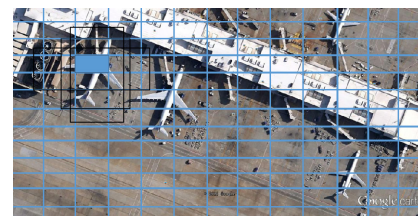




DOI: 10.12086/oe.2018.180350

## 遥感图像中飞机的改进 YOLOv3 实时检测算法

戴伟聪<sup>1,2\*</sup>, 金龙旭<sup>1</sup>, 李国宁<sup>1</sup>, 郑志强<sup>3</sup><sup>1</sup>中国科学院长春光学精密机械与物理研究所, 吉林 长春 130033;<sup>2</sup>中国科学院大学, 北京 100049;<sup>3</sup>长春理工大学, 吉林 长春 130022

**摘要:** 针对遥感图像中的飞机目标, 本文提出一种遥感图像飞机的改进 YOLOv3 实时检测算法。首先, 针对单一的遥感图像飞机目标, 提出一种有 49 个卷积层的卷积神经网络。其次, 在提出的卷积神经网络上应用密集相连模块进行改进, 并提出使用最大池化加强密集连接模块间的特征传递。最后, 针对遥感图像中飞机多为小目标的现实, 提出将 YOLOv3 的 3 个尺度检测增加至 4 个并以密集相连融合不同尺度模块特征层的信息。在本文设计的遥感飞机测试集上进行训练和测试, 实验表明, 该算法的检测精度达到 96.26%、召回率达到 93.81%。

**关键词:** 遥感图像; 飞机目标; 实时检测; 卷积神经网络

**中图分类号:** TP751; O436.3

**文献标志码:** A

**引用格式:** 戴伟聪, 金龙旭, 李国宁, 等. 遥感图像中飞机的改进 YOLOv3 实时检测算法[J]. 光电工程, 2018, 45(12): 180350

## Real-time airplane detection algorithm in remote-sensing images based on improved YOLOv3

Dai Weicong<sup>1,2\*</sup>, Jin Longxu<sup>1</sup>, Li Guoning<sup>1</sup>, Zheng Zhiqiang<sup>3</sup><sup>1</sup>Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, Jilin 130033, China;<sup>2</sup>University of Chinese Academy of Sciences, Beijing 100049, China;<sup>3</sup>Changchun University of Science and Technology, Changchun, Jilin 130022, China

**Abstract:** Focusing on the airplanes in remote-sensing images, a real-time algorithm based on improved YOLOv3 is proposed to detect airplanes in remote-sensing images. Firstly, a convolutional neural network that consists of 49 convolutional layers is proposed to detect airplanes in remote-sensing images specifically. Secondly, dense connection is employed on proposed convolutional neural network, and maxpool is employed to enhance the feature transmit between dense blocks. Finally, to deal with the fact that airplanes in remote-sensing images are small targets mainly, we propose to increase the scale detection from 3 to 4 and employ dense connection to merge feature map among different scales. The algorithm is trained and tested on the designed airplane dataset. The experiment results show that our algorithm obtain 96.26% on precision and 93.81% on recall.

收稿日期: 2018-06-28; 收到修改稿日期: 2018-08-22

基金项目: 国家 863 高技术研究发展计划项目(863-2-5-1-13B)

作者简介: 戴伟聪(1994-), 男, 硕士研究生, 主要从事计算机视觉、目标检测跟踪方面的研究。E-mail: daiweicong16@mailsucas.ac.cn

**Keywords:** remote-sensing image; airplane target; real-time detection; convolutional neural network

**Citation:** Dai W C, Jin L X, Li G N, et al. Real-time airplane detection algorithm in remote-sensing images based on improved YOLOv3[J]. *Opto-Electronic Engineering*, 2018, 45(12): 180350

## 1 引言

遥感图像中飞机目标的检测在军事应用中有着重要的需求,对遥感图像中的飞机的实时监测在现代化战争中可以带来极大的优势;但是遥感图像中存在背景杂斑、光照变化等因素严重影响分类器性能,使得遥感图像中的飞机目标在传统的机器学习框架中难以被准确高效地检测。近几年来,基于深度学习的目标检测方法得到了很大的突破,其主要的方法可以分为两类。一类是基于区域推荐的目标检测算法,首先使用区域推荐产生候选目标,随后使用卷积神经网络进行处理;其精度较高但不能满足实时应用的需求,具有代表性的算法有 RCNN<sup>[1]</sup>、Fast RCNN<sup>[2]</sup>、Faster RCNN<sup>[3]</sup>、Mask RCNN<sup>[4]</sup>等。另一类是基于回归方法的目标检测算法,将检测问题作为回归问题处理,直接预测目标位置与类别。这类方法速度快,但相对而言精度偏低,具有代表性的算法有 YOLO<sup>[5]</sup>、SSD<sup>[6]</sup>、YOLOv2<sup>[7]</sup>、YOLOv3<sup>[8]</sup>等。

