Website: ycyk.brit.com.cn Email: ycyk704@163.com

基于改进 SSD 算法的航拍目标检测算法研究*

黄奕川，李凉海，马纪军，崔慧敏
（北京遥测技术研究所 北京 100076）

摘要：研究基于深度学习技术的无人机航拍图像目标检测算法，首先介绍目标检测算法 SSD（Single Shot MultiBox Detector），并对其特征提取网络进行改进，采用稠密特征提取网络替换原网络的主干特征提取网络，提高算法的特征提取能力，从而提升了算法的检测精度。针对网络实时性问题，在算法中引入分组卷积，极大地减少了网络参数量，提升了网络推理速度。为解决训练中出现的正负样本不均衡问题，利用焦点损失（Focal Loss）改进了原算法的损失函数，进一步提升了网络的收敛速度和精度。最后，通过仿真验证了改进算法在目标检测精度上的优越性。

关键词：人工智能；深度学习；目标检测；图像处理；无人机航拍

中图分类号：V248.1 **文献标识码：**A **文章编号：**CN11-1780(2022)03-0079-08

DOI：10.12347/j.ycyk.20211229001

引用格式：黄奕川, 李凉海, 马纪军, 等. 基于改进 SSD 算法的航拍目标检测算法研究[J]. 遥测遥控, 2022, 43(3): 79–85.

Research on aerial target detection algorithm based on improved SSD algorithm

HUANG Yichuan, LI Lianghai, MA Jijun, CUI Huiming
（Beijing Research Institute of Telemetry, Beijing 100076, China）

Abstract: This paper studies the problem of target detection in UAV aerial images using deep learning technology. Let's start with single shot multibox detector target detection algorithm, and then improve it. Dense feature extraction network is used to improve the feature extraction ability of the algorithm, so as to improve the detection accuracy of the algorithm. Aiming at the problem of network real-time, packet convolution is introduced into the algorithm, which greatly reduces the amount of network parameters and improves the speed of network reasoning. In order to solve the positive and negative problems in training for the sample imbalance problem, this paper improves the loss function of the original algorithm and uses focal loss to improve the original loss function, which further improves the convergence speed and accuracy of the network. Finally, the superiority of the algorithm in target detection accuracy and speed is verified by simulation.

Key words: Artificial intelligence; Deep learning; Object detection; Image processing; UAV aerial photography

DOI：10.12347/j.ycyk.20211229001

Citation: HUANG Yichuan, LI Lianghai, MA Jijun, et al. Research on aerial target detection algorithm based on improved SSD algorithm[J]. Journal of Telemetry, Tracking and Command, 2022, 43(3): 79–85.

引 言

近年来，随着计算平台和深度学习技术的发展，目标检测技术在各个领域都得到了广泛地应用。将目标检测技术搭载在无人机上，通过对无人机拍摄得到图片进行智能图像处理与分析，从而得到目标在图像中的类别与位置信息，可以极大地扩展无人机的应用场景。随着无人机在社会生活中的应用越来越广泛，对航拍图像的分析需求逐年增大，将目标检测技术与无人机航拍技术相互结合逐渐成为当前的研究热点之一[1]。

不同于技术相对成熟的地面静态目标检测，当前无人机航拍目标检测还存在许多未解决的问题，这些问题主要可以归结于四个方面：① 检测目标数量大。与一般的目标检测数据集不同，无人机目标检测往往有可能在一张图片中包含上百个需要检测的模型，如果选用占用资源较大的算法，可能会出现资

*基金项目：国家自然科学基金（61903044）
收稿日期：2021-12-29 收修改稿日期：2022-02-13

源不够用的情况[2]。② 目标尺寸多样性。无人机在拍摄图像时，由于其飞行高度、拍摄角度、物体远近和目标类别的不同，拍摄得到的样本大小也多种多样，这给深度学习算法的训练带来了一定挑战[3]。③ 运动模糊。相比于静止平台上摄像头的成像，无人机飞行过程中摄像头成像容易出现运动模糊，这给目标检测算法带来了一定难度[4]。④ 算力限制。由于无人机整机功耗和尺寸的限制，无人机在进行实时目标检测时对于算法复杂度有较高要求[5]。

