

# 基于改进Yolov5航拍图像目标检测

王海, 孙兆永, 田益民, 张赛赛, 沈立文

北京印刷学院信息工程学院, 北京

收稿日期: 2023年4月8日; 录用日期: 2023年6月7日; 发布日期: 2023年6月19日

## 摘要

本文针对航拍图像检测中可提取目标特征少、小目标多导致检测精度低的问题, 提出了一种基于原始YOLOv5算法的改进目标检测算法。首先, 将CBAM注意力机制部署到原始YOLOv5的特征提取和特征融合网络的C3模块中, 以提升主干网络中网络局部特征捕获与融合能力, 其次, 针对在不断采样过程中导致浅层次特征信息丢失的问题, 在检测头部分, 增加了一个检测头, 用于检测小目标。最后, 对原始YOLOv5的Mosaic数据增强方式进行改进, 将4张图片拼接改为9张图片拼接, 然后根据拼接图片的矩形面积对灰色背景进行裁剪, 加快模型收敛, 提高训练效率, 在不增加模型计算量的前提下, 改善背景复杂等问题。本文使用VisDrone2019数据集, 并基于改进的算法与原始算法分别进行了实验, 结果表明本文的算法能够针对航拍图像目标进行有效的检测。

## 关键词

目标检测, 注意力机制, Mosaic数据增强, YOLOv5

# Target Detection Based on Improved Yolov5 Aerial Image

Hai Wang, Zhaoyong Sun, Yimin Tian, Saisai Zhang, Liwen Shen

School of Information Engineering, Beijing Institute of Graphic Communication, Beijing

Received: Apr. 8<sup>th</sup>, 2023; accepted: Jun. 7<sup>th</sup>, 2023; published: Jun. 19<sup>th</sup>, 2023

## Abstract

In this paper, based on the original YOLOv5 algorithm, an improved target detection algorithm is proposed to solve the problem of low detection accuracy caused by few extracted target features and many small targets in aerial image detection. Firstly, CBAM attention mechanism is deployed in the C3 module of the original YOLOv5 feature extraction and feature fusion network to improve the local feature capture and fusion capability in the backbone network. Secondly, to solve the

problem that shallow level feature information is lost in the process of continuous sampling, a detection header is added to the detection header, used to detect small targets. Finally, the Mosaic data enhancement method of the original YOLOv5 was improved, and the Mosaic of 4 images was changed to 9 images, and then the gray background was cropped according to the rectangular area of the Mosaic images, In this way, the model convergence can be accelerated, training efficiency can be improved and complex background can be improved without increasing the calculation amount of the model. In this paper, we use VisDrone2019 data set to carry out experiments based on the improved algorithm and the original algorithm respectively. The results show that the proposed algorithm can effectively detect aerial image targets.

## Keywords

Target Detection, Attention Mechanism, Mosaic Data Enhancement, YOLOv5

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

近年来,随着深度学习及其机器学习的发展,以深度学习为代表的新兴技术凭借其优秀的泛化能力,在人工智能的研究领域取得了巨大的进展。随着国家对通航飞机以及无人飞行器发展的不断重视,由小型飞行器拍摄的航拍图像目标检测技术日益成为国内外计算机视觉领域研究的热点。无人机拍摄的图像对我们的生活有着很多积极的影响,无人机可以实现高分辨率影像的采集,可以在农业、航拍、快递运输、灾难救援、监控等等领域应用。在完成拍摄之后,需要对无人机机载相机拍摄到的画面进行图像信息的自动化采集和分析,其中对有价值的目标进行检测是完成各项任务的基础,目标检测的精确度将起着至关重要的作用。近年来,由于航拍图像数据集数量的增加和计算机计算能力的提高,检测精度也取得了一些不错的提升,由于无人机高度以及拍摄角度的问题,无人机拍摄的图像普遍都存在高分辨率、背景复杂、目标较小和微小目标聚集等特点,传统用于自然图像的通用深度学习检测器在遥感图像上的检测效果并不能令人满意,为此研究者针对遥感图像特点对这些检测器做出了改进。其中,汪鹏等[1]则以 YOLOv3 为基础,通过引入 DIoU 损失函数使预测框定位更加准确。赵文清等[2]针对光学遥感图像中多尺度物体检测差的问题,使用了一种跨尺度连接操作来增强模型的特征提取能力;李坤亚等[3]在骨干网络中引入通道全局注意力机制来降低冗余信息对模型的干扰。在局部特征表达加强上,赵文清等人[4]针对遥感场景中多尺度目标的检测精度低、目标特征信息易丢失等问题,改进出一种基于高分辨率结构的长短期记忆网络结构;Gong 等[5]还将自注意力机制引入 YOLOv5 以解决小目标密集分布的问题,虽然增加了部分计算量,但取得了不错的检测效果;周华平等[6]根据遥感图像目标密集分布且尺寸较小导致的检测困难的问题,将频率通道注意力引入 YOLOv5 中,提升了模型对小目标的检测效果;李惠惠等[7]为解决遥感图像中目标遮挡与背景复杂的问题,对 YOLOv5 的锚框进行了改进,并将卷积注意力模块引入到模型中,通过强化目标特征以提升目标检测精度。但是由于航拍图像具有背景复杂、小目标较多、遮挡的特点,始终影响算法的检测精度。因此提升航拍图像目标检测精度仍然一个有价值的研究方向。

