

基于改进YOLOv5用于无人机图像的实时车辆检测

何佳桥, 葛晓东

上海理工大学光电信息与计算机工程学院, 上海

收稿日期: 2023年9月5日; 录用日期: 2023年10月11日; 发布日期: 2023年10月23日

摘要

如今, 无人机(UAV)逐渐被用于各个领域, 例如交通监控和智能停车, 其中对车辆的实时监测和分类是关键任务之一。在车辆检测方面存在许多挑战, 例如小型车辆目标和无人机工作时飞行角度的变化导致目标尺度变化从而给车辆检测网络模型的优化带来的负担。此外, 由于高空飞行时航拍图像目标较小、可提取特征较少, 而导致模型检测精度较低。为解决上述问题, 本文旨在基于YOLOv5算法, 提出了一种准确、高效、实时的车辆检测网络。首先, 为了使模型更好地提取小目标特征, 在Neck部分添加了一个新的连接层, 将第一个C3层的高分辨率特征映射连接到Neck部分。其次, 为使模型更加专注小目标, 添加一个步幅为4的输出层作为新的头部。同时优化模型对较大车辆目标的检测较少, 我们去掉了Head中输出特征图为 20×20 的检测头。同时, 考虑到模型的推理速度, 将Neck部分的C3模块替换为更轻量化的DS_C3模块。最后, 为了进一步提高基于IOU损失函数的性能, 将CIoU替换为 α -IOU。本文使用VisDrone2019数据集, 并基于改进算法和原始算法分别进行了实验, 结果表明, 本文的算法能够对小目标进行有效的实时检测。

关键词

车辆检测, YOLOv5, DSConv, 自适应损失函数

Real-Time Vehicle Detection Based on Improved YOLOv5 for Drone Images

Jiaqiao He, Xiaodong Ge

School of Optoelectronic Information and Computer Engineering, University of Shanghai for Science and Technology, Shanghai

Received: Sep. 5th, 2023; accepted: Oct. 11th, 2023; published: Oct. 23rd, 2023

文章引用: 何佳桥, 葛晓东. 基于改进 YOLOv5 用于无人机图像的实时车辆检测[J]. 软件工程与应用, 2023, 12(5): 705-713. DOI: 10.12677/sea.2023.125068

Abstract

Today, unmanned aerial vehicles (UAVs) are gradually being used in various fields, such as traffic monitoring and smart parking, where real-time monitoring and classification of vehicles is one of the key tasks. There are many challenges in vehicle detection, such as changes in flight angle when small vehicle targets and drones are working, resulting in changes in target scale, which puts a burden on the optimization of vehicle detection network models. In addition, due to the small aerial image target and fewer extractable features during high-altitude flight, the model detection accuracy is low. To solve the above problems, this paper aims to propose an accurate, efficient and real-time vehicle detection network based on the YOLOv5 algorithm. First, in order to make the model better extract small target features, a new connection layer is added to the Neck part, connecting the high-resolution feature mapping of the first C3 layer to the Neck section. Second, to make the model more focused on small targets, an output layer with a stride length of 4 is added as a new head. At the same time, the optimization model has less detection of larger vehicle targets, and we remove the detection head with an output feature map of 20×20 in Head. At the same time, considering the inference speed of the model, the C3 module of the Neck part was replaced with a more lightweight DS_C3 module. Finally, to further improve the performance of the IOU-based loss function, the CIOU is replaced with a α -IOU. This paper uses the VisDrone2019 dataset and conducts experiments based on the improved algorithm and the original algorithm, and the results show that the algorithm in this paper can effectively detect small targets in real time.