近年来，深度学习技术在计算机视觉方向上有较大的进步与发展，以卷积神经网络为代表的深度学习算法相较于传统的手工特征目标检测算法，在检测精度与速度上都有显著提升，逐渐取代了传统目标检测算法，成为了当前目标检测技术的主流算法。基于深度学习的目标检测算法主要分为单阶段和两阶段两种[6]。单阶段目标检测算法采用端对端的方式[7]，直接在原图中提取特征，并以此来预测目标的类别和位置。两阶段目标检测算法将检测过程分成了两步进行：首先，通过候选区域生成算法生成候选区域；然后，将生成得到的候选区域输出到后续网络，以确定目标的所属类别和位置[8]。单阶段目标检测算法的结构相对较为简单，且检测速度更快，但是其定位精度往往不如两阶段目标检测算法[9]。

由于无人机本身算力的限制和对实时性的要求，本文选用 SSD 单阶段目标检测算法作为基准算法。并针对无人机航拍图像中目标尺寸多样、算力有限等问题，对 SSD 算法做出相应改进，最后对比改进前后的算法效果，体现改进的有效性。

1　SSD 目标检测算法算法原理

SSD 算法使用 VGG16 作为基础特征提取网络[10]，在保证和 Faster R-CNN 算法具有同样检测精度的同时，检测速度快于 YOLO 算法[11]。算法通过骨干网络（Backbone）后得到 38×38×512 的特征图，然后将特征图进行下采样，依次得到 19×19×1 024、10×10×512、5×5×256、3×3×256、1×1×256 的特征图，然后将这些特征图作为检测预测层的输入，在不同维度下预测目标框的位置，最后再经过极大值抑制算法 NMS（Non-Maxmum Suppression），输出预测目标框的位置和类别[12]，其结构如图 1 所示。

