



吉林大学学报(工学版)

Journal of Jilin University(Engineering and Technology Edition)

ISSN 1671-5497,CN 22-1341/T

《吉林大学学报(工学版)》网络首发论文

题目: 基于改进 YOLOX 的无人机航拍图像密集小目标车辆检测
作者: 张河山, 范梦伟, 谭鑫, 郑展骥, 寇立明, 徐进
DOI: 10.13229/j.cnki.jdxbgxb.20230779
收稿日期: 2023-07-25
网络首发日期: 2023-12-28
引用格式: 张河山, 范梦伟, 谭鑫, 郑展骥, 寇立明, 徐进. 基于改进 YOLOX 的无人机航拍图像密集小目标车辆检测[J/OL]. 吉林大学学报(工学版).
<https://doi.org/10.13229/j.cnki.jdxbgxb.20230779>



网络首发: 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式(包括网络呈现版式)排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

出版确认: 纸质期刊编辑部通过与《中国学术期刊(光盘版)》电子杂志社有限公司签约, 在《中国学术期刊(网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊(网络版)》是国家新闻出版广电总局批准的网络连续型出版物(ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

基于改进 YOLOX 的无人机航拍图像密集小目标车辆检测

张河山^{1,2}, 范梦伟¹, 谭鑫¹, 郑展骥^{1,2}, 寇立明³, 徐进^{1,2}

(1.重庆交通大学, 交通运输学院, 重庆 400074; 2.重庆交通大学 山区复杂道路环境“人-车-路”协同与安全重庆市重点实验室, 重庆 400074; 3.重庆市交通规划研究院, 重庆 400074)

摘要：目前基于无人机航拍图像的目标检测技术广泛应用于智能交通领域，但在无人机航拍视角下对小目标的检测仍存在漏检现象严重、检测精度低等问题。针对这一问题，提出一种改进的 YOLOX 网络用于无人机航拍图像的检测。为了增强网络的特征学习能力，在特征融合部分引入自适应空间特征融合(Adaptively Spatial Feature Fusion, ASFF)模块，并在网络的颈部(Neck)嵌入坐标注意力机制(Coordinate Attention, CA)。为了加强网络对正样本的学习，将二元交叉熵损失(BCE Loss)函数替换为变焦距损失(Varifocal Loss)函数。实验结果表明，改进后的 YOLOX 网络具有更好的检测效能，其 mAP@50 和 mAP@50_95 分别达到了 91.50%和 79.65%。在多种交通场景下的可视化结果表明，相较于其他算法，优化后的网络具有更低的漏检率以及更高的检测精度，能够胜任小目标车辆的检测任务，可为高空视角下的车辆多目标跟踪应用提供参考。

关键词：小目标车辆检测；损失函数；坐标注意力机制；自适应空间特征融合；YOLOX

中图分类号：U491

文献标志码：A

DOI: 10.13229/j.cnki.jdxbgxb.20230779

Dense small object vehicle detection in UAV aerial images using improved YOLOX

ZHANG He-shan^{1,2}, FAN Meng-wei¹, TAN Xin¹, ZHENG Zhan-ji^{1,2}, KOU Li-ming³, XU Jin^{1,2}

(1. School of Traffic & Transportation, Chongqing Jiao Tong University, Chongqing 400074, China; 2. Chongqing Key Laboratory of "Human-Vehicle-Road" Cooperation and Safety for Mountain Complex Environment, Chongqing Jiao Tong University, Chongqing 400074, China; 3. Chongqing Transportation Planning and Research Institute, Chongqing 400074, China)

Abstract: At present, object detection technology based on unmanned UAV aerial images is widely used in the field of intelligent transportation, but there are still problems, such as seriously missed detection and low detection accuracy in the detection of small targets from the perspective of UAV aerial photography. To solve this problem, an improved YOLOX network is proposed for the detection of UAV aerial images. To enhance the feature learning ability of the network, the Adaptive Spatial Feature Fusion (ASFF) module is introduced in the feature fusion part, and the Coordinate Attention (CA) mechanism is embedded in the neck of the network. To enhance the network's learning of positive samples, the binary cross-entropy loss (BCE Loss) function is replaced with the Varifocal Loss function. Experimental results show that the improved YOLOX network has better detection efficiency, and its mAP@50 reaches 91.50% and mAP@50_95 reached 79.65%. The visualization results in various traffic scenarios show that compared with other algorithms, the optimized network has a lower missed detection rate and higher detection accuracy, which can be competent for the detection task of small target vehicles, and can provide a reference for vehicle multi-target tracking applications from a high-altitude perspective.

Keywords: small target vehicle detection; loss function; coordinate attention mechanism; adaptive spatial feature fusion; YOLOX

收稿日期：2023-07-25.

基金项目：国家自然科学基金面上项目(52172340); 重庆市教育委员会青年项目(KJQN202200710); 重庆市博士后科学基金项目(CSTB2022NSCQ-BHX0731); 重庆交通大学研究生科研创新项目(CYS23498).

作者简介：张河山(1988-), 男, 讲师, 博士. 研究方向: 视频图像处理技术, 道路交通安全. E-mail: hszhang@cqjtu.edu.cn

通信作者：徐进(1977-), 男, 教授, 博士. 研究方向: 人-车-路系统协同, 道路安全性设计以及驾驶行为. E-mail: yhn_996699@163.com

0 引言

随着计算机视觉的快速发展,基于无人机航拍图像的目标检测技术在交通领域受到了研究者的广泛关注^[1]。无人机航拍具有视野广、拍摄范围大、灵活性高等优点^[2],经常被作为车辆检测任务的补充视角,应用在特殊路段和突发交通场景中,为后续开展交通事故分析、交通流管控等任务提供数据支持。但在无人机航拍视角下,具有较少特征的小目标车辆占比较大,并且排列密集,在检测网络通道中传递时,会发生特征信息丢失现象,为小目标车辆的检测带来了挑战^[3]。

目前主流的基于深度学习框架的目标检测算法主要分为以 R-CNN^[4]、Fast R-CNN^[5]、Faster R-CNN^[6] 为代表的双阶段模型和以 YOLO、SSD 为代表的基于卷积神经网络的单阶段模型。前者首先生成候选目标框,在提取目标的特征信息后,进一步预测和识别目标的位置和类别。虽然检测精度高,但速度较慢,无法适用实时检测。自 2016 年开始,YOLO 算法经历了快速发展(YOLOv3、YOLOv4、YOLOv5 和 YOLOX 等),2021 年,Ge 等^[7]提出了 YOLOX 算法,利用解耦头、无锚点和标签分配策略 SimOTA 显著提高了算法性能。因此,YOLOX 网络被广泛应用于各种场景下的对象检测。

