

doi: 10.7690/bgzd.2022.12.017

## 基于 YOLOv5 改进的小目标检测算法

刘思诚<sup>1</sup>, 李嘉琛<sup>2</sup>, 邓皓<sup>3</sup>, 王栋栋<sup>3</sup>, 刘娟秀<sup>2</sup>, 张静<sup>2</sup>

(1. 南开大学电子信息与光学工程学院, 天津 300071; 2. 电子科技大学光电科学与工程学院, 成都 610054;  
3. 微智科技-电子科技大学智能显微技术联合研究中心, 四川 宜宾 644000)

**摘要:** 针对传统目标检测算法存在对小目标检测的识别精度低和不稳定的问题, 提出基于 YOLOv5 改进的小目标检测算法。基于卷积神经网络加入额外的检测头, 采用数据增强策略并更改网络卷积步长, 解决了小目标像素低、占比小、易重叠和难以分辨等问题; 同时依托真实检测场景制作一个全新的针对飞机检测的卫星影像数据集, 该数据集的待检测小目标占比达 61%, 飞机姿态及场景丰富, 有助于客观全面地验证网络精度。将改进后的算法与原始的 YOLOv5 模型进行对比, 结果表明, 其平均精确率 AP 值较原始 YOLOv5 模型提升约 3%。

**关键词:** 深度学习; 小目标检测; 检测头; 数据集; 航拍图片

**中图分类号:** TP274 **文献标志码:** A

## Improved Small Target Detection Algorithm Based on YOLOv5

Liu Sicheng<sup>1</sup>, Li Jiachen<sup>2</sup>, Deng Hao<sup>3</sup>, Wang Dongdong<sup>3</sup>, Liu Juanxiu<sup>2</sup>, Zhang Jing<sup>2</sup>

(1. *College of Electronic Information and Optical Engineering, Nankai University, Tianjin 300071, China;*  
2. *School of Optoelectronic Science and Engineering, University of Electronic Science and Technology of China, Chengdu 610054, China;* 3. *MIT-UESTC Joint Institute of Intelligent Microtechnique, Yibin 644000, China*)

**Abstract:** Aiming at the problem of low recognition accuracy and instability of small target detection in traditional target detection algorithm, an improved small target detection algorithm based on YOLOv5 is proposed. Based on the convolutional neural network, an additional detector is added, the data enhancement strategy is adopted and the network convolution step is changed to solve the problems of low pixel, small proportion, easy overlap and difficult resolution of small targets. At the same time, relying on the real detection scene, a new satellite image data set for aircraft detection is produced, in which the proportion of small targets to be detected is 61%, and the aircraft attitude and scene are rich, which is helpful to verify the network accuracy objectively and comprehensively. Comparing the improved algorithm with the original YOLOv5 model, the results show that the average accuracy AP value of the improved algorithm is about 3% higher than that of the original YOLOv5 model.

**Keywords:** deep learning; small target detection; detector; dataset; aerial image

### 0 引言

在军事领域中, 装备设施现代化、信息化是一个非常的重要的趋势。随着信息化社会的发展, 信息作为战略资源的地位将更高。在未来, 围绕信息资源获取、信息化军队建设和占领信息优势高地的竞争将愈演愈烈。目标检测与识别在军事上对于战场监视和侦察具有重要作用, 是现代高科技战争中赢得战争胜利的关键因素之一。目前, 目标检测算法已经应用到多个领域之中, 例如: 对于卫星图像等信息的识别运算, 对特定的目标进行追踪检测, 可以有效帮助军队掌握、识别并监测特定目标的行踪, 如飞机、坦克等军事敏感目标。

