enhance the detection ability of targets of different sizes, and the detection speed can be improved while the detection accuracy is maintained. Finally, the remote sensing dataset of Wuhan University is used to carry out the experiment. The experimental results show that the improved model has a significant increase in detection speed, while the accuracy has increased by 0.22.

Key words machine vision; target detection; unmanned aerial vehicle; convolutional neural network; real-time detection; lightweight

OCIS codes 150.1135; 100.4996; 110.2970

1 引言

无人机因具有无人员伤亡风险、机动灵活和效

费比高等特点,在现代战争中得到了广泛的应用。目标检测是无人机遂行打击和评估等任务的重要前提之一,通过计算机视觉手段可以为无人机提供视

收稿日期: 2020-10-20; 修回日期: 2020-11-04; 录用日期: 2020-12-22

通信作者: *xiao_wandering@163.com

觉定位,能够有效完成地形勘测、精确打击和战场环境侦察等军事任务,为此在军事领域具有重要的研究意义和应用价值。

传统的目标检测方法主要包含选择区域、提取特征和分类特征的过程,方法简单易行,但时间复杂度高,计算量大,鲁棒性差。随着计算机硬件性能的飞速发展,基于深度学习的目标检测算法受到了科研人员的广泛关注,深度学习方法中的卷积神经网络(CNN)是目前主流的目标检测网络之一。R-CNN(Region-CNN)^[1]将 CNN 引入目标检测领域并使用选择搜索生成候选区域,相比于传统目标检测的滑动窗口方法大大减少了计算量,从而使目标检测进入新的发展方向。Faster-RCNN^[2]中的核心部分为 RPN(Region Proposal Network),使用 CNN 生成候选区域可以极大地减少计算量,而且在精度和速度上都得到很大提升,从而实现了真正意义上的端到端目标检测。但上述两阶段的目标检测算法在实现高精度的同时,也带来了速度方面的不足。以 YOLO(You Only Look Once)系列^[3-5]为代表的单阶段目标检测算法,将图像作为 CNN 的输入可以直接回归边界框的位置及其所属类别,使目标实时检测成为了可能。从 LeNet^[6]发展到 VGGNet (Visual Geometry Group Network)^[7], CNN 能够提取越来越复杂的特征,ResNet^[8]能够有效减少因网络层数加深引起梯度消失和梯度爆炸的现象。使 CNN 的网络层数发展到数百层并引入更多的非线性变换,可以大幅度提升模型的表达能力。YOLO-V4^[5]结合了最新研究中大量的目标检测技巧,实现了精度和速度的最优平衡,可以灵活侧重精度或者速度方面。受限于无人机平台的算力,若要实现无人机平台独立对采集到的数据进行目标检测,必须对轻量化的目标检测算法进行研究。

Iandola 等^[9]提出了提高目标检测速度的两个手段,即减少模型中可学习参数的数量和减少整个网络的计算量,开拓了压缩目标检测模型的新方向。Inception-V1^[10]中借鉴了 NIN(Network in Network)^[11]的思想,大量使用大小为 1×1 的卷积核进行降维,在减少参数的同时增强了网络的表达能力。MobileNet-V1 使用深度可分离卷积来减少模型参数,将分组卷积的思想应用到了极致,但是结构太简单而导致精度大幅度下降。Wang 等^[12]认为 DenseNet^[13]需要进行大量计算推理的原因,即使用了大量重复的梯度信息来更新各层的权重,但这会造成计算量的冗余,为此提出了一种高效计算组件

CSPNet (Cross Stage Partial Network)^[12]。ThunderNet^[14]中增加了 CEM (Context Enhancement Module),在减少计算量的同时有效扩大了感受野。上述算法的研究为轻量化目标检测开拓了方向,使目标检测在无人机平台上实现成为了可能。黄梓桐等^[15]在网络中引入轻量级感受野模块和轻量级上下文模块,将最大通道数压缩到 128,提高了小目标检测的准确性和实时性。吕攀飞等^[16]将 Inception-V4 模块^[17]与 DarkNet-53^[4]结合,提高了网络的提取特征能力,但其只是将无人机采集的数据传回地面进行运算,并未将目标检测模型在无人机平台上独立运行。李成跃等^[18]提出了一种基于 YOLO-V3 的深度缩减网络,检测速度为 YOLO-V3 的 5 倍。黄涛等^[19]将残差块引入 DarkNet19 中,检测精度和速度较 YOLO-V2 均有所提升。孙佳等^[20]采用了跳帧检测的方法来弥补检测速度的不足。综上可知,上述文献对网络的轻量化程度还未满足实用需求。