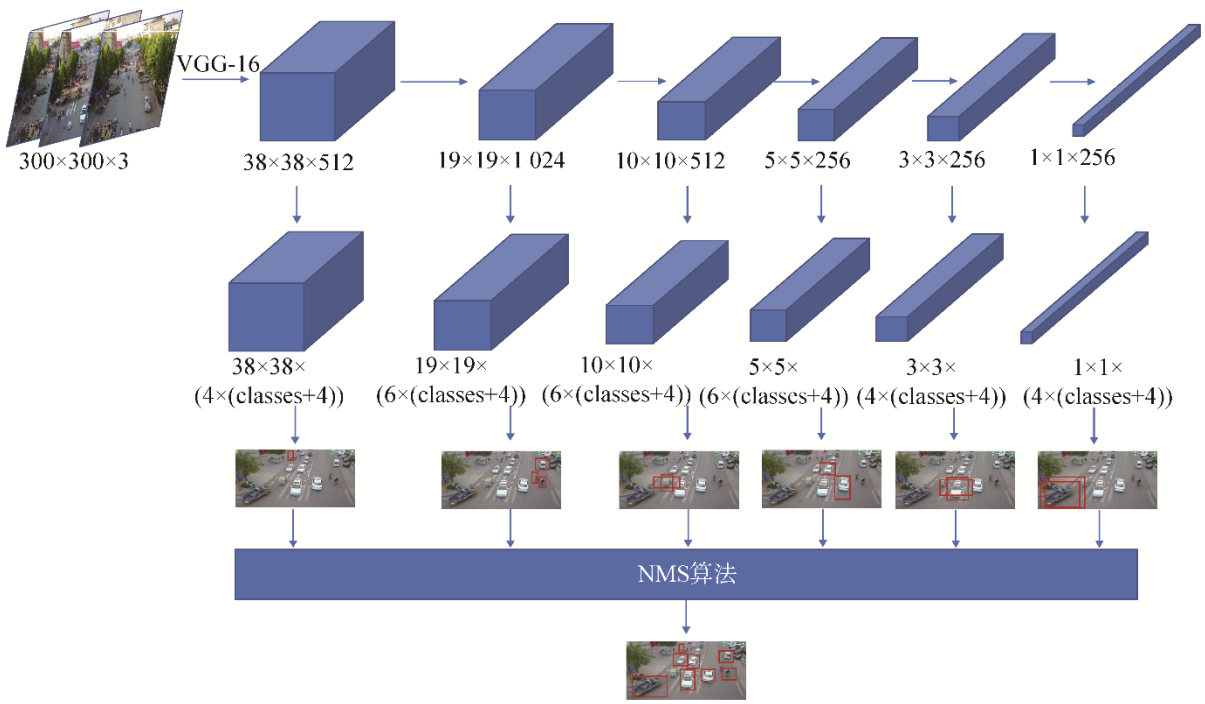


图 1　SSD 网络框架图

Fig. 1　The structure of SSD

SSD 算法通过在基础特征提取网络后添加卷积特征层，这些层的尺寸逐渐缩小，通过在这些不同大小的特征层上进行训练和预测，利用多层特征图的组合作为预测目标类别和位置的依据，以达到多尺度融合效果[13]，确保了检测算法对于多尺度目标的有效性。但是，由于 SSD 算法使用浅层网络的特征信息检测小目标，而浅层特征卷积层数少，缺乏深层语义特征，表征能力不够强，导致对小目标物体的召回率很低。

2　算法改进

2.1　主干网络改进

在 SSD 原始模型中，采用了 VGG16 作为检测算法的主干网络。VGG16 由 13 个卷积层、3 个全连接层和 5 个池化层组成，其突出特点是结构简单，网络通过堆叠若干卷积层和池化层构成，网络的搭建较为容易实现。其缺点在于，VGG16 的层数较浅，特征提取并不充分，特别是在对小目标的检测上，性能并不理想。

本文参考 CVPR2017 上提出的 Densenet-121 结构作为主干网络[14]。稠密网络（Densenet）是由稠密块（Dense block）模块加上过度（Transition）模块组合而成，Densenet 由若干 Dense block 组成，每个 Dense block 由若干卷积层组成，稠密块中的特征图大小一致。相比于其他网络，Densenet 提出了一种稠密连接方法，稠密块中每个卷积层都会接受其前面所有层的输出作为其额外的输入，网络通过这种稠密连接的方式加强了不同网络层之间的特征复用，更有效地利用了特征。传统网络结构可以表示为

$$x_l = H_l(x_{l-1}) \tag{1}$$

而 Densenet 将同一稠密块中前面所有层的输出拼接起来作为输入，可以表示为

$$x_l = H_l(x_0, x_1, \cdots, x_{l-1}) \tag{2}$$

其中，x_l 表示第 l 层的特征图，$H_l(\cdot)$ 代表第 l 层输入到输出的映射，Dense block 示意图如图 2 所示。

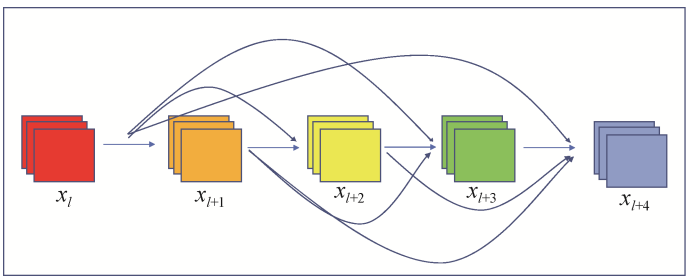


图 2　稠密块示意图

Fig. 2　Schematic of a dense block

在 Densenet 中不同的 Dense block 用 Transition 层连接在一起，Transition 层由 1 个 1×1 的卷积层和 1 个 2×2 的平均池化层组成，其作用是改变特征图的通道数，Densenet 的总体架构如图 3 所示。

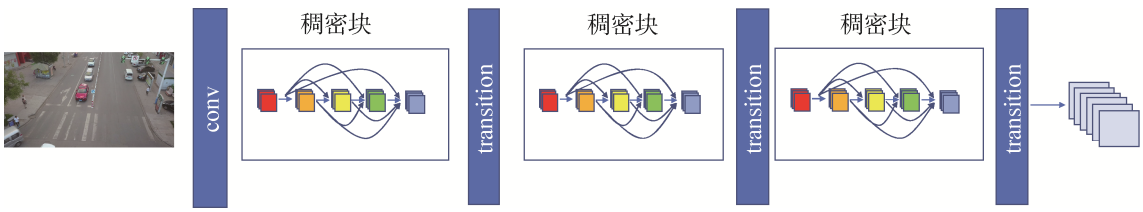


图 3　稠密特征提取网络架构图

Fig. 3　The structure of Densenet

Densenet 相比于 SSD 原始特征提取网络 VGG，使用了更加稠密的网络连接，即互联了同一稠密块中的所有层，其每一层都会接受前面所有层的输出作为其额外输入。这既可以增强特征之间的传播，又在一定程度上增强了不同层次特征的复用，另外，Densenet 也在一定程度上缓解了网络中的梯度消失问题，使得浅层特征层的训练更为有效。通过对特征提取网络的改进，可以使得图像的特征提取更加充分，