目前, 目标检测算法发展十分迅速, 但是对于小目标的检测, 仍旧存在诸多问题, 是学术界的关

注热点。小目标检测之所以受到广泛关注, 是因为在现实场景中, 小目标大量存在, 尤其是在卫星影像、无人机航拍图像中。相对于传统尺寸的目标, 小目标检测对象往往像素低、占比小、容易重叠、难以分辨; 同时, 还更容易受环境的影响, 例如光照剧烈变化等, 往往缺乏充足的信息将其分辨。传统的目标检测算法往往在应对此类对象的识别上能力很弱, 近年, 已经有许多研究开始采取不同的方法用于改进此类问题, 如使用数据增强、将背景合并到检测对象联合检测等方法。其中, 郑晨斌等<sup>[1]</sup>提出的利用双空洞卷积结构来强化浅层上下文信息的模型网络, 在有效节省参数量的同时较好地保留了原始网络结构, 但存在重要上下文信息丢失的问题, 同时也没有考虑到上下文信息可能缺乏的问题。针对小目标分辨率低问题, Haris 等<sup>[2]</sup>提出一

收稿日期: 2022-08-24; 修回日期: 2022-09-28

作者简介: 刘思诚(2001—), 男, 四川人, 从事智能光电检测研究。E-mail: Leonschat@qq.com。

种生成对抗学习方法，该方法利用端到端的联合训练检测方法，提升了低分辨率目标的检测性能。这种方法对于训练数据集要求较高，且对小目标检测性能的提升有限。

同时，在目标检测领域，现有数据集大多针对于大、中尺度目标，小目标数据集样本数量较小，相对比较缺乏。而已有的大规模公开数据集如 MS COCO<sup>[3]</sup>等，尽管小目标已经足够多，高达 40%，但每幅图实例分布不均，有些存在显著差距，训练出的模型会朝着高分布样本偏差。同时，大多数数据集往往针对于特定的检测目标，如 Penn-Fudan 行人检测与分割数据集<sup>[4]</sup>针对行人、EuroCity Persons 数据集<sup>[5]</sup>针对城市交通、Deepscores 数据集<sup>[6]</sup>针对乐谱乐符等。

针对上述问题，笔者在 YOLOv5x 的基础上，提出额外增加小目标的检测头多尺度检测的方法，有效改善小目标检测目标的问题，并制作关于飞机的小目标数据集对算法进行验证。

### 1 算法设计与相关研究

算法流程如图 1 所示。首先设置超参数和模型基本参数，用于后续的训练工作，随后设置冻结网络层并对读取的数据集进行马赛克数据增强；利用 for 循环进行迭代训练，每一次包含向前计算、反向传播、权重更新等操作；每完成一个 epoch，在验证集上进行一次效果评估，选出训练过程中效果较好的模型权重进行保存；最后，在训练结束后，将最优模型权重加载至实例化后的网络上，并在测试集上评估模型的表现性能。

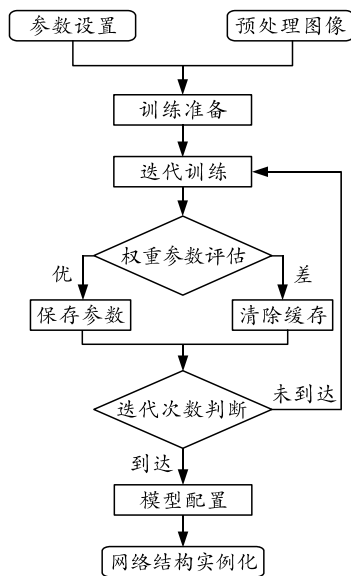


图 1 算法流程

### 1.1 YOLOv5

YOLOv5 是以 CNN 为主干的单阶段目标检测器，快速强大。其网络结构由 3 部分构成：输出端 (head)、主干 (backbone) 以及在之间的插入层 (neck)。与 YOLO 系列中之前的网络不同，YOLOv5 具有 4 种网络模型：YOLOv5s、YOLOv5m、YOLOv5l、YOLOv5x，YOLOv5x 是该系列中网络深度以及宽度最大的模型，具有精度高、检测慢的特点，其他网络较 YOLOv5x 整体架构没有区别。