Keywords

Vehicle Detection, YOLOv5, DSConv, Adaptive Loss Function

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, 无人机(UAV)普遍被应用于各个领域, 例如农业中的无人机农药喷洒、快递行业的无人机送货和智慧交通中的智能停车。由于其灵活性, 无人机可以从不同的位置、高度和角度使用高分辨率相机收集数据。同时, 由于硬件计算能力和人工智能的发展也促进了无人机的实际应用。在交通领域, 迸发了许多创新研究, 本文重点关注智慧交通中的车辆检测, 它也是道路和城市交通检测的第一步[1]。

针对无人机航拍图像中的车辆检测目标, 由于相机角度和无人机高度、目标的大小急剧变化, 同时, 复杂的背景、光照强度以及目标不同外观等因素都会导致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。为此, 针对这些问题。研究者们对目标检测器不断的进行改进。例如, Zhu X 等[2]以 YOLOv5 为基础, 通过在 Neck 部分融入注意力机制并增加检测头, 使得该模型的精准度大大提升。Mahaur [3]通过改进 YOLOv5Neck 部分并融入空洞卷积, 在未显著增大模型的情况下, 提升网络的精准度。Wang T [4]等人针对多尺度目标的检测精度低、目标特征信息易丢失等问题。在 YOLOv4 的基础提出了上一种先验增强的 PETNet 网络。为了解决小目标过采样而导致特征丢失的问题, Li R. & Shen Y [5]在输入端对图像进行超分辨率处理, 并以 YOLOv5 模型为基础进行改进, 提出了 YOLOSr-IST 网络。Huang Y [6]等将 transformer 融入 YOLOv5Backbone 部分, 使模型充分提取小目标特

征。Cui, M [7]等人出基于改进的 YOLOv4-tiny, 该模型使用轻量级和增强的 ShuffleNet 取代默认的骨干网络, 并使用交叉熵损失函数, 使该模型相比较基模型更为准确。此外, Zhou J [8]等人在 YOLOv5 的基础上通过增加特征融合层, 并添加注意力机制, 来解决目标遮挡与复杂背景的问题。以上方法对小目标检测具有重要意义, 大多数方法都提高了准确度, 但存在模型较大导致推理速度慢的问题。再加上无人机上携带的计算资源有限, 因此提升模型检测精度和模型轻量化, 是本文的研究重点。

2. YOLOv5 网络模型介绍

目标检测方法主要分为单阶段两阶段方法, 其中经典的两阶段检测方法主要有 R-CNN 系列[9] [10] [11], 单阶段检测方法的代表为 SSD [12]以及 YOLO [13] [14] [15]系列。YOLO 系列检测方法一直保持更新换代。YOLOv5 (You Only Look Once)模型在 2020 年 5 月由 Ultralytics 提出, 模型在工业界被广泛使用。该模型是以 YOLOv4 为基础改进而来的一阶段检测器, YOLOv5 有 YOLOv5s、YOLOv5m、YOLOv5l、YOLOv5x 四个模型。其中, YOLOv5s 的权重文件仅为上一代模型的 1/9, 其结构也更加小巧, 其网络模型结构如图所示。为了平衡精度和速度, 本文选择 YOLOv5s 作为基础算法模型。