## 2. 算法简介

### 2.1. 基本算法

传统的目标检测算法主要是有四个部分,基于深度学习的目标检测方法相较于传统的算法而言,融

合了手动设计特征提取器以及后处理, 将这四个部分的任务交给神经网络, 神经网络统一实现分类和回归, 实现端到端的学习。在基于深度学习的图像目标检测方法上, 检测方法有单阶段算法和两阶段算法两种。两阶段目标检测算法将检测过程分成了两个阶段, 首先产生候选区域, 然后对候选区域进行分类, 它的特点是算法识别错误率低但是检测速度较慢, 不能够满足实时检测场景。经典的两阶段检测方法有 R-CNN [8]、FastR-CNN [9]和 Faster R-CNN [10]等。单阶段检测方法不需要生成候选框, 它可以直接对整幅图像进行预测, 将图像上的所有区域都视为潜在感兴趣区域用于预测类的概率、边界框位置和置信度分数。经典的单阶段目标检测算法有 SSD [11]以及 YOLO 系列[12][13][14][15]等。相比两阶段检测方法, 单阶段检测方法在一个阶段中完成了两阶段算法做的定位和分类两个步骤, 所以具有更高的效率并且也大大降低了计算资源的消耗。

## 2.2. 网络结构模型

YOLOv5 由 Ultralytics 于 2020 年 5 月提出[16], 是基于 YOLOv4 基础上进行改进得来的一阶段检测器, 它具有不错的检测精度。YOLOv5 有 YOLOv5s、YOLOv5m、YOLOv5l、YOLOv5x 四个模型。YOLOv5s 以及 YOLOv5x 是检测性能与检测精度的两个极端。YOLOv5m 是一个中等大小的模型, 在精度与速度之间提供了一个较好的平衡, 适用于本文的航拍图像目标检测情况, 所以本文选择 YOLOv5m 作为基础的算法模型。YOLOv5 的网络结构分为三个部分, 骨干网络、颈部结构、头部结构。骨干网络是由卷积模块(Conv)、特征提取模块(C3)、空间金字塔模块(SPPF)组成; 随着卷积的加深, 提取到的特征越来越丰富, 颈部结构就负责将这些深层次与浅层次的特征融合起来, 最终就会产生三个不同尺度的输出, 分别为  $80 \times 80$ 、 $40 \times 40$ 、 $20 \times 20$ ; 头部结构只有一个 Detect 模块, 它是由三个简单的  $1 \times 1$  卷积构成的, 三个卷积分别对应着三个检测特征层。

## 3. 改进 YOLOv5 航拍图像目标检测算法

### 3.1. 改进思路

由于航拍图像与自然图像间的较大差异, 航拍图像普遍存在背景信息复杂、目标信息较小、背景信息与目标信息占比差距过大等问题, 研究者们常见的做法有改进损失函数、在检测模型的横向连接处引入各类注意力机制、优化检测锚框等, 而很少尝试将注意力模块与各模块融合使用, 也对航拍图像数据本身使用合理的数据增强方法缺乏关注。YOLOv5 对自然图像检测有较好表现, 但航拍图像相较于自然图像有差别较大, 将原始 YOLOv5 模型直接用于航拍图像检测会产生漏检误检问题, 因此对 YOLOv5 原始模型做了三方面的改进: 1) 为提高模型对图像中感兴趣区域的敏感程度以及提升主干网络中局部网络特征捕获与融合能力, 将 CBAM 通道注意力部署到原始 YOLOv5 的特征提取和特征融合网络的 C3 模块中。2) 为提高模型对小目标的检测效果, 在检测头部分增加针对小目标的预测层。3) 为了调整训练样本中各个目标类别不均衡, 可供模型学习的目标类别语义信息较少的问题, 在 Mosaic 数据增强的基础上将图片的数量由 4 增加到 9, 这样就可以加快模型收敛, 提高训练效率。