当下,有很多学者对 YOLOX 算法进行了研究和改进,并将其应用在各领域中。Wang 等^[8]针对航拍舰船检测精度低等问题,将卷积注意力机制(CBAM)引入 YOLOX 的浅层网络以更好的提取目标特征,并通过焦点损失(Focal Loss)函数调和正负样本数量,提高了对舰船的检测精度。Shen 等^[9]针对复杂背景下的小目标检测缺乏鲁棒性等问题,将 CBAM 注意力机制引入到 YOLOX 的主干中,使网络能够聚焦显著性信息,提升了对小目标的检测效率和检测精度。赵振兵等^[10]在 YOLOX 网络的骨干部分加入全局上下文信息模块,增强模型对外观缺陷特征的学习能力,在预测部分使用 SIoU 损失函数,提高了对变电站仪表外观缺陷检测的准确率。Yi 等^[11]在 YOLOX 的特征金字塔中增加了特征尺度融合检测分支,并提出了一种基于数据域转移的训练策略,将大规模白天数据集与低光照降级后的小规模夜间数据集进行融合,提高了在夜间环境下对车辆/行人的检测精度。Xiong 等^[12]提出了一种改进的轻量级 YOLOX 实时车辆检测算法,对主干网络进行了轻量化设计,尽可能的降低网络参数。此外,

通过引入 α -CIoU 损失函数,提高了预测框的回归精度,也加快了模型的收敛速度。改进算法在参数较少的情况下,检测速度和精度都得到了提高。Luo 等^[13]设计了一种轻量级车辆检测网络 G-YOLOX,该方法将 3×3 卷积和 5×5 卷积交替应用在主干网络中,以达到充分利用不同特征信息的效果。与原 YOLOX 网络相比,G-YOLOX 网络的权重文件和参数量大量减小,并提高了对车辆的检测精度。黄键等^[14]在 YOLOX-S 模型的基础上引入可变形卷积神经网络和焦点损失函数,提高了对车窗目标的识别精度。Han 等^[15]提出一种轻量级的车辆检测方法,该方法将深感受野块应用在骨干网络,以提升目标特征的表现力,并利用图像分割技术,提升检测准确性。实验结果表明,该网络具有较快的检测速度,网络的参数也更少。刘军黎等^[16]提出一种基于 YOLOX 的轻量化无人机航拍目标检测网络 YOLOX-IM,该方法通过增加一个检测头和浅层网络来提高模型对小目标的检测性能,显著提高了对小目标的检测精度。综上所述,目前基于 YOLOX 算法的研究主要侧重于提高对目标的检测精度和网络的轻量化上,针对小目标检测的研究较少。然而,小目标检测一直是计算机视觉领域待解决的难点问题,特别是在无人机航拍视角下,车辆本身像素占比小,且密度较大,同时还伴随着复杂的背景信息。这使得对小目标的检测更加具有挑战性。

目前基于深度学习的目标检测算法对小目标的特征提取能力较弱,特别是在长距离、复杂背景、高密度的场景中,更容易导致漏检。因此,本文提出一种基于优化 YOLOX 网络的密集小目标车辆检测方法。在网络的颈部结合自适应空间特征融和模块,尽可能的融合小目标的特征;为使网络更加聚焦小目标的特征信息,提高对小目标的定位,引入坐标注意力机制;通过优化损失函数,缓解模型训练中正负样本不均衡的问题。

1 YOLOX 网络结构

YOLOX 是旷视科技提出的高性能目标检测网络,结构如图 1 所示。其网络结构主要分为四个部分:输入端(Input)、用于特征提取的主干网络(Backbone)、特征融合的颈部(Neck)以及检测的头部(Prediction)。对输入的图像,采用 Mixup 和 Mosaic 数据增强^[17],在主干网络 CSPDarknet 的起始加入 Focus 操作^[18],将通道维度拼接扩展为四倍,

然后通过卷积提取不同的特征。CSPDarknet 作为其骨干特征提取网络，主要完成对输入图像的特征提取，包括 BaseConv、DWConv、ResUnit、SPP 等模块。在主干网络中获得的三个特征层作为特征提取的输入部分，在网络的颈部进行特征融合。预测部分将检测头改进为 Decoupled Head，利用 SimOTA 技术为不同尺寸的目标动态匹配正样本。虽然 YOLOX 已经是一个优秀的检测模型，但用它来检测无人机航拍图像还存在以下两点问题：

- (1) 无人机拍摄图像区域广，庸杂信息多，小目标数量多且密集，不利于提取特征信息，一些关键信息可能不会被模型所学习。
- (2) 在车辆检测中，目标信息是正样本，背景信息是负样本。由于无人机图像背景复杂，正、负样本数量不平衡，严重影响了模型检测的精度。YOLOX 使用的损失函数是二元交叉熵损失函数，无法调整正负样本之间的平衡性。