由于使用卷积神经网络提取的特征的泛化性远高于传统的人工特征,基于卷积神经网络的目标检测方法被迅速推广应用到各个领域,如图像分类、农业、自动驾驶、零件检测、军事等。薛月菊等人<sup>[9]</sup>在 YOLOv2-tiny 的基础上,结合密集连接,提出了一种用于识别未成熟芒果的方法。王思雨等人<sup>[10]</sup>基于 LeNet5 改进提出了一种基于卷积神经网络的合成孔径图像中的飞机目标检测方法。周敏等人<sup>[11]</sup>设计了一个包含 8 种飞机的数据集和一个 5 层的卷积神经网络用于飞机分类,相对于传统的机器学习方法有很大的提高。谷雨等人<sup>[12]</sup>将卷积神经网络用于提取合成孔径图像中的特征,并结合超限学习机与集成学习提出了一种快速的合成孔径图像目标识别方法。

YOLO 在各个领域应用时相对 RCNN 具有更好的泛化性。为了实现实时高效的遥感飞机目标检测,本文选择基于 YOLOv3 算法,针对遥感图像中的飞机目标对 YOLOv3 算法进行优化改进,并设计搭建一个用于飞机目标检测的卷积神经网络。

本文的主要贡献有:

1) 设计并搭建了一个卷积神经网络,解决了 YOLOv3-tiny 召回率低而 YOLOv3 参数量过大的问题。

2) 使用密集相连模块增强卷积神经网络层间连接,并通过最大池化上一个密集模块间的输出进一步加强密集神经网络区域间的连接,有效解决深度神经网络的退化问题。

3) 对 YOLOv3 的多尺度检测进行改进,首先将 3 个尺度检测扩展为 4 个尺度检测以增大检测遥感图像中的飞机小目标的准确率;其次参照密集神经网络,对 4 个尺度的检测网络使用密集相连复用特征提高 YOLOv3 的多尺度检测的准确率。

4) 针对遥感图像中的飞机目标缺乏数据集的问题,自行设计并标注了一个遥感图像中的飞机目标数据集。

## 2 YOLOv3

YOLOv3 在 YOLOv2 的基础上进行了一些应用性的改进,包括多尺度检测、多标签分类等,并使用基于残差神经网络改进的 Darknet53 网络作为特征提取器,使 YOLO 系列方法不擅长检测小物体的缺陷得到了改善。YOLOv3 也因此成为目前为止最优秀的目标检测算法之一。

YOLOv3 首先将原始图片缩放到 416×416 的大小,使用了类似 FPN 网络的尺度金字塔结构,将原始图像按照特征图的尺度大小划分为  $S \times S$  个等的单元格,在特征图大小为 13×13、26×26、52×52 的三个尺度上进行检测,并使用 2 倍的上采样使特征图在相邻两尺度上传递。每一个单元格借助 3 个锚点框(anchor box)

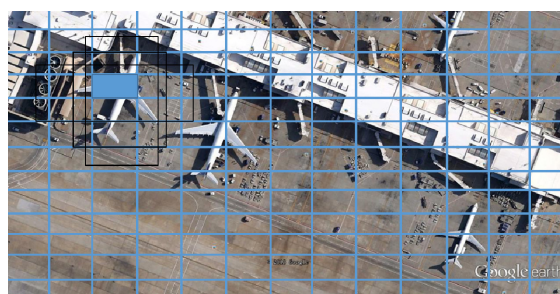


图 1 YOLOv3 在 13×13 的单元格中的预测边界框示意图

Fig. 1 An illustration of predicted bounding boxes on 13x13 grids of YOLOv3

预测 3 个边界框，如图 1 所示。

卷积神经网络在每一个单元格上会为每一个边界框预测 4 个值，即坐标 $(x,y)$ 与目标的宽  $w$  和高  $h$ ，分别记为  $t_x, t_y, t_w, t_h$ 。若目标中心在单元格中相对于图像左上角有偏移 $(c_x, c_y)$ ，并且锚点框具有高度和宽度  $p_w, p_h$ ，则修正后的边界框为

$$\begin{aligned} b_x &= \sigma(t_x) + c_x, \\ b_y &= \sigma(t_y) + c_y, \\ b_w &= p_w e^{t_w}, \\ b_h &= p_h e^{t_h}. \end{aligned} \quad (1)$$

在训练过程中，使用误差平方和作为损失。假设真实坐标为  $\hat{t}^*$ ，则梯度可以通过最小化损失函数求得，梯度为真实坐标值减去预测坐标值： $\hat{t}^* - \hat{t}$ 。

YOLOv3 使用逻辑回归预测锚点框中包含物体的概率。如果锚点框与真实目标边界框的重叠率大于任何其他锚点框，则这个锚点框的概率为 1；如果锚点框与真实目标的边界框的重叠率大于 0.5，但又不是最大的，则忽略这个预测。YOLOv3 在训练时仅对一个物体分配一个锚点框。如果锚点框中没有被认为包含物体，则使其在损失函数中不起作用。YOLOv3 在训练过程中使用二元交叉熵损失和逻辑回归来进行类别预测，这种方式使得 YOLOv3 可以对一个目标进行多标签分类。