提升了网络的检测精度。

2.2　分组卷积结构

本文在原卷积神经网络的结构上引入分组卷积[15]。不同于一般卷积操作，分组卷积在卷积操作之前，首先对输入的特征层进行分组，将经过分组操作得到的不同特征层分组分别进行卷积操作，最后将不同分组的输出结果拼接起来，得到分组卷积的输出，其示意图如图 4 所示。

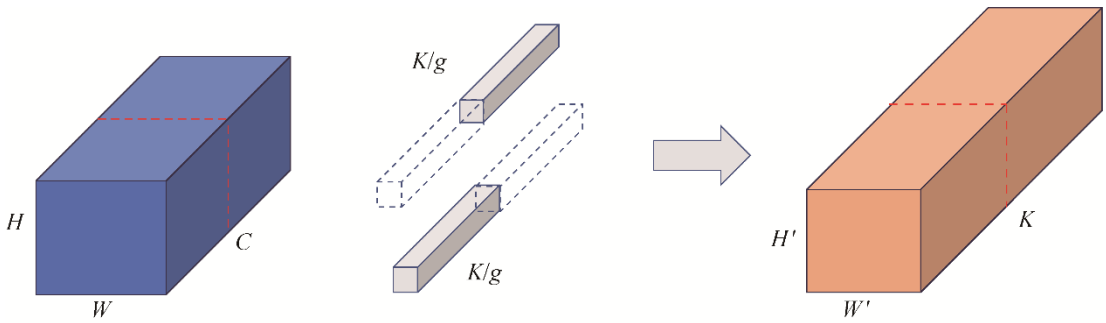


图 4　分组卷积示意图

Fig. 4　Schematic of a group convolution

图 4 中，H 代表输入图像的高度，W 代表输入图像的宽度，C 代表输入图像的深度，g 代表分组卷积所分的组数，H' 代表输出图像的高度，W' 代表输出图像的宽度，K 代表输出特征图的深度。分组卷积相比于一般卷积操作，极大地减少了网络参数量。一般卷积和分组卷积的参数量分别为

$$\begin{aligned} Par &= inc \times outc \times ken^2 \\ g\_Par &= (inc/g) \times (outc/g) \times ken^2 \times g \end{aligned} \tag{3}$$

式（3）中，Par 代表一般卷积操作的参数量，g_Par 代表分组卷积的参数量，inc 代表输入特征图的通道数，$outc$ 代表输出特征图的通道数，ken 代表卷积核的大小，分组卷积的参数量是一般卷积的 g 分之一。在网络中使用分组卷积，可以降低网络的参数量，提升网络的训练和推理的速度。

2.3　引入焦点损失

焦点损失（Focal Loss）主要为了解决单阶段目标检测算法中出现的正负样本不均衡问题[16]。在单阶段目标检测算法中，由于一张图片中往往大部分位置都属于背景，这将导致训练集出现正负样本不均衡的问题。过多的负样本会引起网络权重的偏移，从而影响正样本的预测，并且这些负样本大多往往是易分类的，这会严重影响网络的收敛速度。

为了有效解决类别不均衡的问题，本文在 SSD 中引入 Focal Loss 损失函数

$$\mathrm{FL}(p_t) = -\alpha_t (1 - p_t)^{\gamma} \ln(p_t) \tag{4}$$

式（4）中，$\mathrm{FL}(p_t)$ 为 Focal Loss 损失函数，α_t 为权重因子，p_t 为样本预测正确的概率，γ 为调制系数。Focal Loss 通过权重因子控制正负样本对损失函数的贡献比例，在损失函数中减少负样本的权重，可以在一定程度上缓解正负样本不均衡带来的影响。另一方面，Focal Loss 引入调制系数，减少了易分类样本对损失函数的贡献，使得模型训练时更加专注于难分类的样本。