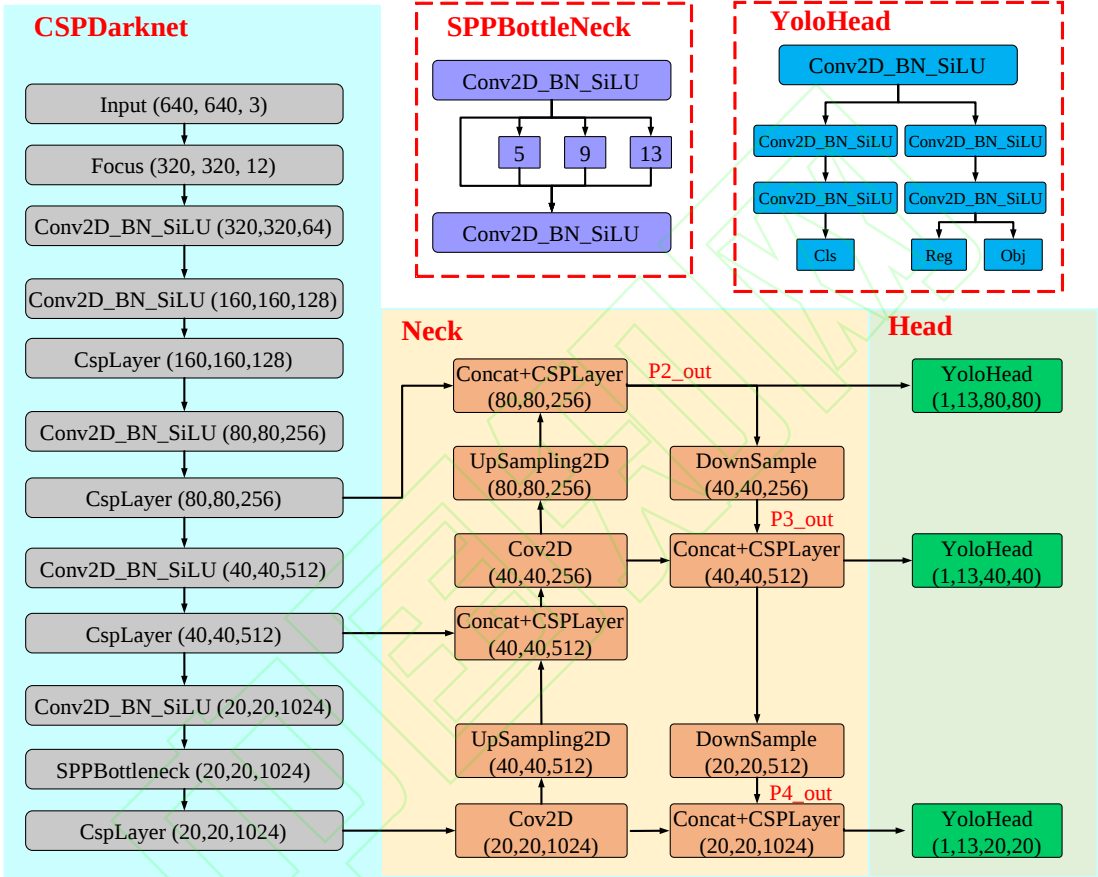


图 1 原始 YOLOX 网络

Fig.1 Original YOLOX network

2 网络的改进与优化

在高空无人机航拍图像中包含了大量小尺寸目标车辆且分布密集，小目标的特征表现不明显。此外，在背景复杂的航拍图像中，引入了更多无关的庸杂信息^[19]，有效特征信息不易提取，导致目标检测的识别准确率低、漏检率较高。因此，本文提出了一种基于改进 YOLOX 网络的密集小目标车辆检测网络，网络结构如图 2，该网络主要由输入端、主

干网络、颈部、头部四部分组成。为了弥补原始网络融合浅层信息的不足，在特征融合部分引入自适应空间特征融合模块，可以自适应的学习每个特征层上的权重参数，使重要特征信息在融合时占据主要地位。将坐标注意力机制嵌入网络的颈部，可以精确捕获位置信息和通道信息，有利于网络定位小目标。最后将二交叉熵损失函数替换为变焦距损失函数，用于改善样本数量不平衡的问题，提高 YOLOX 网络对密集小目标的检测性能。

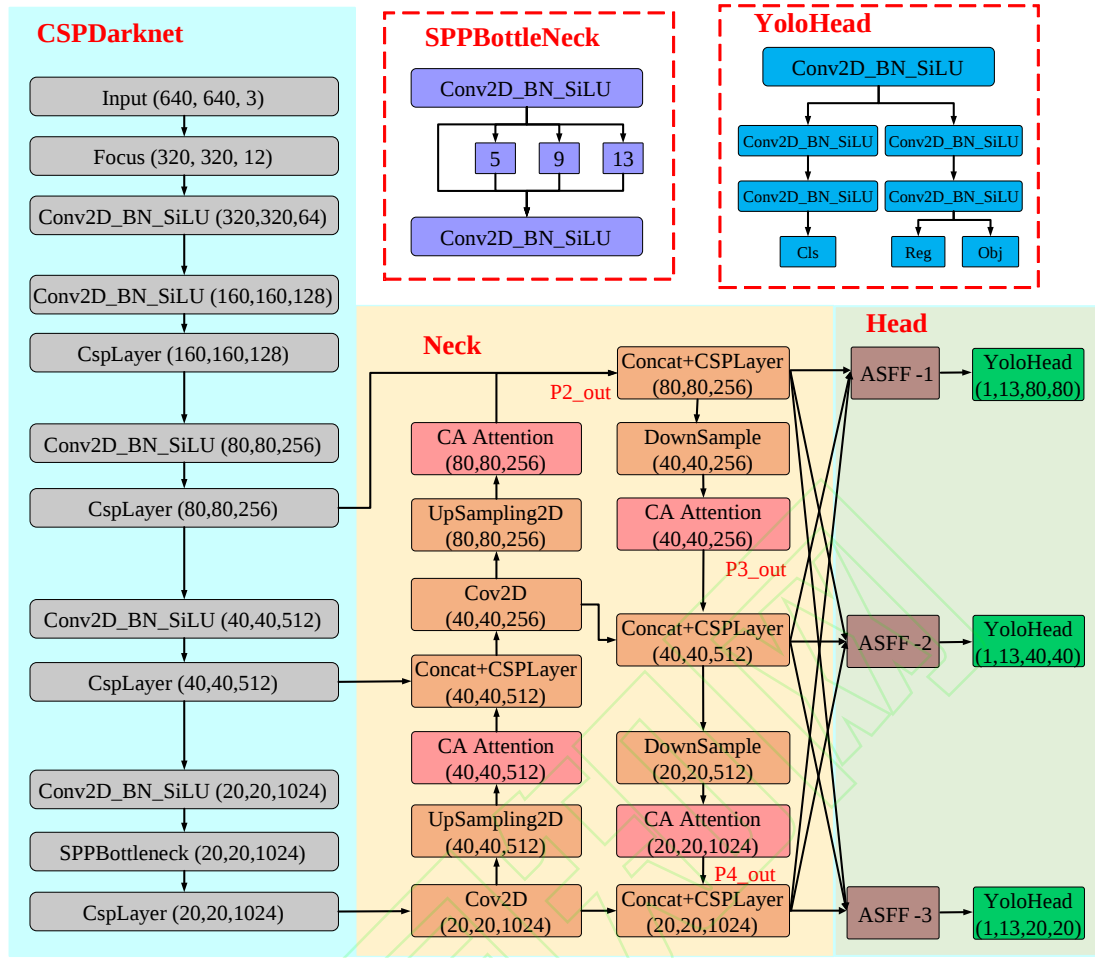


图2 改进的 YOLOX 网络
Fig.2 Improved YOLOX network

2.1 ASFF 模块

在小目标车辆的识别中，不同尺度的特征层所包含的语义信息不同。然而，在无人机航拍图像中，更多的特征信息集中在浅层网络中，在特征融合的过程中容易丢失。本文在颈部网络中使用 ASFF 模块代替原先的 PANet 模块进行特征融合，来缓解上述问题^[25]，结构如图 3。与原先的结构相比，ASFF 结构使网络能够自适应的学习各个特征层上的权重，使重要特征信息在融合时占据主导地位。

ASFF 包括特征调整和自适应融合两部分，特征调整是将其它尺度的特征图缩放到对应尺度的特征图，以保证特征融合时的尺寸不变性；在自适应融合时，将网络训练得到三个不同特征层的重要权

重参数 α 、 β 、 γ 分别与各尺度的特征图相乘。以 ASFF-1 为例，其中 X_1 、 X_2 、 X_3 是分别来自 YOLOX 路径聚合网络输出三个尺度的特征图， $X^{2 \rightarrow 1}$ 、 $X^{3 \rightarrow 1}$ 分别是 X_2 、 X_3 经过特征缩放后生成与 X_1 相同尺寸的特征层。然后将 $X^{1 \rightarrow 1}$ 、 $X^{2 \rightarrow 1}$ 、 $X^{3 \rightarrow 1}$ 分别与对应的权重参数 α 、 β 、 γ 相乘后再相加输出 Y^1 ，通过公式 (1) 进行融合计算：

$$Y_{ij}^1 = \alpha_{ij}^1 \cdot X_{ij}^{1 \rightarrow 1} + \beta_{ij}^1 \cdot X_{ij}^{2 \rightarrow 1} + \gamma_{ij}^1 \cdot X_{ij}^{3 \rightarrow 1} \quad (1)$$