### 3 基于 YOLOv3 的改进

本文针对遥感图像中的飞机目标检测，对 YOLOv3 在网络结构与多尺度检测上提出了三点改进。

#### 3.1 网络结构改进

YOLOv3 中 Redmon 等人借鉴 ResNet 提出了 Darknet53 网络。虽然 Darknet53 通过残差结构使网络的训练难度得到减少，并使用了大量的  $1 \times 1$  卷积核与使用  $3 \times 3$  大小步长为 2 的卷积核替代最大池化减少了参数数量，但对于单类物体的检测，Darknet53 网络有些过于复杂与冗余，过多的参数会导致训练复杂、增大对数据量的需求以及减慢检测速度。

为了实现遥感飞机的实时检测，本文借鉴 Darknet53，以保持准确度并减少参数量为出发点，提出一种参数数量较少、运算复杂度相对较低的卷积神经网络作为特征提取网络。

本文提出的卷积神经网络，称为 Darknet49，在该神经网络中，在过渡模块中使用了  $1 \times 1$  卷积核进一步

降低维数。由于低维度的卷积层使用非线性激活函数会一定程度上破坏图像信息，因此本文在第一个卷积层中使用线性激活函数，如表 1 所示。

表 1 Darknet49 的网络结构  
Table 1 The network structure of Darknet49

Type	Output	Filters	Size
Conv	208×208	16	3×3 conv stride=2
Residual block(1)	208×208	16	1×1 conv stride=1
×4		32	3×3 conv stride=1
Transition module	104×104	16	1×1 conv stride=1
		32	3×3 conv stride=2
Residual block(2)	104×104	32	1×1 conv stride=1
×4		64	3×3 conv stride=1
Transition module	52×52	32	1×1 conv stride=1
		64	3×3 conv stride=2
Residual block(3)	52×52	64	1×1 conv stride=1
×4		128	3×3 conv stride=1
Transition module	26×26	64	1×1 conv stride=1
		128	3×3 conv stride=2
Residual block(4)	26×26	128	1×1 conv stride=1
×4		256	3×3 conv stride=1
Transition module	13×13	128	1×1 conv stride=1
		256	3×3 conv stride=2
Residual block(5)	13×13	128	1×1 conv stride=1
×4		256	3×3 conv stride=1

#### 3.2 密集相连

Huang 等<sup>[13]</sup>借鉴了深度残差神经网络的思想，直接将每个密集模块中的所有层连接起来，使每一层的输入都包含了所有较早层的特征图。DenseNet 通过层间连接有效地解决了梯度消失问题，强化了特征的传递，更有效地复用了卷积神经网络的特征，大幅减少参数的数量，从而减少计算量。相对于小数据集而言，DenseNet 能很好地解决过拟合的问题，尤其适合于训练数据相对匮乏的应用。

本文使用密集相连增强所提出的 Darknet49 网络，Darknet49 的密集相连版本称为 Darknet49-Dense。Darknet49-Dense 由 5 个密集相连模块和 4 个过渡模块组成，每个密集相连模块间有一个过渡模块将特征图的尺寸降低。如图 2 所示，过渡模块中分别对上一个密集模块的输出做步长为 2 的最大池化后与使用步长为 2 的卷积核的输出进行串联拼接后作为下一个密集模块的输入。以这种方式使密集神经网络的模块间连

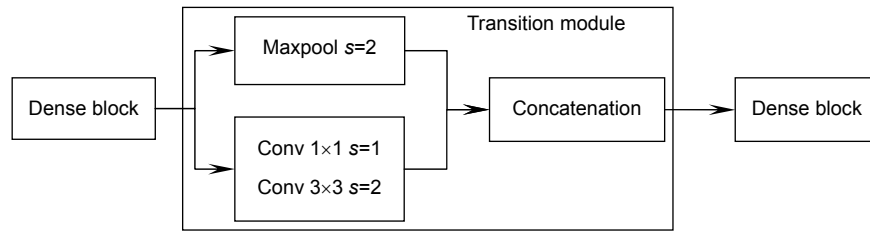


图 2 过渡模块的图解

Fig. 2 An illustration of transition module

接得到增强，减少跨模块间的特征传递损失和增强特征复用。

在密集模块中，第  $l$  层的输入为前  $l-1$  层的输出的串联拼接：

$$x_l = H_l([x_1, x_2, \dots, x_{l-1}]) \quad (2)$$

其中： $H_l$  为激活函数、卷积层、批量归一化的复合函数。

### 3.3 多尺度检测改进

YOLOv3 中，引入 FPN 网络<sup>[14]</sup>，同时利用低层特征的高分辨率和高层特征的高语义信息，并通过上采样融合不同层的特征，在三个不同尺度的特征层上检测物体。针对遥感图像中的飞机目标大多数为小目标的情况，本文对 YOLOv3 中的尺度检测模块进行改进，将原有的 3 个尺度检测扩展为 4 个尺度检测，在较大特征图给小目标分配更为准确的锚点框。

以矩形框的交并比 (IOU，用  $R_{IOU}$  表示) 作为相似度对遥感飞机目标训练集的所有目标标注使用 K-means 聚类获得锚点框的大小，K-means 聚类的距离函数为

$$d(B, C) = 1 - R_{IOU}(B, C) \quad (3)$$

其中： $B$  为矩形框的大小， $C$  为矩形框的中心， $R_{IOU}(B, C)$  表示两个矩形框的交叠比。

本文权衡平均交并比与锚点框的数量 (其关系如图 3)，取 12 个锚点框，分别为 (12, 16)，(16, 24)，(21, 32)，(24, 41)，(24, 51)，(33, 51)，(28, 62)，(39, 64)，(35, 74)，(44, 87)，(53, 105)，(64, 135)。在每个尺度上的每一个单元格借助 3 个锚点框预测 3 个边界框。