3　仿真实验

3.1　实验数据集

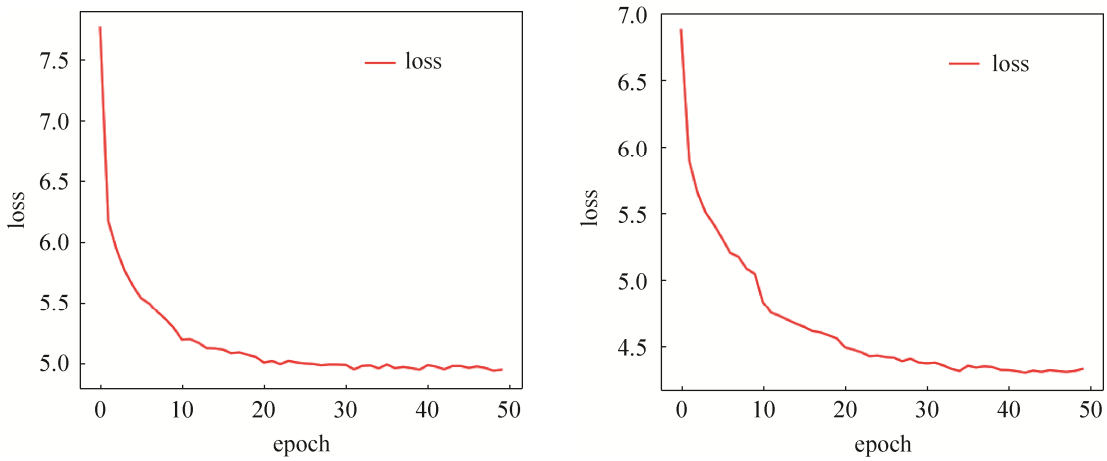
实验选用 VisDrone2019 数据集作为训练样本，该数据集由主办方天津大学负责收集，全部数据由无人机采集得到。检测目标分为 10 类，包括行人、不在行走的人、脚踏车、汽车、面包车、三轮车、带遮阳伞的三轮车、公交车、摩托车、其他。训练集图片尺寸为 1 920×1 080，其中最大目标占图像面积

的 30.29 %，而最小目标仅占图像面积的二十万分之三，数据集中目标有密度分布不均、尺度变化多样、相互遮挡的情况存在。

3.2　网络训练与测试

实验平台配置为操作系统：Ubuntu20.04，深度学习库：Pytorch，CPU：Intel i7-10700，GPU：RTX3080。本文对 SSD 和改进 SSD 网络分别进行训练，SSD 将图像预处理为 300×300 大小作为网络输入，改进 SSD 网络将图像预处理为 512×512 大小作为网络输入，设置网络的学习率为 0.01，动量为 0.9，学习率衰减系数为 0.3，batch-size 为 16，分别迭代 50 次。

SSD 算法和改进 SSD 算法的损失函数收敛曲线如图 5 所示。



（a）原算法损失函数曲线
（a）Loss function curve of original algorithm

（b）改进算法损失函数曲线
（b）Loss function curve of improved algorithm

图 5　网络损失函数曲线
Fig. 5　Network loss function curve

由图 5 可以看出，SSD 算法在训练损失为 5 左右到达收敛，而改进 SSD 算法在训练损失达到 4.3 左右收敛，算法经过改进以后，收敛性能明显好于原算法。用训练得到的网络对航拍图像目标进行检测，其结果如图 6 所示，可以看出改进算法对于行人、摩托车、三轮车等难分类中小目标的检测性能明显高于原算法。

（a）原算法检测效果图
（a）Original algorithm detection rendering

（b）改进算法检测效果图
（b）Improved algorithm detection rendering

图 6　算法效果图
Fig. 6　Algorithm renderings

3.3　实验结果分析

实验比较算法改进前后的预测结果，设置交并比阈值为 50 %，在 Visdrone2019-DET 验证集下，算法对各种目标的平均检测精度见表 1。

表 1　原算法与改进算法各类别平均精度

Table 1　The average percision of the original algorithm and the improved algorithm in each category