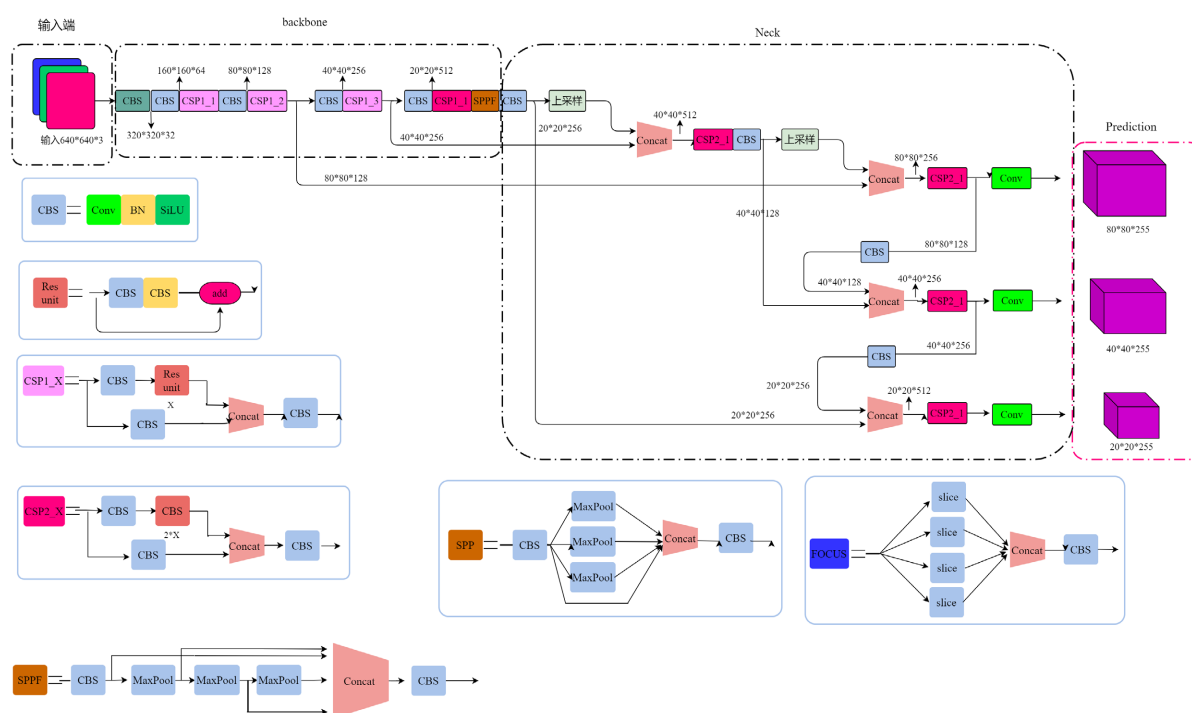


Figure 1. Structure diagram of the YOLOv5 network model

图 1. YOLOv5 网络模型结构图

由上图 1 可知, 该模型主要分为 Input、Backbone、Neck 和 Prediction 四个部分。

Input 端由 Mosaic 数据增强、图片尺寸处理、自适应锚框计算三个部分组成。与 YOLOv4 相同, YOLOv5 使用 Mosaic 进行数据增强, 该方法对小目标检测有显著效果, 符合本文对无人机航拍图像的这种小目标检测的需求。

Backbone 是网络的主干部分, 主要负责特征提取, 其中包括一系列的卷积层和 C3 层, 卷积层由卷积、batch normalization (BN)层和 Sigmoid Linear Unit (SiLU)激活函数三个部分组成。此外, C3 模块包括三个卷积层和一系列的瓶颈模块, 这些模块减少了模型的参数和计算量。在 Backbone 的末端, 为 SPPF

模块, 它融合了不同分辨率的特征, 该模块包含三个最大池化层和卷积层, 在这些模块的作用下, 网络能够提取更多特征信息。

Neck 处于 Backbone 与 Prediction 之间, 由 The Path Aggregation Network (PAN)组成, 它聚合了 Backbone 从不同阶段提取的特征。由于 Backbone 浅层部分, 特征图的空间特征很强, 但语义特征较弱。而 Backbone 深层部分则具有丰富的语义特征而缺乏空间特征。路径聚合网络(PAN)将不同阶段的特征图按两个方向(从上到下和从下到上)进行连接, 从而获得较强的空间信息和语义信息的特征图。

Prediction 单元预测三种不同尺度的特征图, 默认的输出尺寸为 20×20 、 40×40 、 80×80 。多尺度目标检测有助于检测出图像中的大、中、小目标。最后, 使用非极大值抑制(NMS)计算出每个目标最精确的边界框。

3. 改进 YOLOv5 以实现高效的实时航拍车辆检测

3.1. 改进思路

与自然图像相比, 无人机航拍图像的目标信息一般较小, 并且, 由于飞行过程中相机轻微抖动, 导致拍摄图像模糊使小目标特征不能被充分提取等问题。为了解决这些问题, 常用的方法是对模型进行改进, 例如改变损失函数、修改模型结构、增加注意力机制、优化检测锚框等。而很少将模块轻量化融入模型中, 因为往往在修改的过程中往往不能使精度和速度达到平衡。YOLOv5 在进行自然图像检测上有较好的效果, 而航拍图像大多是为小目标, 原始 YOLOv5 模型用于无人机航拍图像检测时, 模型提取到的小目标特征较少, 所以实际效果并不理想。为此, 对 YOLOv5 模型进行了以下改进:

1) 为了提高模型对小目标特征的提取与融合能力, 添加一个步长为 4 的检测头, 同时, 为了由于目标中很少有目标并且防止模型过大影响推理速度, 去除步长为 32 的输出层及在 Neck 部分的对应层。

2) 由于模型添加新的检测头, 模型变大, 为防止严重影响模型的推理速度, 将 C3 原始 BottleNeck 替换成为 DC_C3, 达到模型精度与速度的平衡。

3) 为了进一步提高基于 IOU 的损失函数的效率, 将原始模型的 C-IOU 替换为 α -IoU。

3.2. 小目标检测头

航拍图像由于包含较多小目标, 而 YOLOv5 原始模型包含下采样模型, 随着模型的深入, 小目标特征丢失越来越多, 即模型的 Neck 部分并未充分提取到网络浅层空间特征, 为了改善这个问题, 将第一个 C3 层的高分辨率特征映射连接到 Neck 部分, 即在 Neck 部分添加卷积、上采样、连接和 C3 层, 这有效的保留了空间特征。同时, 增加一个输出特征图为 160×160 的检测头, 并且为了使模型对小目标更为敏感, 去除输出特征图为 20×20 的检测头, 并删除其相关模块。这使模型更加专注小目标同时使模型结构的改变对模型的推理速度影响较小。其具体改进如图 2 所示。

3.3. DS_C3 模块

由于网络层数增加, 减缓了模型的推理速度, 为了使改进对模型的实时检测能力影响最小化, 将原始 YOLOv5 中的 BottleNeck 模块进行改造为 BottleNeck_DSConv, 即将其中的将普通卷积替换为 DSConv。DSConv 是 Marcelo Gennari [16]在 2019 年 ICCV 中所提出的, 该结构为卷积层的变体, 相比原始卷积层, 改变体牺牲了部分精度换来了速度的提升。DSConv 将传统的卷积内核分为两个组件: 可变了化内核 (VQK)和分布偏移, 通过在 VQK 中仅储存整体数值来实现较低的储存器使用并较高的运行速度, 同时通过应用基于内核和基于通道的分布偏移来保持与原始卷积的输出。DSConv 替换 YOLOv5 中的 BottleNeck 中的卷积层, 增强了模型的推理速度, 虽然精度有所下降, 但这是值得的。DSConv 结构如图 3 所示。