笔者采用 YOLOv5x 作为基线模型进行改进，其中网络结构由 CSP-Darknet53 (backbone)、SPPF 和 CSP-PAN (neck) 以及 YOLOv3 Head (head) 组成。同时，还调节了一些参数用以改善模型性能。

### 1.2 改进的目标检测算法

网络结构如图 2 所示，其网络结构相比于 YOLOv5x，新增了一个针对于小目标的检测头，用于提升小目标检测的精度。该检测头基于 CNN 构成。结合原有的检测头，能够更加有效地对小目标进行识别。

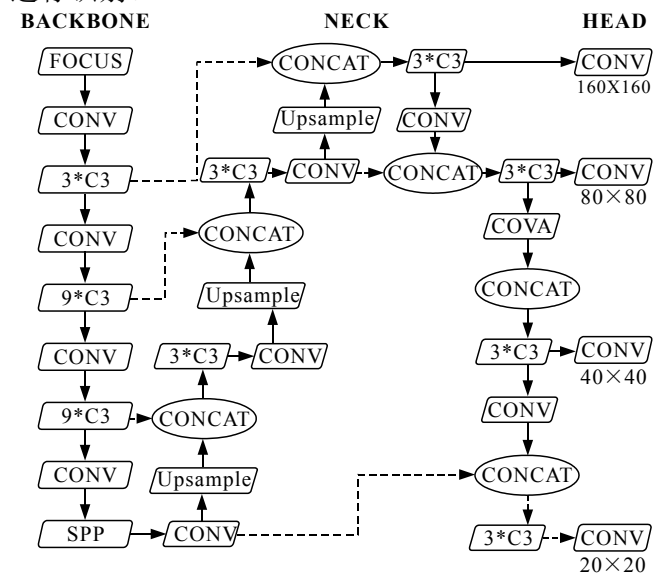


图 2 网络结构

检测头主要用于承担主干网络无法完成的定位任务，通过主干网络中提取到的特征图来检测目标物的位置和类别。不同的检测头有着不同的效果和特征，YOLOv5 本身网络结构层次较深，通过深层次的网络结构后获取到的特征图通常维度较低，原始的 YOLOv5 所带有的 3 个检测头分别在 20×20、40×40、80×80 大小的特征图上进行目标检测。而小目标由于其相对尺度较小，假设一个目标物的像素尺寸为 8 像素，输入图像的尺寸为 640 像素，那么

即使在尺度最大的特征图中,目标物所占据的尺度也仅有 1 像素;而如果目标物尺度小于 8 像素,则整个目标都会在特征图中消失。因此,现有的 YOLOv5 检测头无法满足小目标检测的要求。笔者在模型方面采用增加小目标检测层的方法,虽然一定程度上带来了更大的计算量,但显著改善了小目标识别精度。具体实施办法:保留原 YOLOv5s 的检测头结构,扩增一层卷积层,得到了  $160 \times 160$  大小的特征图,这一大尺度特征图负责检测小目标。并且由于原始训练集中样本特征多样性以及样本类别中独立同分布特征出现频率不同,需要利用低层高维特征分布空间保留的信息与深层低维特征空间的信息进行特征融合。如图 2 所示,第 18 层的特征图通过上采样操作使其特征图尺寸等于第 3 层,并将其特征空间在通道维度进行特征融合。

## 2 实验过程

### 2.1 数据集制作

小目标的定义在不同情景中各不相同,没有统一的标准,通常基于目标在图像中的绝对像素面积或相对像素面积进行定义。MS COCO 数据集采用绝对面积方式,将小于  $32 \times 32$  像素的目标定义为小目标,因为人类对小于该分辨率的图像目标存在辨识困难。Chen 等<sup>[7]</sup>采用相对面积方式,定义小目标的边界框面积与图像总面积之比小于 0.58%。如在  $640 \times 40$  图像中,像素尺寸小于  $42 \times 42$  的目标即可被认为是小目标。笔者将现代信息化战争中非常重要的航空飞机作为自动识别目标,综合 2 种小目标定义方式,收集并制作了名为 dataset-airplaneV1.0 的数据集,其中标注的小目标能够同时满足绝对面积和相对面积的要求。