随后本文借鉴 Densenet 的思想，对四个尺度检测的特征层上采样相应倍数后使用密集相连。尺度检测层的密集相连进一步融合了不同层次的特征，增强各个尺度特征层的语义信息。

图 4 中给出了本文提出的多尺度检测模块，其中  $2\times$  表示步长为 2 的上采样， $4\times$  表示步长为 4 的上采样， $8\times$  表示步长为 8 的上采样。在  $416\times 416$  的大小的图片下，本文提出的基于 Darknet49-Dense 的 YOLOv3 运算量为 14.525 BFLOPS，基于 Darknet49 的 YOLOv3 运算量为 9.695 BFLOPS，而基于 Darknet53 的 YOLOv3 运算量为 65.86 BFLOPS。

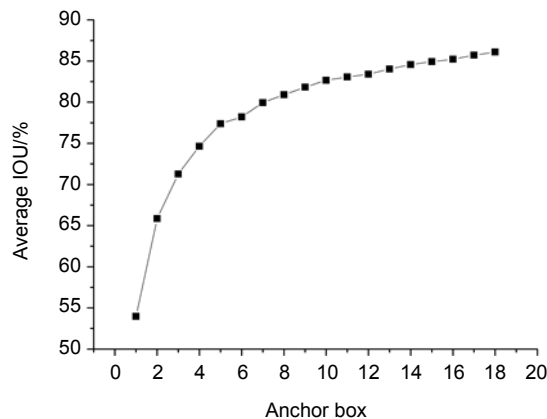


图 3 锚点框数量与平均交并比的关系

Fig. 3 The relationship between the number of anchor boxes and average IOU

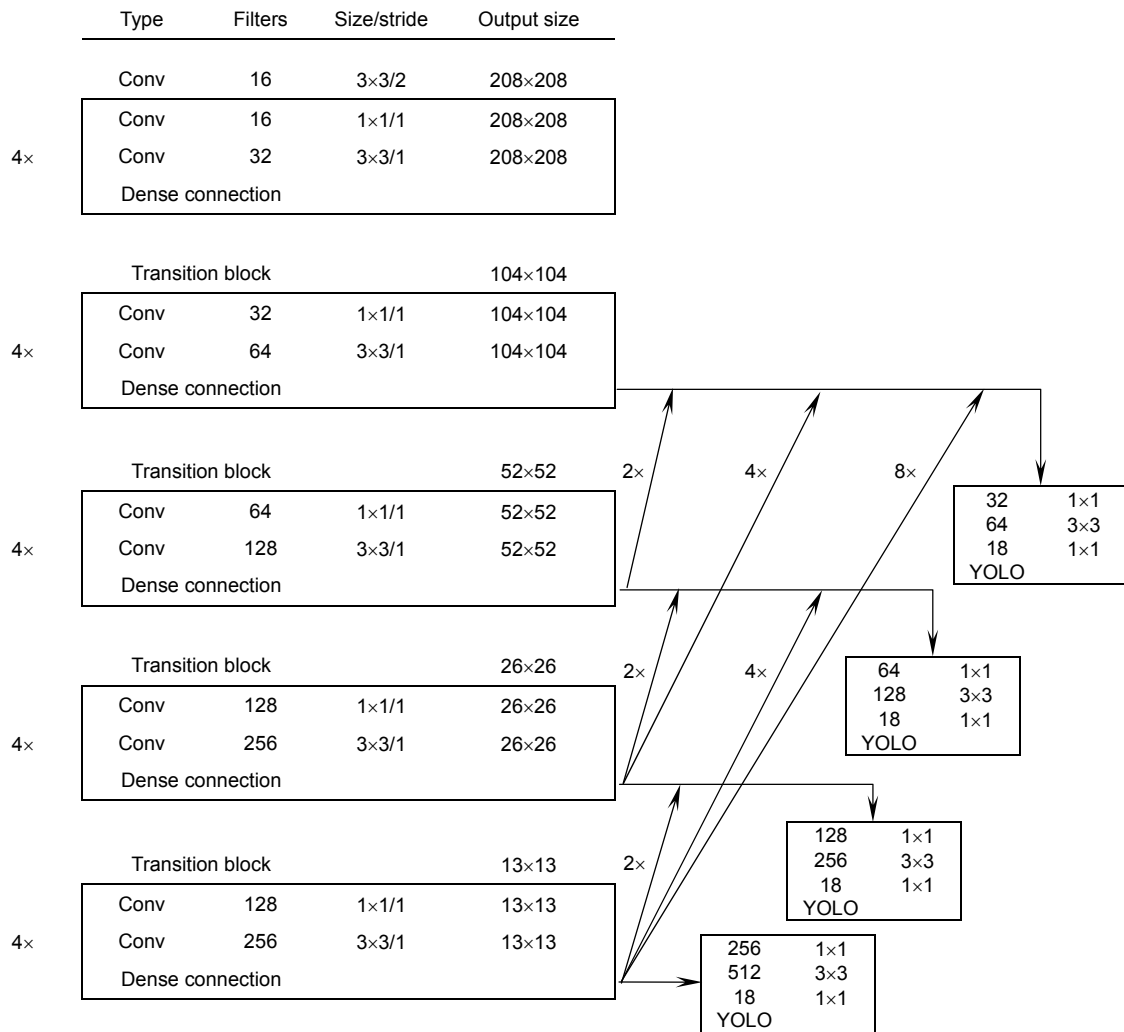


图 4 密集相连的多尺度检测

Fig. 4 Multi-scale detection with dense connection

## 4 实验结果与分析

本文的算法在开源框架 Darknet 上实现，电脑配置为 I7-8700 CPU，16G RAM，显卡 1070Ti，操作系统为 Windows 10。本文中，使用召回率、精度、 $F_1$  值、交叠比 IOU 与 AP 评价模型的性能。

精度：

$$P = \frac{T_p}{F_p + T_p} ;$$

召回率：

$$R = \frac{T_p}{F_p + F_N} ;$$

$F_1$  分数为

$$F_1 = \frac{2P \cdot R}{P + R} .$$

其中： $T_p$  为真正例， $F_p$  为假正例， $F_N$  为假负例。AP 为平均精度，用  $A_p$  表示，采用的是 VOC2007 中的标准，即从 PR 曲线中取召回率  $R$  划分为 (0, 0.1, 0.2, ..., 0.9, 1) 共 11 个等距区间后，求最大精度的平均值可获得平均精度  $A_p$ ：

$$A_p = \frac{1}{11} \sum_{R \in \{0, 0.1, \dots, 1\}} P_{\text{interp}}(R) ,$$

$$P_{\text{interp}}(R) = \max_{\tilde{R}, R \geq \tilde{R}} p(\tilde{R}) , \quad (4)$$

其中： $P_{\text{interp}}(R)$  为召回率满足  $\tilde{R} \geq R$  时的精度的最大值， $\tilde{R}$  表示获得精度最大值时的召回率。