将无人机采集到的画面传回地面进行目标检测会增加信息传输所带来的延迟,容易导致目标丢失等。当无人机数量较多时,这对地面设备的性能要求更高。针对上述问题,为实现目标检测模型在嵌入式计算机上运行,本文对 Tiny YOLO-V3 的主干网络进行改进,可以减少参数的数量和卷积层的运算次数。将两个检测头合并为一个,可以缩短检测时间,大幅度减少模型的时间复杂度(计算量)和空间复杂度(访存量),能够使模型在无人机平台上运行,并且在保持精度的基础上提升检测速度。

2 Tiny YOLO-V3

与全连接神经网络相比,CNN 使用稀疏连接的方式具有权值共享的特点,在大大减少计算量的同时能够很好地提取图像特征。使用 CNN 对输入图像进行卷积、池化和非线性激活等运算,可以预测出图像中每一个目标的位置和类别。对所有输出框进行非极大值抑制(NMS)等处理,可以筛选出较精确的目标检测框。YOLO-V3^[4]采用单阶段回归和多尺度融合的方式可以直接预测目标的位置和类别,具有检测速度快和识别精度高的优点。Tiny YOLO-V3 对其主干网络进行轻量化处理,网络结构如图 1 所示,其中 Y1 和 Y2 为两个输出,即代表两个检测头。轻量化处理后,将 CNN 压缩到 13 个卷积层,池化层虽然会损失一部分特征,但计算量比卷积层少得多,Tiny YOLO-V3 采用池化层作为下

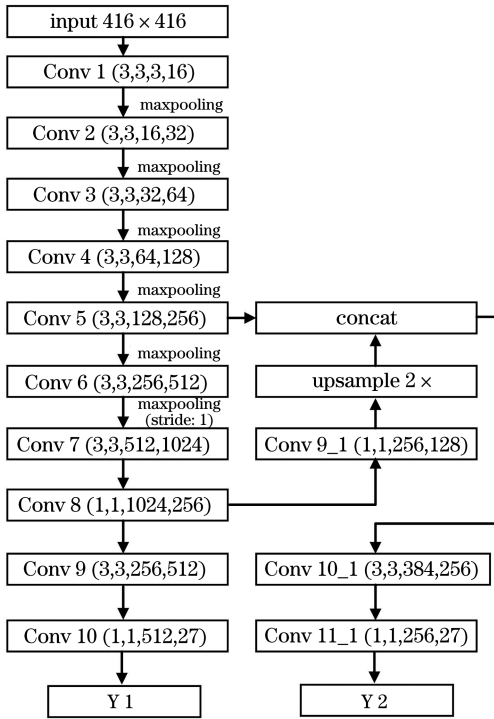


图 1 Tiny YOLO-V3 网络的结构

Fig. 1 Tiny YOLO-V3 network structure

采样层以降低图像的分辨率。检测头采用 FPN (Feature Pyramid Networks)^[21], 使用 32 倍下采样层来检测大尺寸目标, 将 32 倍下采样层进行上采样后与 16 倍下采样层进行融合并用检测小尺寸目标。

CNN 的计算量主要集中在卷积层, 一个卷积层的浮点运算个数为

$$T = D_F^2 \times D_K^2 \times C_{in} \times C_{out}, \quad (1)$$

式中: D_F 表示输出特征图的分辨率; D_K 表示卷积核的大小; C_{in} 表示输入通道数; C_{out} 表示输出通道数。一个卷积层中参数的内存占用量为

$$O = (D_K^2 \times C_{in} \times C_{out} + D_F^2 \times C_{out}) \times 4. \quad (2)$$