目前,主流的航空数据集为 DOTA 数据集<sup>[8]</sup>,关于飞机的目标只有 1 万余个;同时,受限于航空拍摄角度的影响,飞机姿态比较单一。本数据集数据,通过爬虫、视频抽帧等手段获得图片数据后进行手工标注,总共 3 462 张图片,82 926 个目标框。

相对于现有的开源数据集,本数据集具有以下特点:1) 相对尺度、绝对尺度充分满足小目标的要求,同时对于飞机的尺度变化非常剧烈,小目标占总目标框占比大,共计 50 584 个,占总目标框数量的 60% 以上,相比于 COCO 数据集提高了约 20%。2) 场景丰富,背景环境多,数据分布多样化;例如世界各地的机场(美国、中国、英国等;森林、高原、平原、城市等;既有密集分布,也有松散分布)。3)

目标物姿态丰富,不局限于俯视的影像;同时,具有放缩、位移等特征。针对难以分辨的目标物,采用以下方法标注:对受到遮挡的目标,在露出一半以上并带有显著性飞机特征(如机翼)时标注;对模糊的飞机,与近焦图片对比,不予标注无法通过肉眼辨认出显著差距的目标物。

本数据集图例如图 3 所示。



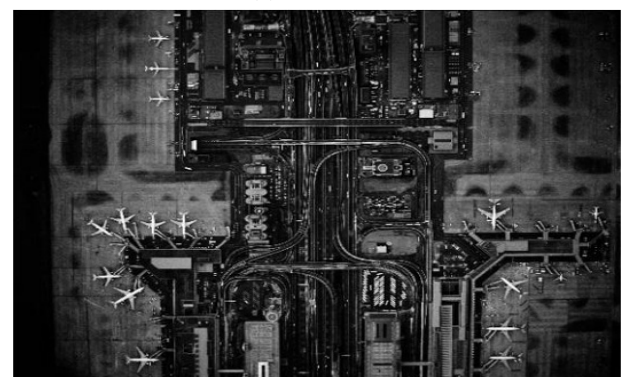
(a) 稀疏分布



(b) 密集分布



(c) 45°侧拍



(d) 俯视



(e) 复杂背景



(f) 重叠遮挡

图 3 数据集图例

目前数据集命名为 dataset-airplaneV1.0, 训练集包含图片 2 353 张, 目标框 56 689 个; 验证集包含图片 809 张, 目标框 19 213 个, 测试集包含图片 300 张, 目标框 7 024 个。

## 2.2 数据增强

数据增强手段是优化小目标检测最简单有效的方法, 其作用在于拓展数据集, 数据集常因高成本的标注训练图片和良好的神经网络模型矛盾, 采用数据增强使模型在训练时能够从不同环境获取的图像具备更高的鲁棒性, 降低对某些特定属性的依赖, 从而具有更好的泛化性。此前, 除了传统的裁剪、调色、调亮等方法之外, 还有一些利用多个图像进行数据增强的方法, 能够得到非常好的结果。笔者采用 Mosaic 方法<sup>[9]</sup>进行数据增强, Mosaic 是物体检测中常用的数据增强方法, 是 CutMix<sup>[10]</sup>的升级版, 其核心是将 4 张图片随机裁剪缩放后, 再通过随机组合拼接成一张图片, 不仅丰富了数据集, 也增加了小样本目标。同时, 在归一化计算时批量计算 4 张图片数据, 可降低训练过程中对内存的需求。