式中： Y_{ij}^1 为通过 ASFF-1 得到的新特征图，

$\alpha_{ij}^1 + \beta_{ij}^1 + \gamma_{ij}^1 = 1$ ， α_{ij}^1 、 β_{ij}^1 和 $\gamma_{ij}^1 \in [0, 1]$ ， $X_{ij}^{n \rightarrow l}$ 表示经过上述相关操作得到的第 l 层特征图融合前的第 n 层特征图上 (x, y) 处的特征向量。

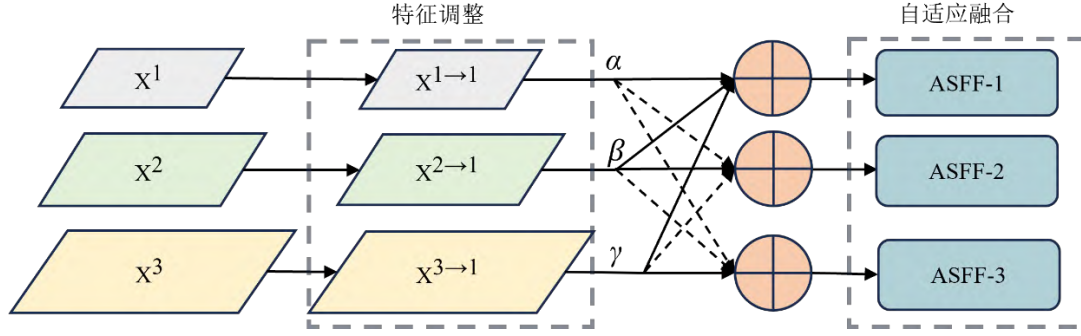


图 3 ASFF 模块

Fig.3 ASFF module

2.2 坐标注意力机制

注意力机制在计算机视觉和深度学习中得到了广泛的应用，以提高网络的鲁棒性和泛化能力。在多种注意力形式中，通道注意力机制（SENet）^[20]和 CBAM^[21]已被用于车辆检测的网络中，以提升网络的检测效能。SENet 在通道维度进行特征提取，形

成每个通道的加权值。CBAM 使用两个独立的模块对特征进行细化。然而，两者在小目标检测领域表现不佳。因此，引入坐标注意力机制^[22]，结构如图 4，不仅获取了通道间信息，还考虑了方向相关的位置信息，有助于网络更好地定位和识别小目标。

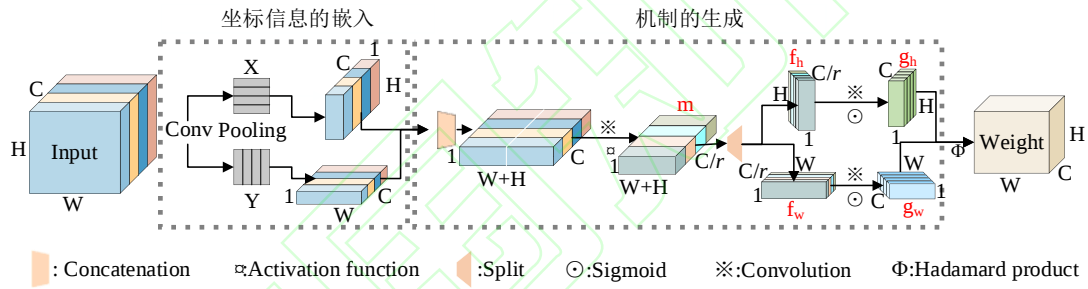


图 4 坐标注意力机制

Fig.4 Coordinate Attention mechanism

坐标注意力机制包括坐标信息的嵌入和机制的生成两部分^[23]。坐标信息的嵌入，通过对于输入维度为 $H \times W \times C$ 的特征图，分别在水平方向和垂直方向用 $(H, 1)$ 和 $(1, W)$ 的池化核进行平均分化对通道进行编码，产生 $C \times H \times 1$ 和 $C \times 1 \times W$ 维度的特征图。计算公式如式（2）和（3）。

$$Z_c^h(h) = \frac{1}{W} \sum_{0 \leq i \leq w} x_c(h, i) \quad (2)$$

$$Z_c^w(w) = \frac{1}{H} \sum_{0 \leq j \leq h} x_c(j, w) \quad (3)$$

式中： $Z_c^h(h)$ 表示高度为 h 第 c 通道的输出，

$Z_c^w(w)$ 表示宽度为 w 第 c 通道的输出。

坐标机制的生成，通过将上述生成的两个特征图在空间维度上进行堆叠，并使用 1×1 卷积进行降维和激活操作，使通道数由 C 降为 C/r ，其中 r 用于控制减少率，然后通过非线性操作对空间信息进行编码处理，生成维度为 $1 \times (W+H) \times C/r$ 的特征图 m ，其计算公式如式（4）。

$$m = \delta \left(F_1 \left([z^h, z^w] \right) \right) \quad (4)$$

式中： F_1 表示 1×1 的卷积变换， $[z^h, z^w]$ 表示堆叠操作， δ 表示非线性激活函数。

为了生成坐标注意力，在空间维度上通过 Split 操作将 m 分解成两个独立的单张量 $f^h \in \mathbb{R}^{C/r \times H \times 1}$

和 $f^w \in \mathbb{R}^{C/r \times 1 \times W}$ ，再分别用两个卷积变换函数 F_h

和 F_w 转换通道数,通过结合 Sigmoid 激活函数得到注意力向量 $g^h \in \mathbb{R}^{C \times H \times 1}$ 和 $g^w \in \mathbb{R}^{C \times 1 \times W}$ 。计算公式如式(5)和(6)

$$g^h = \sigma(F_h(f^h)) \quad (5)$$

$$g^w = \sigma(F_w(f^w)) \quad (6)$$

式中: F_h 和 F_w 表示 1×1 的卷积变换, σ 表示 Sigmoid 激活函数,代表 0-1 的值,赋予重要度。

最后,对 g^h 和 g^w 扩张、相乘后,分别作为注意力权重^[24]。坐标注意力机制的输出公式可以写成:

$$y_c(i, j) = x_c(i, g) \times g_c^h(i) \times g_c^w(j) \quad (7)$$

2.3 优化损失函数

YOLOX 使用 BCE Loss 函数来计算对象的分类损失和置信度损失,使用 IoU 作为预测框回归的损失。在无人机拍摄图像中进行车辆目标检测时,会存在前后背景差别较大的问题,使得目标类与背景类处在极度不平衡状态。Varifocal Loss 函数可以缓解目标类和背景类之间数量不平衡产生的影响^[26]。因此,本文用 Varifocal Loss 函数代替 BCE Loss 函数。BCE Loss 函数计算公式如式(8)。

$$L_{BCE} = -\log P_t = \begin{cases} -\log(R), & y=1 \\ -\log(1-R), & y=0 \end{cases} \quad (8)$$