整个网络的时间复杂度(所有卷积层的浮点运算总数)为 2.77×10^9 , 空间复杂度(所有卷积层中参数的内存占用总量)为 5.92×10^7 。Tiny YOLO-V3 对 YOLO-V3^[4] 的主干网络和检测头分别进行压缩, 以此作为基础并在保持精度的前提下进行进一步压缩。

3 Tiny YOLO-V3 的改进模型

3.1 主干网络的改进

对于轻量级网络, 输入图像的尺寸太小会丢失图像特征, 输入图像的尺寸太大会增加第一个卷积层的运算量。为了在保留图像特征的同时提升检测速度, 选择 $320 \text{ pixel} \times 320 \text{ pixel}$ 作为输入图像的尺寸。

改进后的主干网络如图 2 所示, 其中 c 为每一层的通道数。

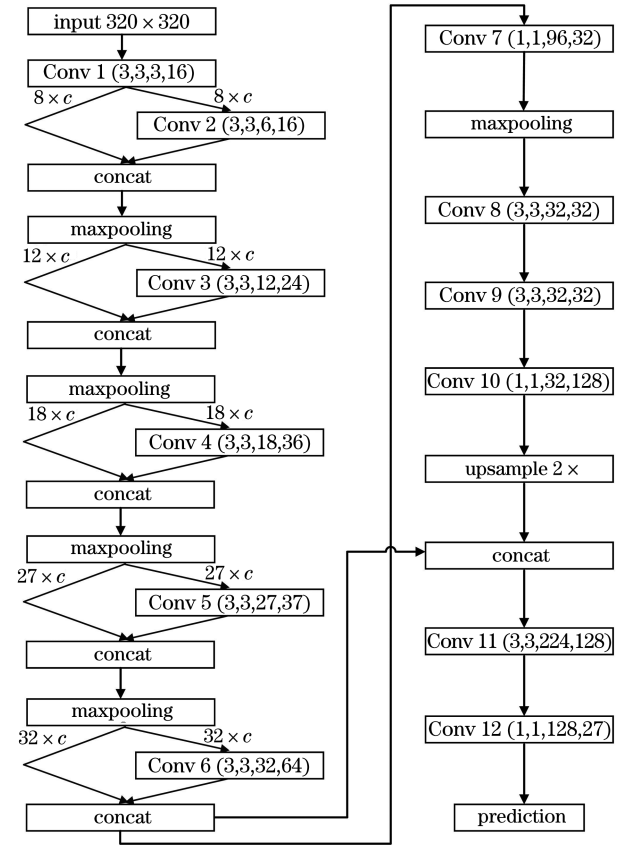


图 2 改进后的主干网络结构

Fig. 2 Improved backbone network structure

Tiny YOLO-V3 的主干网络采用大小为 3×3 的卷积核来提取图像特征, 使用平均池化层来降低图像的分辨率, 最大通道数为 256。由(1)式可知, 降低通道数可以有效减少计算量。将上一卷积层的输出通道分为 part 1 和 part 2 两个部分, 对 part 2 进行卷积后与 part 1 进行融合可以减少下一卷积层的输入通道数, 同时减少下一卷积层的输出通道数, 从而大幅度减少下一卷积层的运算量。此外, part 1 保留了上一层的特征, 使下一层在提取图像深层特征的同时保留了更多的空间信息, 能够有效提升小目标的检测效果。

通道数过大会大幅度增加卷积层的计算量。使用大小为 1×1 的卷积核在通道数大的卷积层之前进行降维, 不仅能够大幅度减少浮点运算次数, 而且能够充分整合各个通道之间的信息, 并且在降维的同时可以引入更多的非线性激活层, 从而增强网络的表达能力。改进后的主干网络在第 8 个卷积层之前进行降维, 在第 9 个卷积层之后进行升维, 减少了约为 2.7×10^7 次浮点运算, 使运算速度提升约为 11%。

提取图像深层特征的过程中,使用两个连续大小为 3×3 的卷积核能够增加网络的深度和表达能力,在增大感受野的同时可以减少浮点运算次数,能够提升网络的推理速度,而且使网络在缩减计算量的基础上能够保持对大尺度目标检测的精度和召回率。