增强图例如图 4 所示, 数据增强的图来源为图 3 中的(a)、(b)、(d)、(f)。



图 4 数据增强图例

## 2.3 实验设置及评估指标

在云端使用 JupyterLab 进行运算, 实验环境为 NVIDIA RTX 2080 TI GPU(显存 12 GB)进行训练、验证。环境框架: Pytorch 1.10, CUDA 11.2。在训练阶段, 使用预训练网络 YOLOv5x, 迭代次数 100 次, 并只保留迭代次数最大的结果和在验证集上的最优结果。由于 GPU 显存较小且训练图片分辨率较高, batch size 综合考虑设置为 4; 同时, 为提高训练速度, 采用 cache 方式。

测试平台采用 Windows 10 操作系统, CPU 为 Intel(R)Core(TM)i7-8550UCPU@1.80 GHz 1.99 GHz, 运行内存 16 GB 的笔记本电脑。

目前, 主流的目标检测评价指标为平均精确率 (average precision, AP), 即为精确度-召回曲线 (precision-recall curve, PRC) 下面的面积, 如式(1)所示。

$$AP = \sum_{i=1}^{n-1} (r_{i+1} - r_i) P_{\text{inter}}(r_{i+1}). \quad (1)$$

AP@0.5 即为 detector IoU<sup>[11]</sup>阈值设置为 0.5 的 AP 值, AP@0.5:0.95 即为 IoU 阈值以 0.05 为间隔, 从 0.5 到 0.95 的平均 AP 值。AP 值能够很好地综合考虑 Precision 和 Recall 值, 反映模型的性能, 其中  $r$  为各 Precision 插值段对应的 Recall 值。P 函数表示所有大于指定 Recall 值的所有 Recall 值所对应的 Precision 最大值。

同时, 为充分说明本算法对于小目标检测的作用, 设计一个检测类算法, 将符合小目标的目标框挑选出来, 对此类目标框进行评估计算, 将小目标的 mAP 值设定为 APs。同时, 在比较改进原算法时, 使用了对抗训练网络 RobustDET<sup>[12]</sup>、YOLO 衍生算法 YOLOF<sup>[13]</sup>、对密集目标有显著提升的算法 GFocalV2<sup>[14]</sup>, 以及 VariFocalNET<sup>[15]</sup>, 其中后 2 种

算法基于 MMDetection<sup>[16]</sup>构成。以上 4 种算法，均使用默认的预训练模型进行训练；其中，根据代码需求的配置环境不同，对实验环境进行重新适配 (Pytorch 1.5.0, CUDA 10.1)，训练采用默认的训练轮次，训练结果以收敛结果为最终结果，即未收敛的模型还会继续进行训练。最终的各算法训练结果及评估如表 1 所示。

表 1 检测结果 %

检测网络	AP@0.5	AP@0.5:0.95	Aps
YOLOv5x	91.73	61.74	52.13
YOLOv5x+CN	94.25	64.30	55.40
Nhead(本文中)			
YOLOF	65.60	38.89	18.61
RobustDET	20.00	7.40	1.40
GFocalV2	89.50	59.30	47.60
VariFocalNet	86.00	54.60	41.80

相比于原始的 YOLOv5x，笔者所提方法 mAP 提升了约 3%，mAP@0.5 提升了约 3%，APs 提升了约 3.5%，模型大小由 165 M 增加至 175 M，参数数量由 86 173 414 增加至 90 844 152 (由 444 增加至

564 层)，检测速度方面每张图增加了 0.1 ms；相比于其他算法，笔者所提方法在测试数据集上表现效果最佳。可以得出，笔者提出的增加额外检测头的检测方法对于小目标物体检测具有显著的提升效果。将测试结果可视化后可以发现，对于相同的目标物，采用笔者所提方法检测得到的置信度高于原始的 YOLOv5，漏检、错检的情况也得到改善。