在本文中，使用了两个评价算法处理速度的指标，一个为算法每秒处理的图像数量  $N$ ，其单位为帧每秒 (f/s)，另一个为算法处理每张图像所需的时间  $t=1/N$ ，其单位为 ms。一般而言，算法的处理速度达到 30 f/s 即可认为达到实时。

### 4.1 数据集

本文的数据集取材于谷歌地球, 选取 990 张不同天气情况、不同地区的机场遥感图像(如图 5), 使用 LableImg 对遥感图像中的飞机目标进行标注。随机选取其中 850 张用作训练集, 140 张用作测试集。数据集中的遥感图像的分辨率均为 1372×941。

机场遥感图像中包括光照剧烈(图 5(a))、图像模糊(图 5(b))、光照不足、背景斑杂、雾(图 5(c))等情况。

### 4.2 训练方法

在训练阶段, 采用的动量为 0.9, 使用小批量随机梯度下降进行优化, 初始学习率为  $\eta_{lr} = 0.001$ , 衰减系数设置为 0.005, 前 1000 批次训练中采用学习率为  $\eta_{learning\_rate} = \eta_{lr} \times (N_{batch} / 1000)^2$  以稳定网络,  $N_{batch}$  为当前的批次数, 随后学习率为  $10^{-3}$ 。通过调整饱和度和曝光量、色调来增加训练样本。

本文采用多尺度训练策略, 增强对不同尺寸图像的鲁棒性, 每经 10 批次训练随机选择新的图片尺寸进行训练, 图片尺寸为从 320×320 到 608×608, 间隔为 32 进行采样。

### 4.3 检测结果定量评估

将本文提出的算法 YOLOv3-air 与 YOLOv3、YOLOv3-tiny、YOLOv2 进行比较, 图像在训练前均缩放为 416×416, 性能对比见表 1。表中的数据使用的模型是取 AP 值最大的模型, 并且召回率、精度是通过调整阈值使  $F_1$  值最大获得。本文提出的算法在 140 张

的测试集中得到了精度 96.26%、召回率 93.81%、平均精度 89.31%, 相对于 YOLOv3-tiny 分别提升了 6%、13%、13%。YOLOv3-air 的平均精度相对于 YOLOv3-tiny 有大幅提升是因为 4 个尺度检测可以准确检测大部分的小目标使召回率上升。但是出乎意料之外的是, YOLOv3 的 AP 竟然逊色于 YOLOv3-air 不少, 只与 YOLOv3-tiny 相近。YOLOv3 在本实验中运行速度为 33.2 f/s, YOLOv3-tiny 为 215.2 f/s, 而 YOLOv3-air 为 58.3 f/s。YOLOv3-air 针对飞机数据集使用聚类算法重新计算了锚点框的大小, 使得的 IOU 超过 YOLOv3 和 YOLOv3-tiny 接近 4%。本文也提供了 YOLOv2 与 YOLOv2-tiny 在本文中的遥感图像测试集的检测结果, 可以看出 YOLOv2 缺少多尺度检测导致在召回率上远逊于 YOLOv3。