3.2 检测头的改进

主干网络作为整个模型的骨架,可以充分提取图像的特征。检测头作为模型的头部,可以使用主干网络提取到的特征生成相应的输出并对图像的类别以及位置进行预测。深层特征图具有较大的感受野,包含更多的语义信息,能够较好地检测大尺寸目标,但是相比于浅层特征缺少了太多的空间位置信息,为此不容易检测出小尺寸目标。FPN^[21]中将深层特征进行上采样处理并与浅层信息融合,能够有效提高小目标的检测性能。YOLO-V3^[4]将 FPN 作为检测头并分为小、中和大三个尺寸,32 倍下采样的特征图作为大尺度采样框并用来检测图像中的大尺寸目标,小尺寸特征图经过上采样后与 16 倍下采样特征图融合作为中尺度采样框并用来检测图像中的中尺寸目标,中尺寸特征图经过上采样后与 8 倍下采样特征图融合作为小尺度采样框并用来检测图像中的小尺寸目标。采用 FPN 可以使网络有效检测图像中各个尺寸的目标,从而提升网络的精度。Tiny YOLO-V3 的检测头如图 3 所示,其中 c_5 和 c_8 为 Conv 5 和 Conv 8 的输出特征图。将 3 个尺度缩减为 2 个尺度,即去掉 8 倍下采样的检测头,在轻量化网络中仍能较好地检测大和小尺寸的目标。

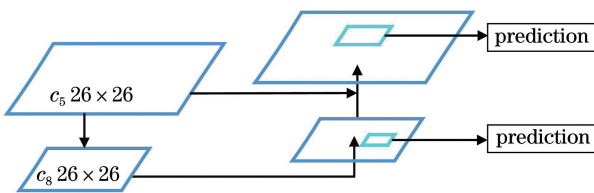


图 3 Tiny YOLO-V3 中检测头的结构

Fig. 3 Structure of detection head in Tiny YOLO-V3



图 5 数据集中的部分样本图像。(a)飞机;(b)油桶;(c)立交桥;(d)操场

Fig. 5 Partial sample images in dataset. (a) Aircraft; (b) oiltank; (c) flyover; (d) playground

使用多个检测头会增加卷积层的数量,同时会产生大量多余的检测框,从而增加了计算量,降低了检测速度。经过多次实验,将 Tiny YOLO-V3 中的两个检测头单独进行训练后并进行对比,结果发现去除 32 倍下采样的检测头经过训练后,精度几乎没有下降,比之前减少了约为 1.18×10^8 次浮点运算,但是去除 16 倍下采样的检测头后,精度会大幅度下降,因为 16 倍下采样的检测头中包含了更多的上下文信息,表达能力更强。综上,实验使用图 4 的检测头,将 32 倍下采样层经过上采样后与 16 倍下采样层进行融合,将其作为唯一的检测头并进行检测,可以减少大量的浮点运算,在精度上也不会有太大损失,其中 c_6 和 c_{10} 为 Conv 6 和 Conv 10 的输出特征图。

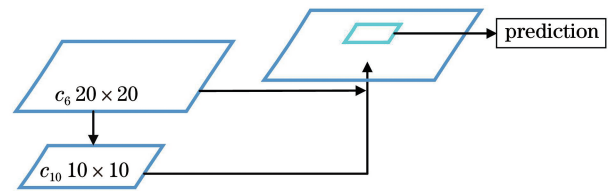


图 4 改进后的检测头结构

Fig. 4 Improved detection head structure

4 实验

4.1 数据集

实验过程中,使用武汉大学开源提供的 RSOD-Dataset^[22],数据集中共有 936 张图像,其中飞机有 446 张,油桶有 165 张,立交桥有 176 张,训练场有 149 张。部分图像如图 5 所示,图像中包含不同尺寸的样本,尤其是立交桥和训练场两个类别,背景复杂,样本较少,会增加学习难度。因数据集较小,故实验采用随机翻转、随机裁剪和随机平移等方法对数据集进行数据增强。按照各类别的比例,随机选取数据集的 85% 作为训练集,10% 作为验证集,验证集和剩下的 5% 共同作为测试集,具体划分情况如表 1 所示。采用 K-means 算法对数据集中的框进行聚类,得到大小分别为 (8 pixel, 8 pixel)、