### 3.2. CBAM 注意力模块

CBAM 是一种轻量的注意力模块, 它是混合注意力机制(通道注意力机制与空间注意力机制)中的一种。它能够直接应用在网络结构当中。经过 CBAM 后, 新的特征图将得到通道和空间维度上的注意力权重, 通道注意力与空间注意力的顺序是自由组合的, 但是实验表明, 将通道注意力放在前面可以取得更好的效果。图 1 是 CBAM 注意力模块总体结构。通过引入注意力机制, 可以针对地解决航拍数据集中复杂背景, 使更多的注意力聚焦在有用的信息上, 降低对无关信息的关注度, 从而提高处理的效率和准确性。

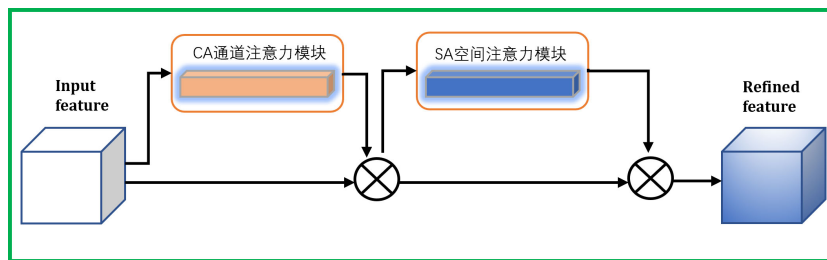


Figure 1. CBAM attention module

图 1. CBAM 注意力模块

在主干网络 C3 模块中加入 CBAM 注意力模块如图 2 所示, 可以增强网络对图像特征的提取, 极大地减少参数和计算量。

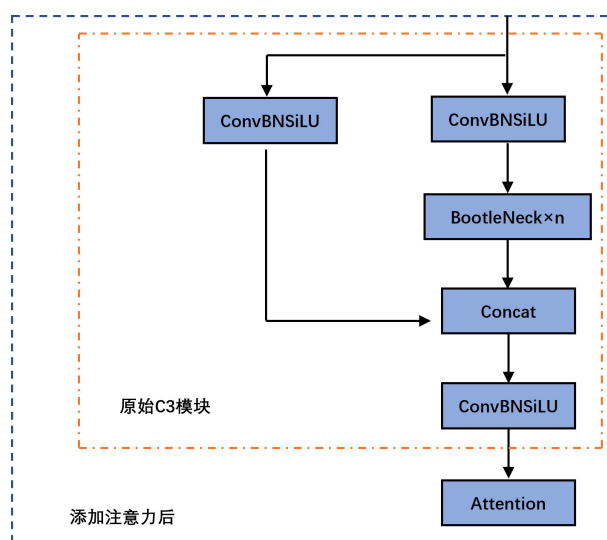


Figure 2. Adding modules

图 2. 添加模块方式

### 3.3. 小目标预测层

航拍图像目标检测难度较大的一个原因是由于拍摄的角度与高度问题导致图像中存在大量的小目标, 由于原始 YOLOv5 模型具有多层下采样, 随着下采样的深入, 在深层次的特征图上不容易得到小目标的特征信息, 所以在网络中增加针对小目标的预测层来解决该问题, 图 3 给出了在 YOLOv5 结构基础上添加针对小目标的预测层后的网络结构图。具体来说, YOLOv5 模型结构在特征融合模块(FPN 模块)并没有充分利用特征提取模块(Backbone 模块)中的浅层特征图, 而小目标的特征信息大多是在浅层当中。所以, 在 FPN 模块经过上采样操作得到  $80 \times 80$  尺度的特征图后, 再进行一次上采样操作, 将生成的  $160 \times 160$  特征图与 Backbone 模块中浅层的  $160 \times 160$  尺度的特征图进行特征融合, 并将得到的特征图直接输入到预测模块, 得到一个针对小目标的预测层, 具体改进如图 3 框中所示。