式中: $y=1$ 表示正样本, $y=0$ 表示负样本, R 表示预测后的置信度。BCE Loss 函数被广泛用在简单的检测和分类问题中,但它无法调和正负样本间的比例,因此,在正负样本数量差别较大的任务中表现不佳。焦点损失(Focal Loss)函数可以改善目标类与背景类之间不平衡的问题。计算公式如式(9)。

$$L_f = -\alpha_t(1-P_t)^\gamma \log P_t \\ = \begin{cases} -\alpha(1-R)^\gamma \log R, & y=1 \\ -(1-\alpha)R^\gamma \log(1-R), & y=0 \end{cases} \quad (9)$$

式中: R 为目标类的预测概率, y 为正负样本的类别,取值为 1 或 0, α 是可调比例因子因子, $(1-R)^\gamma$ 为目标类调制因子, R^γ 为背景类调制因子,两者用于调整目标类与背景类间的平衡。Focal Loss 能够使用加权的方法解决类别不平衡的问题。

Focal Loss 采用平等的方法处理正负样本失衡的问题,而在实际的检测中,样本中的正例对预测效果的影响更为重要。Varifocal Loss 函数在 Focal Loss 函数的基础上增加了一种动态加权运算,计算公式如式(10)。

$$L_{vf} = \begin{cases} -q(q \log(R) + (1-q) \log(1-R))^\gamma, & q > 0 \\ -\alpha R^\gamma \log(1-R), & q = 0 \end{cases} \quad (10)$$

式中: R 是预测值,表示目标得分。 $q=0$ 表示背景类, $q>0$ 表示目标类,对于目标类,将正样本的 q 值设置为预测框和目标框的 IoU 值,否则设置为 0。对于背景类,所有目标的 q 值均设置为 0。

Varifocal Loss 使用 R^γ 缩放因子来降低负样本的损失贡献,但不会用同等的方式处理正样本,这主要考虑到正样本对检测效果的贡献更为重要。同时,为了动态调整正负样本之间的比例,还在损失函数中增添了可调比例因子 α 。Varifocal Loss 函数的引入可以提高对密集小目标的分类精度,缓解正负样本间的不平衡性^[27]。

3 实验

3.1 实验环境

为保证实验的严谨性,所有实验均使用相同的环境来运行,具体环境配置细节如表 1 所示。

表 1 参数配置

实验平台	型号参数
CPU	12 vCPU Intel(R) Xeon(R) Platinum 8255C CPU @2.50GHz
GPU	GeForce RTX 3080(10GB)
操作系统	Windows 10
框架	Pytorch
编程环境	Python

3.2 数据集

以 大疆御 mavic 2 pro 专业版作为本文图片数据的采集平台。该无人机搭载了哈苏 L1D-20c 相机,在精准识别、轨迹预测和高速跟随等方面具有较好的表现,并支持 8 公里距离控制、1080P 高清图传、4K HDR 视频拍摄及 f/2.8-f/11 可调光圈,可适应多种光照环境。利用高清数字图传技术,可以实时查看航拍画面,无人机及飞行控制界面如图 5 所示。



图 5 无人机及其操作界面

Fig.5 UAV and its operation interface

本文以重庆市典型的山地城市干道复杂交织区为数据采集对象，在无风良好的时间段，使用无人机在距离地面 120-250m 的高空悬停拍摄车辆运行视频，在获得车辆运行视频之后，按照设定的适当帧率，提取图片后，并运用图片切割算法将分辨率为 3840×2160 像素图像分割为 768×768 像素进行标注。数据集共 1992 张图片，包含汽车、货车、公共汽车三个类别的车辆，按照 7: 1: 2 的比例将车辆图片划分为训练集、验证集和测试集。其中每张图片包括了多个训练对象，且排列较为密集，更符合无人机航拍视角下的密集小目标车辆检测情况，部分数据集如图 6 所示。本文研究所用的数据是 CQSkyEyeX 数据集：<http://www.cqskyeyex.com>。中用于训练图片的一部分。



图 6 数据集中的部分图片

Fig.6 Some images in the dataset

3.3 评价指标

为了直观、全面地评价改进网络的性能^[28]，本文采用目标检测任务中常用的评价模型性能好坏的指标，表 2 列出了常用的几种用于评估模型检测效果的指标。TP (True Positive)：真正例，即本就是正样本并且被正确识别为正样本的部分；TN (True Negative)：真负例，即本身是负样本被识别为负样本的部分；FP (False Positive)：假正例，即负样本被错误的识别为正样本；FN (False Negative)：假负例，即正样本被错误的识别为负样本。

表 2 网络检测性能的评价指标

Table 2 Evaluation index of network detection performance

指标	公式	意义
Precision	$P = \frac{TP}{TP + FP}$	在所有检测到的目标中 检测到正确的概率
Recall	$R = \frac{TP}{TP + FN}$	在所有阳性样本中正确 识别的概率
F1_score	$F1 = \frac{2 \times R \times P}{R + P}$	准确率和召回率的 调和均值
AP	$AP = \int P \cdot R \cdot dR$	同一召回率下精度的 平均值
mAP	$mAP = \frac{1}{C} \sum_k AP_k$	数据集中所有平均精度 之和的平均值

3.4 网络训练

在模型训练过程中，所有参数设置均保持一致，所用图像经过 Mixup 及 Mosaic 处理后输入到网络中，具体网络参数设置详见表 3

表 3 训练过程中的参数设置

Table 3 Parameter settings during training

参数	值
训练轮次	200
类别	3
批次	8
初始学习率	0
动量	0.9
权重衰减	5e-4
检测阈值	0.35
解冻训练	30
非极大值抑制	0.65

3.5 结果分析

在模型训练中，改进 YOLOX 和原始 YOLOX 的损失曲线如图 7 所示。图 7 (a) 中，可以看出，本文提出的改进网络的总体收敛速度优于原始网络，

在解冻训练后，损失值收敛明显加快。最终的损失值更小，网络更稳定。从图 7 (b) 看出，前 100 迭代分类损失值波动较大，改进网络的损失值在 100 轮次后的迭代中表现更低，原始网络的损失函数曲线存在较大畸变。图 7 (c) 中，两模型的置信度损失在前期波动较大，改进网络的曲线后期没有剧烈

波动，改进后的置信度损失收敛更快且具有更低的损失值。从图 7 (d) 看出，在前 50 个迭代轮次中，两模型的收敛速度相对较快，随后两曲线趋于稳定波动，改进后的模型相对稳定，在解冻后两曲线损失值均呈现急速下降状态。