表 1 数据集的划分数量

Table 1 Number of partitions in dataset

Dataset	Aircraft	Oiltank	Flyover	Playground
Training set	390	136	141	130
Test set	56	29	35	19

(13 pixel, 13 pixel)、(18 pixel, 19 pixel)、(25 pixel, 27 pixel)、(37 pixel, 42 pixel) 和 (112 pixel, 138 pixel) 共 6 种锚框。

4.2 实验环境

实验分为训练和测试两个阶段, 实验流程如图 6 所示。训练是在 Intel(R) Core(TM) i7-9700 CPU 3.00 GHz 和内存为 16 GB 的 Windows10 系统中所搭建的 TensorFlow 框架下进行, 显卡为 GeForce GTX 1660Ti, 显卡内存为 6 GB。测试使用的是模拟的无人机平台, 在 Cortex-A72 (ARM v8) 1.5 GHz、内存为 8 GB 和 Linux raspberrypi 4.19.118-v71+ 系统中所搭建的 TensorFlow 框架下进行, 具体配置信息如表 2 所示。

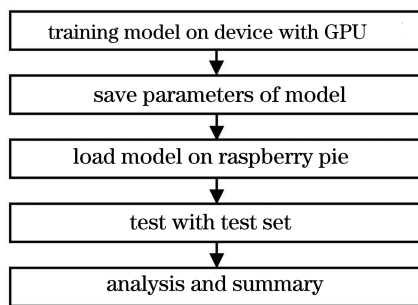


图 6 实验流程

Fig. 6 Experimental process

表 2 实验环境的配置

Table 2 Configuration of experimental environment

Facilitie	Version
Operating system of training	Windows10.0
Processor of training	Intel(R)Core(TM) i7-9700
	CPU 3.00 GHz
Operating system of test	Linux raspberrypi
	4.19.118-v71+
Processor of test	Cortex-A72(ARM v8)
PyThon	3.6.5
TensorFlow	1.14.0

4.3 实验结果

模型训练的过程中, 使用显卡 GeForce GTX 1660Ti 对其进行训练, 训练过程中未使用预训练的

权重, 为此需从头开始训练。为了避免训练初始出现梯度爆炸的现象, 将预热阶段的迭代次数设为 6, 预热阶段的学习率呈线性增长, 最大值为 1×10^{-4} 。随后在训练过程中, 采用余弦退火下降的方式来降低学习率。训练过程中学习率的变化曲线如图 7 所示。

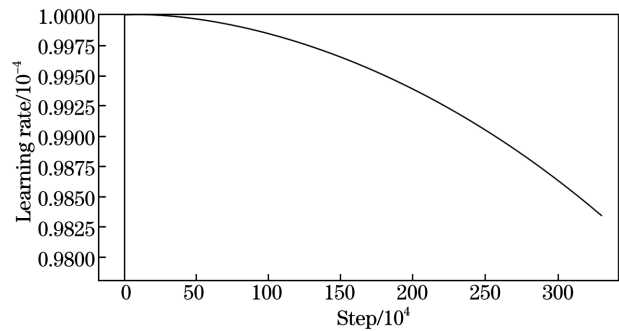


图 7 训练过程中的学习率变化曲线

Fig. 7 Change curve of learning rate in training process

训练过程中, 损失值的变化曲线如图 8 所示。从图 8 可以看到, 批处理量设为 6, 每轮中包括 133 个批, 迭代 5768 次后, 损失值收敛到 4.22, 此时终止训练。