### 3.4. 改进 Mosaic 数据增强

Mosaic 数据增强的方法就是将数据集的图片进行随机裁剪, 再拼接到其他图片上进行训练数据。所以 Mosaic 数据增强使训练成本可以大大降低。为了更好的检测效果, YOLOv5 引入了 Mosaic 数据增强。

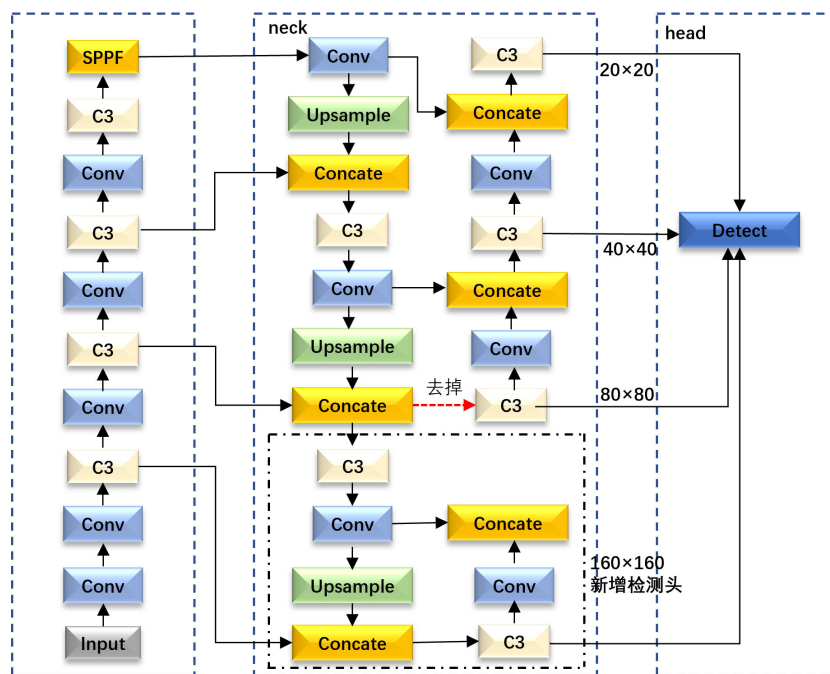


Figure 3. Adding a detection header  
图 3. 添加检测头

由于原始 YOLOv5 的 Mosaic 数据增强是随机裁剪图片，图像中的小目标以及模糊密集的目标被裁剪的概率很高，从而使输入模型的样本只有背景；此外，由于航拍图像尺寸大小也不同，随机裁剪后，会造成拼接后的图像大部分是灰色背景，这样会导致模型的检测速度和精度都会降低。因此，本文对 Mosaic 数据增强方式进行改进，将 4 张图片拼接改为 9 张图片拼接，然后根据拼接图片的矩形面积对灰色背景进行裁剪，改进后的 Mosaic 数据增强方式增加了检测目标不被裁剪的概率，从而提高了模型的检测性能。

## 4. 模型训练

### 4.1. 实验平台

所有实验都基于深度学习框架 Pytorch1.10.0，采用 Windows10 操作系统，使用单个 NVIDIA GeForce GPU 3060 显卡进行模型训练，表 1 展示了具体的信息。

Table 1. Experimental environment configuration  
表 1. 实验环境配置

项目	环境
CPU	11th Gen Intel(R) Core(TM)i7-11700@2.5GHz
GPU	NVIDIA GeForce GPU 3060
内存	12G
Python 版本	3.8
CUDA	11.3
cuDNN	8.2

## 4.2. 训练参数设置

将 VisDrone2019 数据集中的数据随机划分成训练集(80%)、验证集(10%)和测试集(10%)，并进行原始 YOLOv5 算法和改进 YOLOv5 算法对比实验。具体参数如表 2 所示：

**Table 2.** Training parameters  
**表 2.** 训练参数

参数名称	参数大小
权重衰减系数	0.0005
初始学习率	0.01
batch size	16
epoch	300
图片像素大小	640 × 640

## 5. 实验结果与分析

### 5.1. 评价指标

本文实验用来评估模型检测性能的指标分别是精确率 *Precision*、召回率 *Recall* 和平均精确率 *mAP*：

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

$$mAP = \frac{\sum_{i=1}^c AP_i}{C} \quad (3)$$

公式中：*TP* 表示正确检测目标的数量；*FN* 表示错误检测目标的数量；*FP* 表示漏掉的检测目标的数量；*C* 表示一共有多少个检测类别；*AP<sub>i</sub>* 表示第几个类别的精确率。