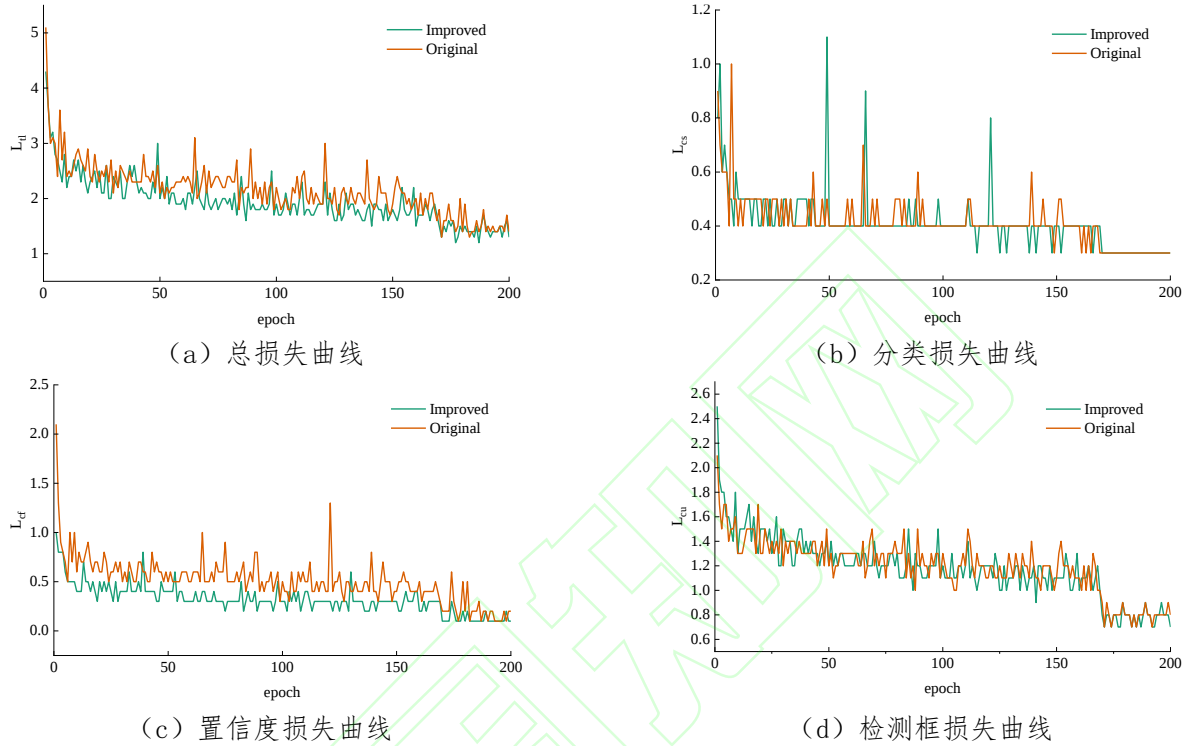


图 7 训练过程中的损失曲线对比

Fig.7 Comparison of the descent curve of the loss value in training

为了更清楚地分析每个改进模块对模型检测性能的影响，创建了四组消融实验^[29]，并在相同的参数设置下迭代了 200 轮次，进行了验证实验。所得结果如图 8 所示，曲线显示了每添加一个模块后的 mAP 值的变化趋势，包括 $mAP@50$ 和 $mAP@50_{95}$ 。可

以看出，mAP 值在不同模块的相继添加后均呈现不同程度的提升，并在最后的解冻阶段 mAP 值继续上升。从局部放大图可以看出，相比原始 YOLOX，本文提出的改进网络的 $mAP@50$ 和 $mAP@50_{95}$ 的值更突出，且在迭代周期内曲线具有更好的收敛性。

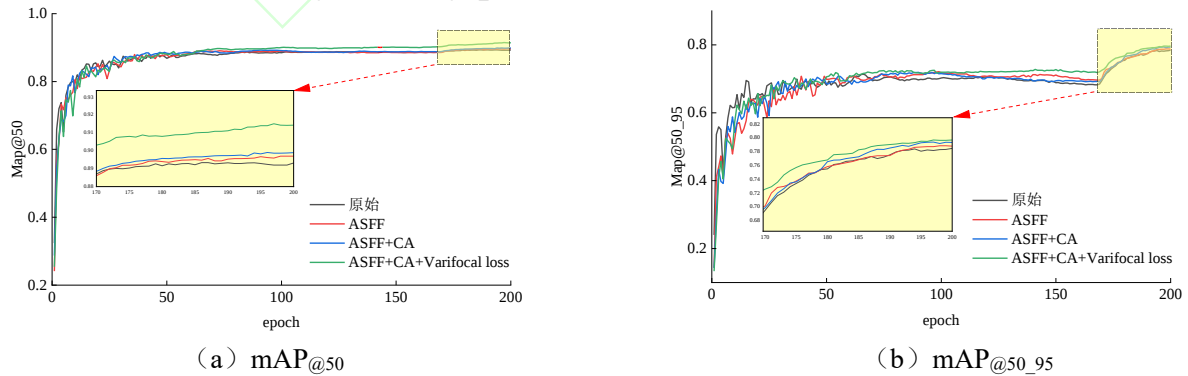


图 8 消融实验中的(a) $mAP@50$, (b) $mAP@50_{95}$ 变化

Fig.8 (a) $mAP@50$, (b) $mAP@50_{95}$ in ablation experiment

此外，通过表 4 可以更加直观的看出，在各模块加入后，各个子项的具体数值表现（ $mAP_{@50}$ 、 $mAP_{@50_{95}}$ ）。其中，Parameters 表示网络内的参数总数，用于度量模型的大小，参数量的单位为 M，检测结果的精度均为百分数。

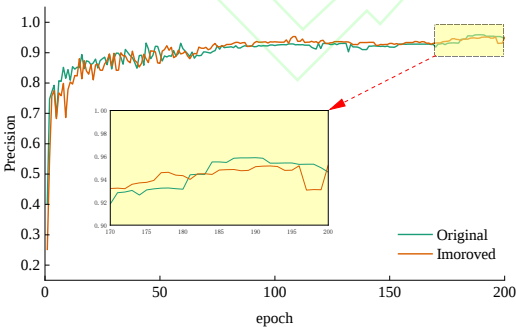
对表中实验结果的分析表明，本文提出的改进 YOLOX 网络，具有最高的 mAP 值，相比原始网络， $mAP_{@50}$ 和 $mAP_{@50_{95}}$ 分别提高了 2.2%、1.21%。在不同结构加入原始 YOLOX 网络后，均出现了不同程度的性能提升，说明本文的改进模块均对小目标检测有一定的效果。