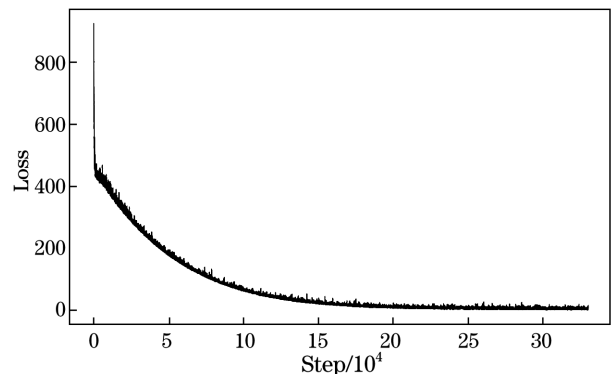


图 8 训练过程中损失值的变化曲线

Fig. 8 Change curve of loss value during training

训练结束后, 在树莓派微型电脑上加载训练后的模型并进行测试。目标检测模型的性能指标主要有精度、召回率、mAP (mean Average Precision) 和 FPS (Frames Per Second)。精度和召回率的定义公式为

$$P_A = \frac{x_{TP,A}}{x_{TP,A} + x_{FP,A}}, \quad (3)$$

$$R_A = \frac{x_{TP,A}}{x_{TP,A} + x_{FN,A}}, \quad (4)$$

式中: $x_{TP,A}$ 表示 A 类中被正确识别的目标; $x_{FP,A}$ 表示 A 类中被错误识别为正例的目标; $x_{FN,A}$ 表示 A 类中未正确识别的正例; P_A 表示对于所有识别

为类别 A 的目标中正确识别的目标占比; R_A 表示图片中所有类别为 A 的目标中被正确识别的目标占比。置信度是某个目标属于类别 A 的概率,将测试集中每个类别被检测出的目标按照置信度从高到低进行排序并绘制该类别的精度-召回率曲线,将曲线下的面积作为此类别的 AP,接着对所有类别的 AP 求均值即可得到 mAP,mAP 反映模型对各个类

别的平均检测精度。FPS 指目标检测模型每秒处理图像的帧数,反映目标的检测速度。对模型的性能指标进行测试,主要记录 mAP 和 FPS 两项指标,用来对比 Tiny YOLO-V3 和改进后模型的性能。各个类别的精度如表 3 所示,部分检测结果如图 9 所示。从表 3 可以看到,改进后模型的检测精度较之前模型提升了 0.022。

表 3 Tiny YOLO-V3 与改进后模型的 mAP 值

Table 3 mAP values of Tiny YOLO-V3 and improved model

Model	AP				mAP
	Aircraft	Oiltank	Flyover	Playground	
Tiny YOLO-V3	0.608	0.397	0.007	0.023	0.253
Improved model	0.635	0.431	0.012	0.022	0.275