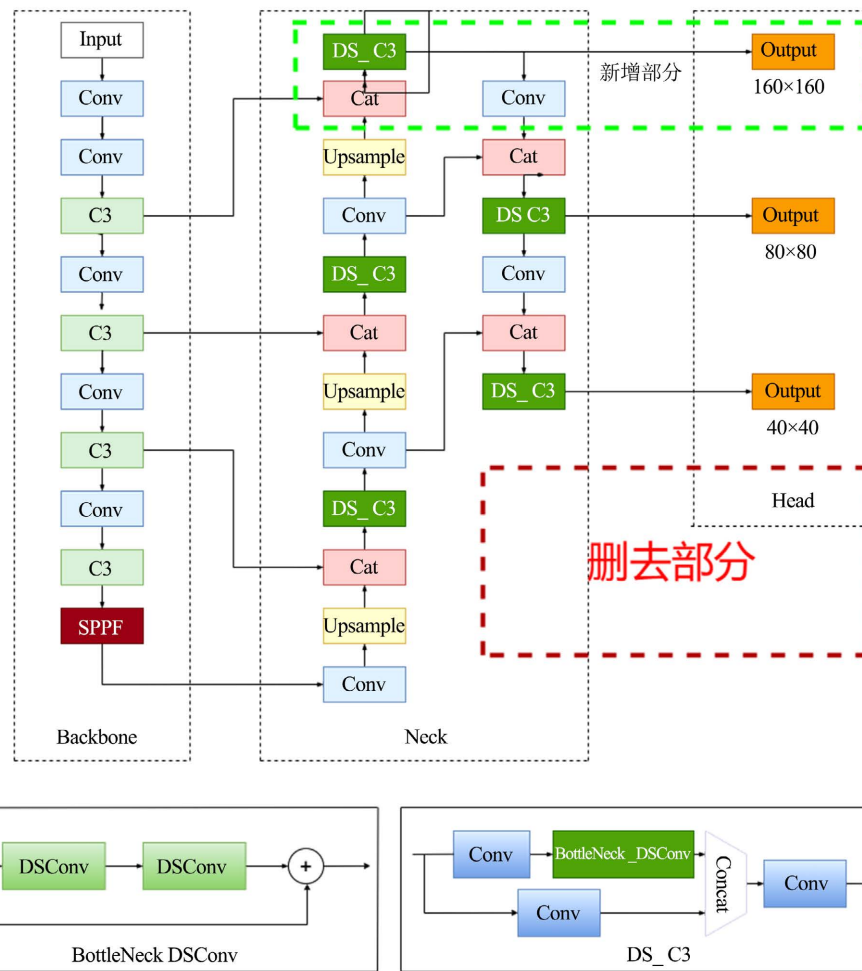


Figure 2. YOLOv5 improves the network model structure diagram
图 2. YOLOv5 改进网络模型结构图

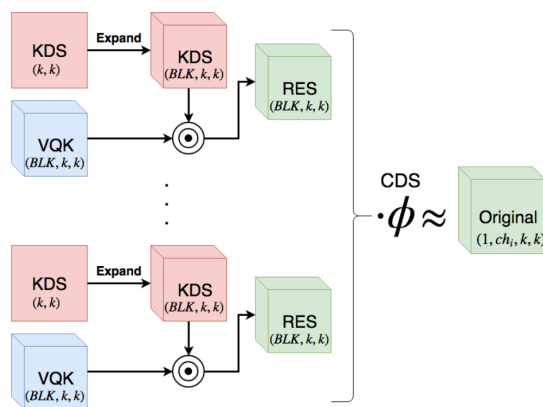


Figure 3. DSConv model structure diagram
图 3. DSConv 模型结构图

3.4. 替换损失函数

YOLOv5 的损失函数由三个不同部分组成, 边界框回归损失、置信度损失和分类损失, 置信度损失

和分类损失采用二元交叉熵损失函数, 而边界框回归损失函数默认为 C-IoU。C-IoU 的计算公式如下:

$$L_{C-IoU} = 1 - IoU + \frac{\rho^2(b, b_{gt})}{c^2} + \beta v \tag{1}$$

$$v = \frac{4}{\pi^2} \left(\arctan \frac{\omega^{gt}}{h^{gt}} - \arctan \frac{\omega}{h} \right)^2 \tag{2}$$

$$\beta = \frac{v}{(1 - IoU) + v} \tag{3}$$

其中, b 和 b_{gt} 分别为预测框和 Ground Truth 的中心坐标, ρ 表示两点的欧氏距离, c 为两框之间最小的边界框的对角线长度, β 为正参数, v 表示两框长宽比的一致性, ω 、 h 和 ω^{gt} 、 h^{gt} 分别为预测框的长宽和 Ground Truth 的长宽。虽然 C-IoU 的收敛速度和精度相比 IOU 有所提高, 但在小目标数据集上, α -IoU 效果优于 C-IoU, 并且在模型上具有更好的鲁棒性。所以, 为了提升模型精度, 将 C-IoU 进行替换, 其中, α 为权重系数。同时, 经实验分析, $\alpha = 2$ 时模型精度提升最为显著。 α -IoU 的计算公式如下:

$$L_{\alpha-IoU} = 1 - IoU^\alpha + \frac{\rho^{2\alpha}(b, b_{gt})}{c^{2\alpha}} + (\beta v)^\alpha \tag{4}$$