### 4.4 定性评估与误差分析

图 6(a)、图 6(b)、图 6(c)三个对比图中, 仅有 YOLOv3-air 可以检测出所有飞机目标, YOLOv3-tiny 则因为只有两个尺度检测因此难以有效检测小目标, YOLOv3 能检测出大部分的飞机目标, 但对于极小的飞机目标也无法处理。图 7(b)、图 7(d)中登机桥与飞机的外形十分相似, 并且在标注飞机目标时容易将与飞机邻接的登机桥一并标注, 使得算法容易将登机桥误认为飞机, 这也是限制算法性能的主要原因之一。

从图 6(d)、图 6(e)、图 6(f)可以看出, YOLOv3 的泛化能力相对于 YOLOv3-tiny 和 YOLOv3-air 有一定



图 5 飞机数据集实例

Fig. 5 Three samples of airplane dataset

表 1 5 种算法的性能对比

Table 1 Performance comparison of 5 algorithms

	$P/\%$	$R/\%$	$F_1/\%$	$AP/\%$	$R_{iou}/\%$	速度/(f/s)	处理时间/ms
YOLOv3	93.56	78.9	85.61	78.97	68.80	33.2	30.1
YOLOv3-tiny	90.82	83.05	86.76	78.99	67.05	215.2	4.6
YOLOv3-air	96.26	93.81	95.02	89.31	72.46	58.3	17.2
YOLOv2	87.11	62.27	72.62	60.92	60.28	47.5	21.1
YOLOv2-tiny	67.44	54.41	60.23	46.87	45.83	207.5	4.8

的差距,无法处理外表剧烈变化的飞机目标。分析可知,简单的卷积神经网络模型具有更好的泛化性,尤其是在数据集较小、数据复杂多变的情况下;YOLOv3 参数数量过大,导致 YOLOv3 过拟合。本文提出的 YOLOv3-air 通过减少参数数量和增加多尺度检测上结合了 YOLOv3-tiny 和 YOLOv3 的优点,在小目标检测

与泛化性上表现优异,并且通过密集连接复用特征减少了训练数据不足的影响。

为了进一步探讨训练集大小与算法性能的关系,表 2 中给出了在训练集只有 300 张遥感图像与只有 500 张遥感图像时,在相同测试集下 YOLOv3-air 的性能。在表 2 中, YOLOv3-air-500 表示训练集中只有 500 张