类别	SSD 算法/(%)	改进 SSD/(%)
行走的人	17.2	24.9
人（非行人）	9.2	15.1
脚踏车	5.4	11.6
汽车	53.8	57.0
面包车	37.5	34.8
货车	46.2	50.7
三轮车	16.1	11.1
带遮阳伞的三轮车	14.1	37.3
公交车	55.5	66.3
摩托车	16.8	24.1
总体	24.0	29.3

经过对表 1 的分析可以看出，数据集的总体平均精度均值 MAP（Mean Average Precision）从 24.0 %提升至 29.3 %，这是由于改进 SSD 算法使用更优的主干网络和更高分辨率图像作为输入，对于大多数类别目标的检测性能都有一定程度的提升。表 1 中行人、人（非行人）、脚踏车、带遮阳伞的三轮车、公交车、摩托车目标检测性能提升较大，其中带遮阳伞的三轮车的平均精度从 14.1 %提升到 37.3 %，提升幅度达到 23.2 %。这说明改进后的 SSD 算法对于航拍图像目标的检测性能有了较大提升，而面包车、三轮车的检测精度小幅度降低，是由于改进 SSD 算法使用的分组卷积在减小算法运算量的同时，也在一定程度上引入了稀疏连接，减弱了卷积的表达能力，从而在一定程度上影响了部分对分辨率改变不敏感类别目标的检测精度。

表 2　原算法与改进算法各尺度平均检测精度

Table 2　The average precision of the original algorithm and the improved algorithm at each scale

目标尺度	SSD 算法/(%)	改进 SSD/(%)
小目标	4.1	10.6
中等目标	31.2	36.7
大目标	60.4	62.2

为了进一步说明改进算法对不同尺度目标的检测能力，从数据集中分别筛选出面积小于 322 像素的目标作为小目标，大于 322 像素小于 962 像素的目标作为中等目标，大于 962 像素的目标作为大目标。对这三种目标的检测精度进行统计，结果见表 2。

分析表 2 可以看出，改进 SSD 算法主要在小目标和中等目标的检测上相比于原始 SSD 算法有较大提升，小目标平均检测精度均值从 4.1 %提升到了 10.6 %，中等目标平均检测精度均值从 31.2 %提升到了 36.7 %。这主要是由于本文对网络的改进，增强了对难分类中小目标的识别能力。而大目标检测也有小幅度提升，本文认为这是因为优化了特征提取网络结构，提升了对不同尺度目标的特征提取能力。

综上，改进后的模型能显著提高对航拍目标的检测能力。

4　结束语

本文针对无人机航拍中的问题和研究现状，在 SSD 算法的基础上对原算法进行改进，提出了基于改进 SSD 算法的航拍目标检测算法。训练阶段，通过学习 50 轮训练数据集，原算法和改进算法分别在训练损失为 5 和 4.3 处达到收敛，改进算法的收敛效果好于原算法，这表明改进算法在训练集上的检测精度高于原算法。验证阶段，原算法和改进算法的平均检测精度均值分别为 24.0 %和 29.3 %，算法经过改进后平均精度均值提升 5.3 个百分点，进一步验证了算法改进的有效性。

参考文献

[1]　李诚龙, 屈文秋, 李彦冬, 等. 面向 eVTOL 航空器的城市空中运输交通管理综述[J]. 交通运输工程学报, 2020, 20(4): 35–54.

LI Chenglong, QU Wenqiu, LI Yandong, et al. Overview of traffic management of urban air mobility(UAM)with eVTOL aircraft[J]. Journal of Traffic and Transportation Engineering, 2020, 20(4): 35–54.

[2]　卢伟. 基于深度学习的无人机航拍图像目标检测[D]. 厦门: 厦门大学, 2019.

[3] 邢姗姗, 赵文龙. 基于 YOLO 系列算法的复杂场景下无人机目标检测研究综述[J]. 计算机应用研究, 2020, 37(S2): 28–30.

[4] 江波, 屈若锟, 李彦冬, 等. 基于深度学习的无人机航拍目标检测研究综述[J]. 航空学报, 2021, 42(4): 137–151.
JIANG Bo, QU Ruokun, LI Yandong, et al. Object detection in UAV imagery based on deep learning: Review[J]. Acta Aeronautica ET Astronautica Sinica, 2021, 42(4): 137–151.