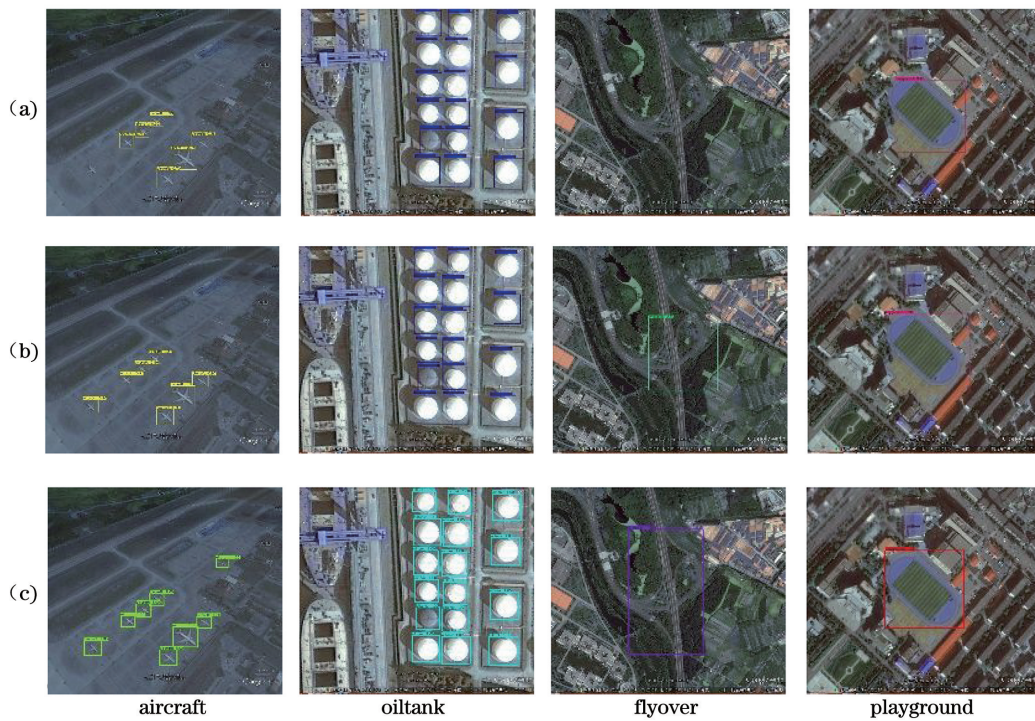


图 9 不同模型的部分检测结果。(a) Tiny YOLO-V3; (b) 改进后模型检测结果; (c) 数据集标签

Fig. 9 Partial test results of different models. (a) Tiny YOLO - V3; (b) test results of improved model; (c) dataset labels

改进后网络的 AP 值较之前网络均有提升,上升了约 2%。从图 9 可以看到,飞机样本最多且特征相对简单,所以精度最高;油桶的颜色和大小不一,这会增加学习难度;立交桥和训练场两个类别的数据集较小且每张图像中包含的样本较少,背景复杂,为此学习难度更大,所以这两个类别的 AP 值过低,最终模型的 mAP 值较小,对此可对这两个类别进行额外的数据增强来改善结果。实验结果一方面说明改进前网络的卷积层中的部分通道是冗余的;另一方面说明改进后的模型中包含了更多的空间信

息,能够有效提升模型的表达能力。

目标检测模型的时间复杂度表示处理一张图像所作的浮点运算次数,空间复杂度表示模型参数所需的内存占用大小。实验过程中发现,尽管模型的时间复杂度较之前减少了 87.7%,空间复杂度减少了 67.9%,检测目标的时间缩短了,但读取图像和非极大值抑制的时间并不会因为模型的变化而缩短,而且图像的分辨率越大,读取图像的速度越慢,在实验前需将图像的分辨率调整为 320 pixel \times 320 pixel 并进行测试。此外需要说明的是,模型的

性能好坏还要根据计算机的最大算力和带宽上限等因素考虑,不能一概而论。实验中模型的复杂度和帧率如表 4 所示。从表 4 可以看到,改进后的网络在树莓派微型电脑上将帧率从 1.83 提升到了 3.01。

表 4 不同模型的复杂度和帧率

Table 4 Complexity and frame rate of different models

Model	Time complexity	Space complexity	FPS
Tiny YOLO-V3	2767598080	59176252	1.83
Improved model	423320000	19023708	3.01

5 结 论

无人机上使用的计算平台算力较低,通常难以运行复杂的目标检测模型。为了在无人机平台上独立运行目标检测模型,对轻量级网络 Tiny YOLO-V3 进行改进并将主干网络和检测头重新设计,使其可以在保持检测精度的基础上进一步压缩时间和空间复杂度以提升检测速度。实验结果证明,改进后的模型在树莓派微型电脑上将帧率从 1.83 提升到了 3.01,同时检测精度较之前模型提升了 0.022,较好地实现预期目标。结果也表明高效的目标检测模型能够以较少的参数和运算量尽可能多地提取图像的特征,从而在保证精度的同时降低模型对设备性能的要求。另外,图像的读取和检测框的处理仍是帧率提升的瓶颈,后期需继续研究影响模型运行速度的其他因素。