部分可视化结果如图 5 所示。图 5(a)可以看出，较原算法，额外的检测头增强了对小目标的特征分析能力，能更加准确地识别出小型飞机和建筑物、车辆等干扰目标的区别；通过图 5(b)可以看出，添加的检测头改善了特征图中小目标由于高深度网络易消失、被忽略的现象；图 5(c)的结果说明了对于重叠、被遮挡的目标，额外的检测头增强了对大规模聚集目标的分辨能力。图 5(d)表明本文中算法能对不同姿态的目标依旧具有高置信度的识别能力。

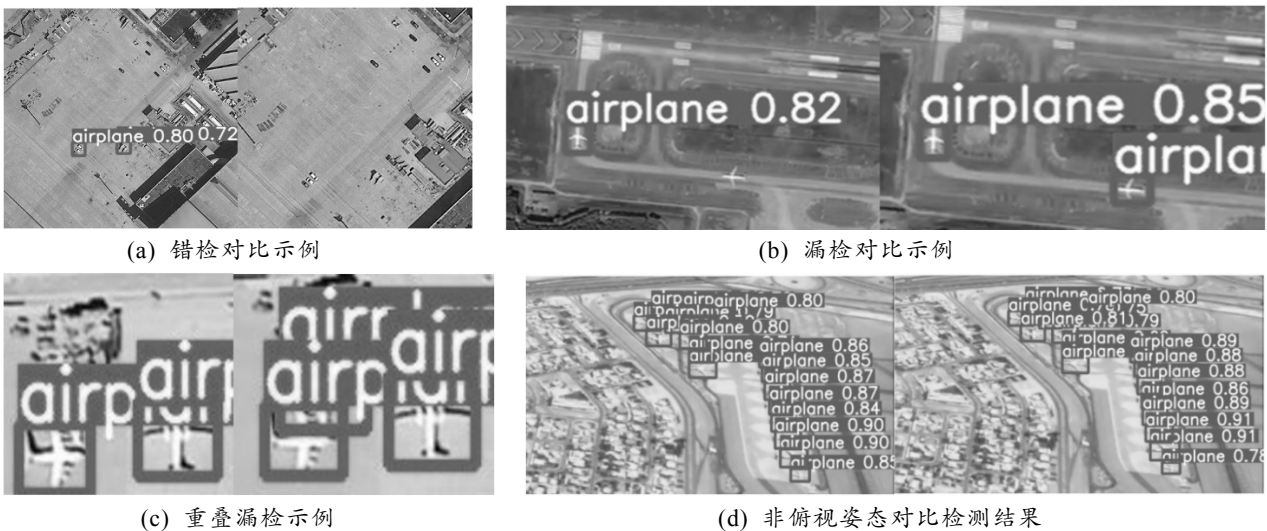


图 5 检测图例

### 3 结论

笔者提出并验证了一种基于 YOLOv5 改进的目标识别方法，通过数据增强、增加小尺寸目标检测头等手段提升了 YOLOv5 针对于小目标的检测效果。

1) 实验结果表明，该方法有效提高了小目标检测性能，同时对于多尺度目标的准确率较高，在复杂环境中可较好地完成多目标检测任务，所提模型的 AP 值相比原始的 YOLOv5x 模型提升约 3%。

2) 笔者基于真实目标检测场景制作了一个全新的飞机检测的航拍影像数据集，其特点为小目标占比高 (占比约为 60%)、飞机数量多 (约 83 000 个)、

飞机姿态及场景丰富，该数据集有助于客观全面地进行模型评价。

### 参考文献：

[1] 郑晨斌, 张勇, 胡杭, 等. 目标检测强化上下文模型[J]. 浙江大学学报(工学版), 2020, 54(3): 529-539.

[2] HARIS M, SHAKHNAROVICH G, UKITA N. Task-Driven Super Resolution: Object Detection in Low-resolution Images[C]// International conference on neural information processing. ArXiv, 2018.

[3] LINTY. Microsoft COCO: Common Objects in Context[C]// European conference on computer vision. ArXiv, 2014.