[5] 裴伟, 许晏铭, 朱永英, 等. 改进的 SSD 航拍目标检测方法[J]. 软件学报, 2019, 30(3): 738–758.
PEI Wei, XU Yanming, ZHU Yongying, et al. The target detection method of aerial photography images with improved SSD[J]. Journal of Software, 2019, 30(3): 738–758.

[6] 鲁博, 瞿绍军. 融合 BiFPN 和改进 Yolov3-tiny 网络的航拍图像车辆检测方法[J]. 小型微型计算机系统, 2021, 42(8): 1694–1698.
LU Bo, QU Shaojun. Vehicle detection method in aerial images based on BiFPN and improved Yolov3-tiny network[J]. Journal of Chinese Computer Systems, 2021, 42(8): 1694–1698.

[7] 程旭, 宋晨, 史金钢, 等. 基于深度学习的通用目标检测研究综述[J]. 电子学报, 2021, 49(7): 1428–1438.
CHENG Xu, SONG Chen, SHI Jingang, et al. A survey of generic object detection methods based on deep learning[J]. Acta Electronica Sinica, 2021, 49(7): 1428–1438.

[8] 吴雪, 宋晓茹, 高嵩, 等. 基于深度学习的目标检测算法综述[J]. 传感器与微系统, 2021, 40(2): 4–7, 18.
WU Xue, SONG Xiaoru, GAO Song, et al. Review of target detection algorithms based on deep learning[J]. Transducer and Microsystem Technologies, 2021, 40(2): 4–7, 18.

[9] 鞠默然, 罗海波, 王仲博, 等. 改进的 YOLO V3 算法及其在小目标检测中的应用[J]. 光学学报, 2019, 39(7): 245–252.
JU Moran, LUO Haibo, WANG Zhongbo, et al. Improved YOLO V3 algorithm and its application in small target detection[J]. Acta Optica Sinica, 2019, 39(7): 245–252.

[10] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]//European Conference on Computer Vision, 2016: 21–37.

[11] 唐聪, 凌永顺, 郑科栋, 等. 基于深度学习的多视窗 SSD 目标检测方法[J]. 红外与激光工程, 2018, 47(1): 290–298.
TANG Cong, LING Yongshun, ZHENG Kedong, et al. Object detection method of multi-view SSD based on deep learning[J]. Infrared and Laser Engineering, 2018, 47(1): 290–298.

[12] 卢健, 何金鑫, 李哲, 等. 基于深度学习的目标检测综述[J]. 电光与控制, 2020, 27(5): 56–63.
LU Jian, HE Jinxin, LI Zhe, et al. A survey of target detection based on deep learning[J]. Electronics Optics & Control, 2020, 27(5): 56–63.

[13] 黄豪杰, 段先华, 黄欣辰. 基于深度学习水果检测的研究与改进[J]. 计算机工程与应用, 2020, 56(3): 132–138.
HUANG Haojie, DUAN Xianhua, HUANG Xinchen. Research and improvement of fruits detection based on deep learning[J]. Computer Engineering and Applications, 2020, 56(3): 132–138.

[14] HUANG G, LIU Z, MAATEN L V D, et al. Densely connected convolutional networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2017: 2261–2269.

[15] ZHANG Xiangyu, ZHOU Xinyu, LIN Mengxiao, et al. ShuffleNet: An extremely efficient convolutional neural network for mobile devices[EB/OL]. arXiv: 1707.01083, 2017.

[16] LIN Tsung-Yi, GOYAL Priya, GIRSHICK Ross, et al. Focal loss for dense object detection[EB/OL]. arXiv: 1708.02002, 2017.

[作者简介]

黄奕川　1997 年生，硕士研究生，主要研究方向为模式识别和图像处理。

李凉海　1965 年生，硕士，研究员，主要研究方向为雷达系统设计。

马纪军　1986 年生，硕士，高级工程师，主要研究方向为控制技术和系统。

崔慧敏　1990 年生，博士，高级工程师，主要研究方向为高精度伺服控制和基于深度学习的图像处理。

(本文编辑：傅　杰)