### 5.2. 实验结果与分析

为了验证改进算法的效果，在 VisDrone2019 数据集上进行原始算法与改进算法的对比实验，除改进部分以外，每组实验使用的训练参数不变，不同的改进方法对模型检测性能的影响如表 3 所示，取所得到的检测 *mAP* 的平均值作为衡量标准。在 VisDrone2019 数据集上的 10 个类别当中，对车辆的检测效果最好，并且改进算法相对于原始算法在小目标，密集场景下取得了更好的效果，如图 4 及图 5。

**Table 3.** Experimental results (mAP)

**表 3.** 实验结果(mAP)

方法	CBAM	小目标检测层	Mosaic 数据增强	VisDrone2019 数据集
YOLOv5				0.730
YOLOv5_A	√			0.733
YOLOv5_B		√		0.740
YOLOv5_C			√	0.738

Continued

本文方法一	✓		✓	0.740
本文方法二		✓	✓	0.743
本文综合方法	✓	✓	✓	0.745

从表格中可以明显看到，每个单独的模块均有利于航拍图像目标检测任务精度的提升。在航拍数据集中小目标数量较多，由于传统的 YOLOv5 算法只有 3 个目标检测头，这三个检测头检测大中小目标。但是由于航拍图像数据集的小目标数量太多且密集。在逐层的卷积之后，在小目标上能提取到的有效特征很少，所以在单独添加小目标检测层后，模型提高了对小目标的特征提取能力，因此表现出来的检测效果比较好；在添加注意力机制之后，直接改善了 neck 网络部分的整体特征提取能力，并且使模型在检测任务中更注意有效区域，提高检测性能；添加改进的 Mosaic 数据增强，使用 9 张图片拼接，使得更多的特征保留下来以至于不被裁减，因此获取更好的检测效果。由此可见，通过增加这些模块来提高检测效果是合理的。为了得到更好的效果，将每两个模块综合使用，实验结果显示，两两组合的改进带来了良好的精度提升；最后在三个模块同时的作用下，可以将目标检测的精度提升至 0.745，较原始的 YOLOv5 算法模型获得了 1.5% 的精度性能提升，说明改进的方法是有效的。

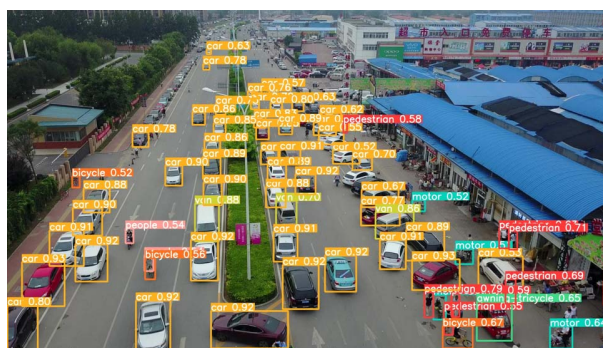


Figure 4. Original algorithm detection

图 4. 原始算法检测

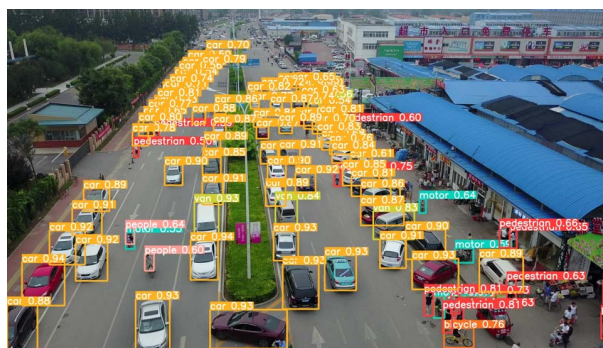


Figure 5. Improved algorithm detection

图 5. 改进算法检测

## 6. 结论

针对航拍图像检测中可以提取较少的特征、密集目标且复杂的背景导致检测精度低的问题，基于原

始 YOLOv5 目标检测算法改进得到新的航拍图像目标检测方法, 该方法通过引入 CBAM 注意力机制和增加小目标检测头来提高对航拍图像目标的检测效果, 同时使用改进 Mosaic 数据增强方法来提升检测精度与速度, 从原始算法与改进算法在 VisDrone2019 数据集上的实验结果来看, 改进的目标检测算法对检测精度有一定的提升。在后续工作中, 可以围绕着轻量化结构来优化网络结构, 实现在 APP 端的实时检测。