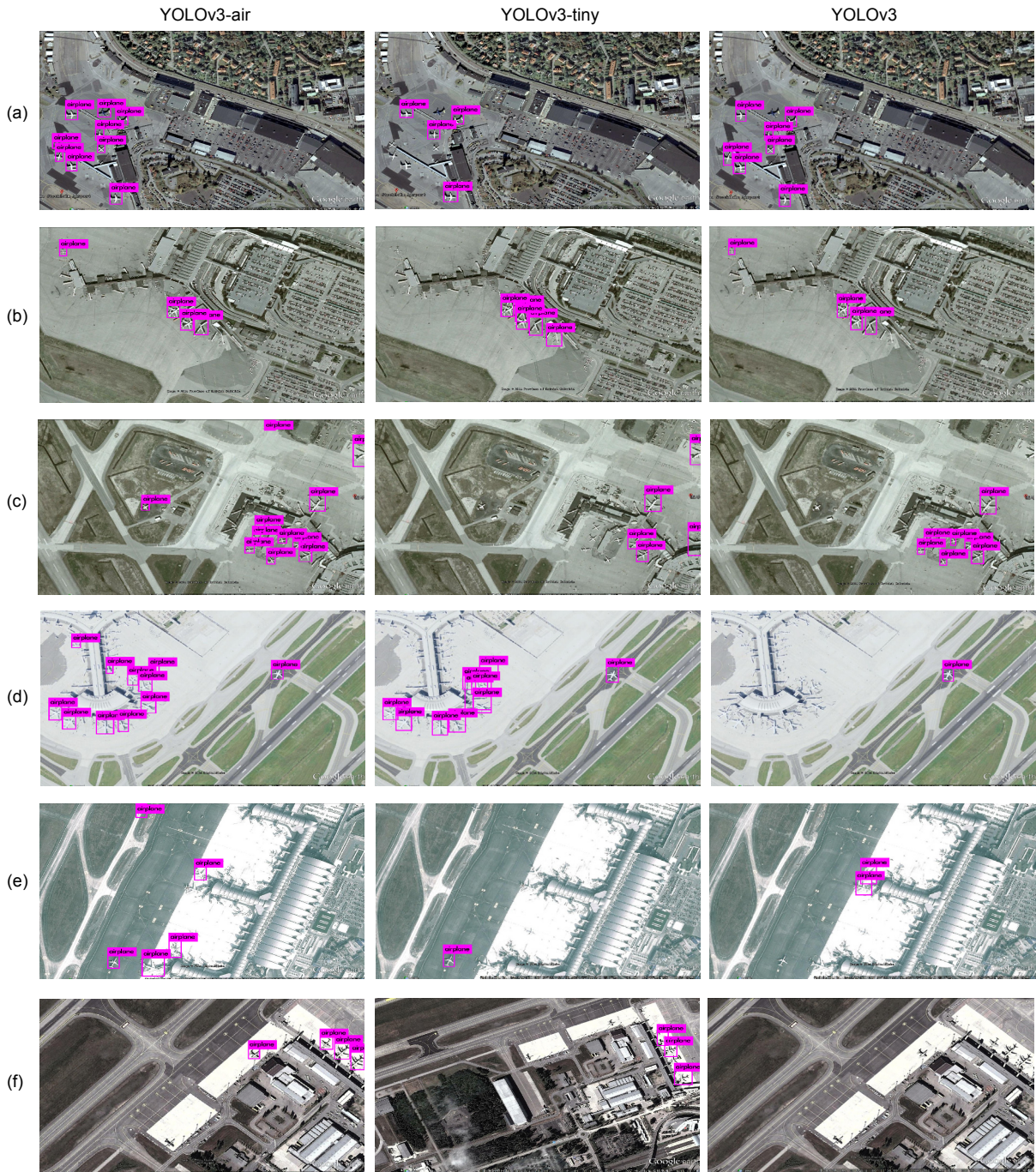


图 6 从左到右分别是 YOLOv3-air、YOLOv3-tiny、YOLOv3 的检测结果。(a) P883; (b) P902; (c) P903; (d) P909; (e) P866; (f) P867

Fig. 6 The detection results of YOLOv3-air, YOLOv3-tiny, YOLOv3 in order. (a) P883; (b) P902; (c) P903; (d) P909; (e) P866; (f) P867

表 2 在不同大小的训练集中, YOLOv3-air 的性能对比

Table 2 Performance comparison of YOLOv3-air with different number images in training set

	P/%	R/%	F <sub>1</sub> /%	A <sub>P</sub> /%	R <sub>IOU</sub> /%
YOLOv3-air	96.26	93.81	95.02	89.31	72.46
YOLOv3-air-500	93.47	87.25	90.25	86.53	70.74
YOLOv3-air-300	92.62	74.49	82.57	78.12	67.15

图像时的 YOLOv3-air, YOLOv3-air-300 表示训练集中只有 300 张图像时的 YOLOv3-air。

从表 2 可以看出, 当训练集的大小发生改变时, 在同样的测试集上, 算法的性能有一定的下降, 当训练集中只有 300 张遥感图像时, 算法的性能退化到与 YOLOv3, YOLOv3-tiny 相近。实验表明, 对于同一卷积神经网络, 训练集的大小会对算法的性能产生较大影响。在处理相同的问题时, 在训练集的数据量较大时, 可以使用参数量偏多的卷积神经网络进行处理, 因为充足的数据可以防止模型过拟合; 但在训练集的数据量较小时, 用参数量少的卷积神经网络进行处理会更为合适。当训练集的数据量偏少时, 会造成参数量大的卷积神经网络没有得到充分的训练, 过度拟合当前的训练集, 从而在测试集上的性能略差。

## 5 结论

本文主要基于 YOLOv3 对网络结构和多尺度检测进行改进提出了一种遥感图像飞机目标的实时检测算法, 并且自行设计标注了一个遥感图像飞机目标的训练与测试集。实验结果表明, 本文提出的 YOLOv3-air 对遥感图像中的飞机目标具有良好的鲁棒性, 在运行速度达到 58.34 f/s 的同时, 精度、召回率和平均精度分别达到了 96.26%、93.81% 和 89.31%。以上的工作验证了 YOLOv3 用于遥感图像中飞机目标检测的可行性和高效性, 但存在数据较少、目标标注引入干扰背景的问题, 后续的工作可以针对数据集进行优化扩展。

## 参考文献

- [1] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]// *Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, 2014: 580–587.
- [2] Girshick R. Fast R-CNN[C]// *IEEE International Conference on Computer Vision*. IEEE, 2015:1440–1448.
- [3] Ren S Q, He K M, 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 Systems*, Montreal, Canada, 2015: 91–99.