4. 模型训练

4.1. 数据集

使用的数据集为公开数据集 VisDrone2019, 该数据集包含 6471 张训练图像, 548 张验证图像和 1610 张测试图像。该数据集包含 10 个类。为了专注于车辆检测, 只使用其中 Car、Van、Bus 和 Truck 4 个类别对网络进行训练和评估, 数据集中没有删除任何图像。

4.2. 实验环境

本实验所使用的软硬件环境如表 1 所示, 训练参数如表 2 所示。

Table 1. Experiment environment configuration table

表 1. 实验环境配置表

参数	配置
CPU	Xeon(R) Platinum 8255C @2.5Hz
GPU	2080Ti
内存	11G
Pytorch	1.11.0
Ubuntu	20.04
CUDA	11.3

Table 2. Table of training parameters

表 2. 训练参数表

参数	参数大小
权重衰减系数	0.0005
初始学习率	0.01
batch size	16
epoch	300

5. 实验结果与分析

5.1. 评估指标

使用的评估指标为精确率(Precision)、召回率(Recall)、平均精确率(Mean Average Precision)。其中 TP 表示被模型正确识别为车辆的目标数量, FP 表示被模型错误识别为车辆的目标数量, FN 表示模型遗漏识别车辆的目标数量。因此, 精确率表示模型正确识别车辆占所有识别为车辆的所以目标的比例, 召回率表示模型正确识别为车辆的目标数量占数据集中所有车辆的比例, N 表示为类的数量。 AP_i 表示第 i 类在不同 IoU 阈值上的平均精度。我们使用 mAP50 和 mAP95 来评估模型的准确度。

$$P = \frac{TP}{TP + FP} \quad (5)$$

$$R = \frac{TP}{TP + FN} \quad (6)$$

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (7)$$

5.2. 实验结果与分析

为了验证改进算法的效果, 模型的超参数设置都相同, 不同改进对模型检测车辆能力的影响如表 3、表 4 所示。在 VisDrone2019 数据集的四个类别中, 相比于原始模型, 改进后的模型对车辆的检测能力和在密集场景下的效果显著提升。效果如图 4、图 5 所示。



Figure 4. Original model renderings

图 4. 原始模型效果图



Figure 5. Improved model renderings

图 5. 改进模型效果图

从表 3 中可以看出, 每个模型对模型检测性能均有提升。由于 VisDrone2019 数据集中小目标较多, 原始模型在进行逐层卷积下采样后, 小目标特征丢失较多, 将第一个 C3 层的高分辨率特征映射连接到 Neck 部分后并增加小目标检测头, 模型的小目标检测性能大大提升。改进后, 模型的 mAP50 和 Map95 分别提高了 6.24% 和 4.93%。这表明了新的高分辨率特征图拼接层和小目标检测头的重要性。 α -IoU 替换 C-IoU 后提升了模型的检测精度, 并且改进后对模型的推理速度影响较为微小。DS_C3 的替换, 虽然模型的检测精度有所下降, 但模型的推理速度有所提升, 所以改进是有效的。为了得到更好的效果, 将两两改进组合使用, 实验结果表明, 两两改进组合结合了各自的优点, 模型的检测精度均有明显的提升。最后, 将三种改进同时融入模型, 与原始模型相比, 新模型的精确率提高了 3.31%、召回率提高了 7.03%、mAP50 和 Map95 分别提高了 6.41% 和 5.32%, 并且, 修改后的模型相比于原始模型虽然 FPS 有所下降, 但与未加入 DS_C3 的模型相比, FPS 提升了 4.28, 同时, 新模型的 FPS 达到了 94.26, 具有实时检测能力。说明总体改进方法是有效的。

Table 3. Table of the effects of different modifications on model accuracy

表 3. 不同修改对模型精度的影响表

Model	精确率	召回率	mAP50	mAP95	FPS
YOLOv5s	68.81	45.33	51.12	34.22	120.85
+小目标检测头	70.02	52.62	57.36	39.15	91.20
+ α -IoU	67.71	47.92	52.92	35.80	120.08
+DS_C3	67.82	45.42	50.86	33.87	125.02
+ α -IoU, 小目标检测头	71.84	52.95	58.34	40.17	89.85
+ α -IoU+DS_C3	65.20	48.38	52.22	35.40	123.67
+DS_C3, 小目标检测头	70.94	51.50	56.71	38.67	96.23
Proposed model	72.12	52.36	57.53	39.54	94.26