表 4 消融实验对比结果

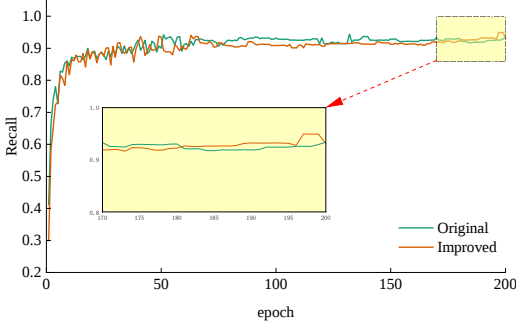
Table 4 Comparative results of ablation experiments

网络	$mAP_{@50}/\%$	$mAP_{@50_{95}}/\%$	Parameters/M
Original YOLOX	89.3	78.44	8.94
ASFF	89.7	78.83	14.38
ASFF+CA	89.9	79.35	14.40
Improved YOLOX	91.5	79.65	14.40

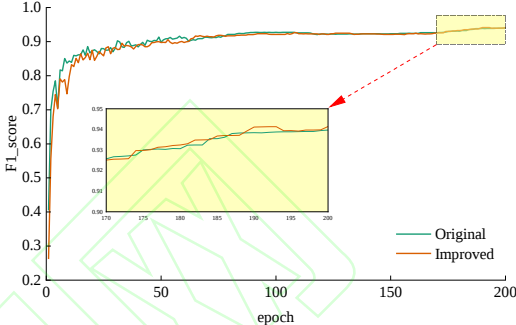
图 9 显示了原始 YOLOX 和改进 YOLOX 的精确度（Precision）、召回率（Recall）、F1 测量值（F1_score）值随着迭代次数变化的曲线。在前 30 个轮次，三个指标的数值急速增加，波动较大。随着迭代次数的增加，各项指标的数值增大放缓，曲线的形态较为平稳。图 9（a）中，改进 YOLOX 网络的 Precision 在解冻后的第 180 轮次出现略微的下降，但最终改进网络的精确度高于原始网络。从图 9（b）、（c）看出，在解冻后，改进 YOLOX 网络的 Recall、F1_score 值均高于原始 YOLOX 网络。



(a) 精确度



(b) 召回率



(c) F1 测量值

图 9 训练中的精确度、召回率和 F1 测量值的变化

Fig.9 Changes in precision, recall, and F1 measurements in training

改进的 YOLOX 网络的 P-R 曲线如图 10 所示，曲线和坐标轴包围的区域是表示 AP 值。可以看出，三个目标的 P-R 曲线的包围面积都比较大，反映出 AP 值比较高。由于货车和公共汽车在数据集中的数量较少，它们的 P-R 曲线并不平滑。汽车在数据集中的比例很高，汽车的 P-R 曲线是连续和平滑的并且具有最高的包围面积。

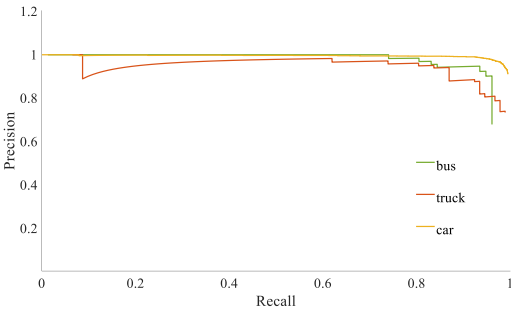


图 10 精度-召回率曲线

Fig.10 Precision-recall curves.

为了检验提出的改进网络对密集小目标车辆检测的有效性和先进性，本文进一步选取了具有代表性的单阶段目标检测模型 SSD、YOLOv5 和双阶段目标检测模型 Faster R-CNN 进行对比实验。实验结

果如表 5 所示。由表 5 可以得出，在对密集小目标车辆的检测中，本文算法在参数量和 Gflops 增加较少的前提下，具有最高的 mAP、召回率、F1_score

值，与其他算法相比 mAP 值提高了 2.2%-28.7%，相较于其他算法具有一定的优越性

表 5 不同目标检测算法对比结果

Table 5 Compare the results of different object detection algorithms

网络	mAP/%	Parameters/M	Precision/%	Recall/%	F1_score	Gflops/G
Original YOLOX	89.3	8.94	95.89	93.69	93.97	26.8
Faster R-CNN	62.8	137.00	46.20	72.47	56.33	185.1
SSD	88.6	26.29	78.38	93.75	85.33	140.9
YOLOv5	89.3	7.10	84.91	91.33	81.00	16.5
Improved YOLOX	91.5	14.40	95.33	94.94	94.13	35.2

同时，为了验证改进 YOLOX 算法在实际检测场景中的效果，选取了多张不同场景下的图像进行对比测试，测试图像均由无人机在不同高度拍摄得到，检测效果如图 11 所示。图 11 (a)、11 (b)、11 (c)、11 (d)、11 (e) 分别为原始 YOLOX 算法、Faster R-CNN 算法、SSD 算法、YOLOv5 算法和本文算法的可视化结果图。检测框中有预测目标的类别及其置信度大小，置信度越高表示算法的性能越好。图中蓝色四角形和蓝色矩形框表示模型未检测到的车辆，两者数量越少表示算法的检测性能越好，在检测效果较好的原始 YOLOX 算法和改进 YOLOX 算法结果图上添加局部放大图进行更详细的对比。每张图片均包含了大量的密集小目标车辆，并且在匝道上行驶的车辆角度也不尽相同。根据图

中蓝色矩形框数量、面积大小以及四角形数量可以看出，在图 11 (b)、(c)、(d) 中的每个场景下均出现了大量的漏检，改进后网络的车辆漏检数量显著减少，检测效果更好。此外，根据局部放大图可知，改进后的算法在检测小目标车辆时具有更高的置信度，并且可以很好的区分车辆类别。

综上所述，本文提出的改进 YOLOX 网络对于无人机航拍视角下的密集小目标车辆具有较好的检测效果，优化后的算法降低了无人机图像中小目标车辆的漏检率，并且置信度得到了进一步的提升，特别是在车辆密集分布的场景中，改善效果更显著。从直观数据和场景应用效果上验证了改进网络的可行性，更适合无人机航拍图像中密集小目标车辆的检测。



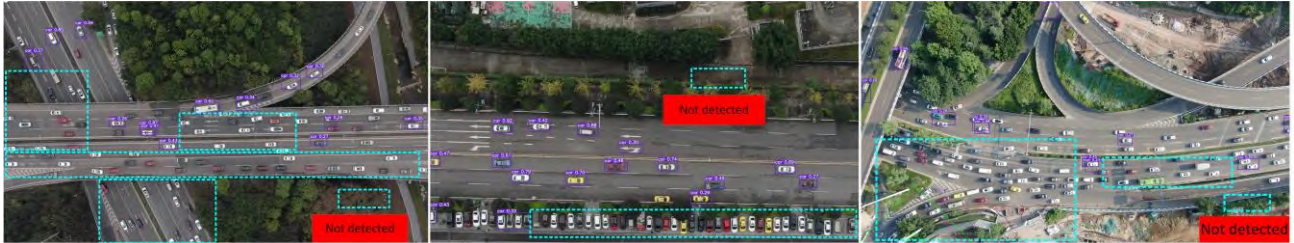
(a) 原始 YOLOX



(b) Faster R-CNN



(c) SSD



(d) YOLOv5



(e) 改进 YOLOX

图 11 不同目标检测算法可视化结果对比

Fig.11 Comparison of visualization results of different object detection algorithms