## 基金项目

BIGC20190122011, BIGC27170122003。

## 参考文献

- [1] 汪鹏, 辛雪静, 王利琴, 刘瑞. 基于 YOLOv3 的光学遥感图像目标检测算法[J]. 激光与光电子学进展, 2021, 58(20): 509-517.
- [2] 赵文清, 康怿瑾, 赵振兵, 翟永杰. 改进 YOLOv5s 的遥感图像目标检测[J]. 智能系统学报, 2023, 18(1): 86-95.
- [3] 李坤亚, 欧鸥, 刘广滨, 于泽峰, 李林. 改进 YOLOv5 的遥感图像目标检测算法[J/OL]. 计算机工程与应用: 1-9. <http://kns.cnki.net/kcms/detail/11.2127.TP.20221212.1832.022.html>, 2023-06-12.
- [4] 赵文清, 孔子旭, 周震东, 赵振兵. 增强小目标特征的航空遥感目标检测[J]. 中国图象图形学报, 2021, 26(3): 644-653.
- [5] Gong, H., Mu, T., Li, Q., *et al.* (2022) Swin-Transformer-Enabled Yolov5 with Attention Mechanism for Small Object Detection on Satellite Images. *Remote Sensing*, **14**, Article 2861. <https://doi.org/10.3390/rs14122861>
- [6] 周华平, 郭伟. 改进 YOLOv5 网络在遥感图像目标检测中的应用[J]. 遥感信息, 2022, 37(5): 23-30.
- [7] 李惠惠, 范军芳, 陈启丽. 改进 YOLOv5 的遥感图像目标检测[J]. 弹箭与制导学报, 2022(4): 17-23. <https://doi.org/10.15892/j.cnki.djzdx.2022.04.004>
- [8] Girshick, R.B., Donahue, J., Darrell, T. and Malik, J. (2013) Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. *Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, 23-28 June 2014, 580-587. <https://doi.org/10.1109/CVPR.2014.81>
- [9] Girshick, R.B. (2015) Fast R-CNN. *Proceedings of 2015 IEEE International Conference on Computer Vision (ICCV)*, Santiago, 7-13 December 2015, 1440-1448. <https://doi.org/10.1109/ICCV.2015.169>
- [10] Ren, S., He, K., Girshick, R.B. and Sun, J. (2015) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**, 1137-1149. <https://doi.org/10.1109/TPAMI.2016.2577031>
- [11] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S.E., Fu, C. and Berg, A.C. (2015) SSD: Single Shot MultiBox Detector. In: Leibe, B., Matas, J., Sebe, N. and Welling, M., Eds., *ECCV 2016: Computer Vision—ECCV 2016, Lecture Notes in Computer Science*, Vol. 9905. Springer, Cham, 21-37. [https://doi.org/10.1007/978-3-319-46448-0\\_2](https://doi.org/10.1007/978-3-319-46448-0_2)
- [12] Redmon, J., Divvala, S.K., Girshick, R.B. and Farhadi, A. (2015) You Only Look Once: Unified, Real-Time Object Detection. *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, 27-30 June 2016, 779-788. <https://doi.org/10.1109/CVPR.2016.91>
- [13] Redmon, J. and Farhadi, A. (2016) YOLO9000: Better, Faster, Stronger. *Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, 21-26 July 2017, 6517-6525. <https://doi.org/10.1109/CVPR.2017.690>
- [14] Redmon, J. and Farhadi, A. (2018) YOLOv3: An Incremental Improvement. <https://arxiv.org/abs/1804.02767>
- [15] Bochkovskiy, A., Wang, C. and Liao, H.M. (2020) YOLOv4: Optimal Speed and Accuracy of Object Detection. <https://arxiv.org/abs/2004.10934>
- [16] Zhu, X., Lyu, S., Wang, X. and Zhao, Q. (2021) TPH-YOLOv5: Improved YOLOv5 Based on Transformer Prediction Head for Object Detection on Drone-Captured Scenarios. *Proceedings of 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, Montreal, 11-17 October 2021, 2778-2788. <https://doi.org/10.1109/ICCVW54120.2021.00312>