Table 4. Each category is in the mAP50 table on two models

表 4. 每个类别在两个模型上的 mAP50 表

类别	Car	Van	Bus	Truck
YOLOv5s	76.82	41.40	34.86	51.44
改进后的模型	85.50	48.68	37.64	59.27

6. 结论

本文提出了一种基于 YOLOV5s 的改进模型, 用于无人机航拍图像中的实时车辆检测, 并且在 VisDrone2019 数据集中进行了评估。与原始模型相比, 所提出的方法进行了以下修改: 1) 添加一个新的检测头来检测小目标同时改进了网络的 Neck 部分以获得更好的特征图。2) 边界框回归损失函数替换为自适应损失函数来提高模型的检测性能。3) 将 Neck 部分的 C3 替换为 DS_C3, 提高模型的推理能力。总体来说, 修改的模型的检测性能相比原始模型具有显著提升。在后续的工作中, 继续着重对模型的轻量化, 实现在计算能力较差的设备上运行。

参考文献

- [1] Srivastava, S., Narayan, S. and Mittal, S. (2021) A Survey of Deep Learning Techniques for Vehicle Detection from UAV Images. *Journal of Systems Architecture*, 117, Article ID: 102152. <https://doi.org/10.1016/j.sysarc.2021.102152>

-
- [2] 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. 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>
- [3] Mahaur, B. and Mishra, K. (2023) Small-Object Detection Based on YOLOv5 in Autonomous Driving Systems. *Pattern Recognition Letters*, **168**, 115-122. <https://doi.org/10.1016/j.patrec.2023.03.009>
- [4] Wang, T., Ma, Z., Yang, T. and Zou, S. (2023) PETNet: A Yolo-Based Prior Enhanced Transformer Network for Aerial Image Detection. *Neurocomputing*, **547**, Article ID: 126384. <https://doi.org/10.1016/j.neucom.2023.126384>
- [5] Li, R. and Shen, Y. (2023) YOLOSr-IST: A Deep Learning Method for Small Target Detection in Infrared Remote Sensing Images Based on Super-Resolution and YOLO. *Signal Processing*, **208**, Article ID: 108962. <https://doi.org/10.1016/j.sigpro.2023.108962>
- [6] Huang, Y., He, J., Liu, G., Li, D., Hu, R., Hu, X. and Bian, D. (2023) YOLO-EP: A Detection Algorithm to Detect Eggs of *Pomacea canaliculata* in Rice Fields. *Ecological Informatics*, **77**, Article ID: 102211. <https://doi.org/10.1016/j.ecoinf.2023.102211>
- [7] Cui, M., Lou, Y., Ge, Y. and Wang, K. (2023) LES-YOLO: A Lightweight Pinec-Onedetection Algorithm Based on Improved YOLOv4-Tiny Network. *Computers and Electronics in Agriculture*, **205**, Article ID: 107613. <https://doi.org/10.1016/j.compag.2023.107613>
- [8] Zhou, J., Jiang, P., Zou, A., et al. (2021) Ship Target Detection Algorithm Based on Improved YOLOv5. *Journal of Marine Science and Engineering*, **9**, Article 908. <https://doi.org/10.3390/jmse9080908>
- [9] 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>
- [10] 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>
- [11] 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>
- [12] 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*, Springer, Cham, 21-37. https://doi.org/10.1007/978-3-319-46448-0_2
- [13] 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>
- [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] Do Nascimento, M., Fawcett, R. and Prisacariu, V. (2019) DSConv: Efficient Convolution Operator. 2019 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, 27 October-2 November 2019, 5147-5156. <https://doi.org/10.1109/ICCV.2019.00525>