参 考 文 献

- [1] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2014, Columbus, OH, USA. New York: IEEE Press, 2014: 580-587.
- [2] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
- [3] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 6517-6525.
- [4] Redmon J, Farhadi A. YOLOv3: an incremental improvement[EB/OL]. (2018-04-08)[2020-10-19]. <https://arxiv.org/abs/1804.02767>.
- [5] Bochkovskiy A, Wang C Y, Liao H Y. YOLOv4: optimal speed and accuracy of object detection[EB/OL]. (2020-04-23)[2020-10-19]. <https://arxiv.org/abs/2004.10934>.
- [6] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [7] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2015-04-10)[2020-10-19]. <https://arxiv.org/abs/1409.1556>.
- [8] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.
- [9] Iandola F N, Han S, Moskewicz M W, et al. SqueezeNet: AlexNet-level accuracy with 50× fewer parameters and <0.5 MB model size [EB/OL]. (2016-11-04)[2020-10-19]. <https://arxiv.org/abs/1602.07360>.
- [10] Szegedy C, Liu W, Jia Y Q, et al. Going deeper with convolutions[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE Press, 2015: 15523970.
- [11] Lin M, Chen Q, Yan S. Network in network[C]//International Conference on Learning Representations, April 14-16, 2014, Banff, AB, Canada. [S.l.: s.n.], 2014.
- [12] Wang C Y, Mark Liao H Y, Wu Y H, et al. CSPNet: a new backbone that can enhance learning capability of CNN[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), June 14-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 1571-1580.
- [13] Huang G, Liu Z, van der Maaten L, et al. Densely connected convolutional networks [C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 2261-2269.
- [14] Qin Z, Li Z M, Zhang Z N, et al. ThunderNet: towards real-time generic object detection on mobile devices[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea. New York:

- IEEE Press, 2019: 6717-6726.
- [15] Huang Z T, Arifu K. Real-time detection of pedestrian and vehicle targets under uav platform [J/OL]. Computer Engineering and Application. (2020-08-06) [2020-10-19]. <http://kns.cnki.net/kcms/detail/11.2127.TP.20200806.1353.018.html>.
黄梓桐, 阿里甫·库尔班. 无人机平台下的行人与车辆目标实时检测 [J/OL]. 计算机工程与应用. (2020-08-06) [2020-10-19]. <http://kns.cnki.net/kcms/detail/11.2127.TP.20200806.1353.018.html>.
- [16] Lü P F, Wang S G. Intelligent target recognition method of unmanned aerial vehicle combat platform [J]. Laser & Optoelectronics Progress, 2019, 56(7): 071001.
吕攀飞, 王曙光. 无人机作战平台的智能目标识别方法 [J]. 激光与光电子学进展, 2019, 56(7): 071001.
- [17] Szegedy C, Ioffe S, Vanhoucke V, et al. Inception-v4, inception-ResNet and the impact of residual connections on learning [EB/OL]. (2016-08-23) [2020-10-19]. <https://arxiv.org/abs/1602.07261>.
- [18] Li C Y, Yao J M, Lin Z X, et al. Object detection method based on improved YOLO lightweight network [J]. Laser & Optoelectronics Progress, 2020, 57(14): 141003.
李成跃, 姚剑敏, 林志贤, 等. 基于改进 YOLO 轻量化网络的目标检测方法 [J]. 激光与光电子学进展, 2020, 57(14): 141003.
- [19] Huang T, Zhao S F, Bai Y R, et al. Method of real-time road target depth neural network detection for UAV flight control platform [J]. Laser & Optoelectronics Progress, 2020, 57(4): 041509.
黄涛, 赵栓峰, 拜云端, 等. 面向无人机飞控平台的实时道路目标深度神经网络检测方法 [J]. 激光与光电子学进展, 2020, 57(4): 041509.
- [20] Sun J, Guo D B, Yang T T, et al. Real-time object detection based on improved YOLOv3 network [J]. Laser & Optoelectronics Progress, 2020, 57(22): 221505.
孙佳, 郭大波, 杨甜甜, 等. 基于改进的 YOLOv3 网络的实时目标检测 [J]. 激光与光电子学进展, 2020, 57(22): 221505.
- [21] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection [C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 936-944.
- [22] Xiao Z F, Liu Q, Tang G F, et al. Elliptic Fourier transformation-based histograms of oriented gradients for rotationally invariant object detection in remote-sensing images [J]. International Journal of Remote Sensing, 2015, 36(2): 618-644.