- [4] He K, Gkioxari G, Dollár P, et al. Mask R-CNN[C]// *IEEE International Conference on Computer Vision*. IEEE, 2017: 2980–2988.
- [5] Redmon J, Divvala S, Girshick R, et al. You Only Look Once: Unified, Real-Time Object Detection[C]// *Computer Vision and Pattern Recognition*. IEEE, 2016: 779–788.
- [6] Liu W, Anguelov D, Erhan D, et al. SSD: Single Shot MultiBox Detector[C]// *European Conference on Computer Vision*. Springer International Publishing, 2016: 21–37.
- [7] Redmon J, Farhadi A. YOLO9000: Better, Faster, Stronger[C]// *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2017: 6517–6525.
- [8] Redmon J, Farhadi A. YOLOv3: An Incremental Improvement[J]. arXiv preprint arXiv:1804.02767, 2018.
- [9] Xue Y J, Huang N, Tu S Q, et al. Immature mango detection based on improved YOLOv2[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2018, **34**(7): 173–179.  
薛月菊, 黄宁, 涂淑琴, 等. 未成熟芒果的改进 YOLOv2 识别方法[J]. *农业工程学报*, 2018, **34**(7): 173–179.
- [10] Wang S Y, Gao X, Sun H, et al. An aircraft detection method based on convolutional neural networks in high-resolution SAR images[J]. *Journal of Radars*, 2017, **6**(2): 195–203.  
王思雨, 高鑫, 孙皓, 等. 基于卷积神经网络的高分辨率 SAR 图像飞机目标检测方法[J]. *雷达学报*, 2017, **6**(2): 195–203.
- [11] Zhou M, Shi Z W, Ding H P. Aircraft classification in remote-sensing images using convolutional neural networks[J]. *Journal of Image and Graphics*, 2017, **22**(5): 702–708.  
周敏, 史振威, 丁火平. 遥感图像飞机目标分类的卷积神经网络方法[J]. *中国图象图形学报*, 2017, **22**(5): 702–708.
- [12] Gu Y, Xu Y. Fast SAR target recognition based on random convolution features and ensemble extreme learning machines[J]. *Opto-Electronic Engineering*, 2018, **45**(1): 170432.  
谷雨, 徐英. 基于随机卷积特征和集成超学习机的快速 SAR 目标识别[J]. *光电工程*, 2018, **45**(1): 170432.
- [13] Huang G, Liu Z, Maaten L V D, et al. Densely Connected Convolutional Networks[C]// *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE Computer Society, 2017: 2261–2269.
- [14] Lin T Y, Dollar P, Girshick R, et al. Feature Pyramid Networks for Object Detection[C]// *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE Computer Society, 2017: 936–944.



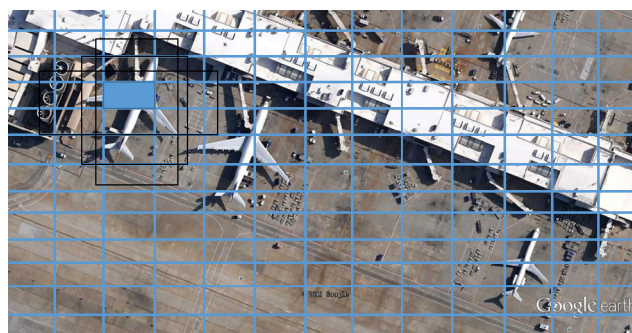
# Real-time airplane detection algorithm in remote-sensing images based on improved YOLOv3

Dai Weicong<sup>1,2\*</sup>, Jin Longxu<sup>1</sup>, Li Guoning<sup>1</sup>, Zheng Zhiqiang<sup>3</sup>

<sup>1</sup>Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, Jilin 130033, China;

<sup>2</sup>University of Chinese Academy of Sciences, Beijing 100049, China;

<sup>3</sup>Changchun University of Science and Technology, Changchun, Jilin 130022, China



An illustration of predicted bounding boxes on 13x13 grids of YOLOv3

**Overview:** The detection of airplanes in remote-sensing images has many important applications in many domains. However, limited to the performance of traditional machine learning methods, the airplanes in remote-sensing images are difficult to be detected. Recently, deep convolutional neural networks are employed to solve object detection problem and reach excellent accuracy. YOLO is one of the most famous real-time object detection algorithms based on regression. Compared with other algorithms, YOLO is more generalized when applied to many domains. Focusing on the airplanes in remote-sensing images, a real-time algorithm based on improved YOLOv3 is proposed to detect airplanes in remote-sensing images. Firstly, a convolutional neural network that consists of 49 convolutional layers is proposed to detect airplanes in remote-sensing images specifically. In the transition blocks of proposed convolutional neural network, we employ  $1\times 1$  convolution kernels to further reduce the parameters. Secondly, dense connection is employed on proposed convolutional neural network, and the maxpool is employed to enhance the feature transmit between two dense blocks. In this way, the feature transmit between two dense blocks is reconnected after a undersampling convolutional layer. The dense connection in proposed convolutional neural network enable the network to avoid over-fitting and reach high accuracy although the network is trained by relative few training data. Finally, to deal with the fact that airplanes in remote-sensing images are small targets mainly, we propose to increase the scale detections from 3 to 4 and employ dense connection to merge feature map among different scales. The anchor boxes in our work are obtained by running k-means clustering on the training set bounding boxes. The algorithm is trained and tested on the designed airplane dataset, which have 990 remote-sensing images. The qualitative experiment results show that our algorithm has stronger robustness than other existing algorithms, and our algorithm also shows especially high recall on small targets. The quantitative experiment results show that our algorithm obtains 96.26% on precision, 93.81% on recall and 89.31% on AP. Our algorithm reaches a relative improvement of 13.1% with respect to the YOLOv3 on AP. The detector proposed in this study is proven to perform real-time speed of more than 58.3 frames per second on a 1070Ti GPU. This study demonstrates the high effectiveness and accuracy of deep convolutional neural network in detecting airplanes on remote-sensing images. Meanwhile, the research also shows the fact that the performance of convolutional neural networks is decided by their structure and the number of training data.

**Citation:** Dai W C, Jin L X, Li G N, *et al.* Real-time airplane detection algorithm in remote-sensing images based on improved YOLOv3[J]. *Opto-Electronic Engineering*, 2018, 45(12): 180350

Supported by National High Technology Research and Development Program ("863"Program) of China (863-2-5-1-13B)  
\*E-mail: daiweicong16@mails.ucas.ac.cn