4 结束语

针对高空无人机拍摄图像的检测任务，原有 YOLOX 算法存在漏检率高、检测精度低等问题。本文对 YOLOX 网络进行了多方面的改进，最后提出了一种改进的 YOLOX 检测网络。首先，在颈部的末端加入自适应空间特征融合模块，充分利用了不同尺度的信息，改善小目标的检测效果。其次，引入坐标注意力机制，有助于强化小目标的特征，使模型能够更准确地提取小目标车辆特征，从而提高目标检测的准确性。为了平衡正负样本之间的数量，将二元交叉熵损失函数替换为变焦距损失函数。

从 Loss 可视化、mAP 值、精确度、召回率以及实际场景检测效果等多个方面，对优化算法和原始算法训练得到的两模型进行详细的对比分析。经实验表明，本文所构建的改进模型，相比于主流的双阶段和单阶段目标检测算法，mAP 值提高了 2.2%-28.7%，多项指标均为最高，具有一定的优势。实际场景下的检测可视化验证了本文算法对于无人机航拍图像中密集小目标车辆的检测具有较好的检测效果。相比其他算法，改进后的 YOLOX 算法具有更

低的漏检率以及更高的检测精度，对小目标车辆的位置检测也更准确，可以胜任无人机航拍图像中密集小目标的检测任务。

参考文献

- [1] Liu C, Yag D, Tang L, et al. A Lightweight Object Detector Based on Spatial-Coordinate Self-Attention for UAV Aerial Images[J/OL]. Remote Sensing, 2023, 15(1): 83.
- [2] Liu W, Wang M, Zhang S, et al. Research on vehicle target detection technology based on UAV aerial images[C/OL]/2022 IEEE International Conference on Mechatronics and Automation (ICMA). 2022: 412-416.
- [3] 李旭, 宋世奇, 殷晓晴. 基于目标空间分布特征的无人机航拍车辆实时检测技术研究[J]. 中国公路学报, 2022, 35(12): 193-204.
- [4] Li Xu, Song Shi-qi, Yin Xiao-qing. Research on real-time detection technology of UAV aerial photography vehicles based on target spatial distribution characteristics[J]. China Journal of Highway and Transport, 2022, 35(12): 193-204.
- [4] Girshic R, Donahue J, Darrell T, et al. Rich Feature

- Hierarchies for Accurate Object Detection and Semantic Segmentation[C/OL]//2014 IEEE Conference on Computer Vision and Pattern Recognition. 2014: 580-587.
- [5] Girshick R. Fast R-CNN[C/OL]//2015 IEEE International Conference on Computer Vision (ICCV). 2015: 1440-1448.
- [6] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J/OL]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
- [7] Ge Z, Liu S, Wang F, et al. YOLOX: Exceeding YOLO Series in 2021[M/OL]. arXiv, 2021[2023-06-27]. <http://arxiv.org/abs/2107.08430>.
- [8] Wang Y, Li J, Chen Z, et al. Ships' Small Target Detection Based on the CBAM-YOLOX Algorithm[J/OL]. Journal of Marine Science and Engineering, 2022, 10(12): 2013.
- [9] Shen C, Ma C, Gao W. Multiple Attention Mechanism Enhanced YOLOX for Remote Sensing Object Detection[J/OL]. Sensors, 2023, 23(3): 1261.
- [10] 赵振兵, 马迪雅, 石颖等.基于改进 YOLOX 的变电站仪表外观缺陷检测算法[J/OL].图学报:1-10[2023-07-17].
Zhao Zhen-bing, Ma Di-ya, Shi Ying, et al. Appearance defect detection algorithm for substation instrument based on improved YOLOX[J/OL].Journal of Graphics:1-10[2023-07-17].
- [11] Yi K, Luo K, Chen T, et al. An Improved YOLOX Model and Domain Transfer Strategy for Nighttime Pedestrian and Vehicle Detection[J/OL]. Applied Sciences,2022, 12(23):12476.
- [12] Xiong C, Yu A, Yuan S, et al. Vehicle detection algorithm based on lightweight YOLOX[J/OL]. Signal, Image and Video Processing, 2023, 17(5): 1793-1800.
- [13] Luo Q, Wang J, Gao M, et al. G-YOLOX: A Lightweight Network for Detecting VehicleTypes[J/OL]. Journal of Sensors, 2022, 2022: 1-10.
- [14] 黄键, 徐伟峰, 苏攀等.基于 YOLOX-S 的车窗状态识别算法[J].吉林大学学报(理学版),2023,61(04):875-882.
Huang Jian, XU Wei-feng, SU Pan, et al. Window state recognition algorithm based on YOLOX-S[J].Journal of Jilin University(Science Edition),2023,61(04):875-882.)
- [15] Han S, Yoo J, Kwon S. Real-Time Vehicle-Detection Method in Bird-View Unmanned-Aerial-Vehicle Imagery[J/OL]. Sensors, 2019, 19(18): 3958.
- [16] 刘军黎, 刘晓锋, 邱洁等.YOLOX-IM:一种无人机航拍视频的轻量化交通参数提取模型[J].国外电子测量技术,2023,42(01):159-169.
- Liu Jun-li, Liu Xiao-feng, Qiu Jie,et al. YOLOX-IM:A lightweight traffic parameter extraction model for UAV aerial video[J].Foreign Electronic Measurement Technology,2023,42(01):159-169.)
- [17] 张稀柳, 张晓玲, 何敏军.基于改进 YOLOX-s 的车辆检测方法研究[J/OL].系统仿真学报:1-10[2023-07-14].
Zhang Xi-liu, Zhang Xiao-ling, He Min-jun. Research on vehicle detection method based on improved YOLOX-s[J/OL].Journal of System Simulation:1-10[2023-07-14].
- [18] Peng H, Tan X. Improved YOLOX's Anchor-Free SAR Image Ship Target Detection[J/OL].IEEE Access, 2022, 10: 70001-70015.
- [19] Luo X, Wu Y, Zhao L. YOLOD: A Target Detection Method for UAV Aerial Imagery[J/OL]. Remote Sensing, 2022, 14(14): 3240.
- [20] Hu J, Shen L, Albanie S, et al. Squeeze-and-Excitation Networks[J/OL]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011-2023.
- [21] Woo S, Park J, Lee J Y, et al. CBAM: Convolutional Block Attention Module[C/OL]//FER.Computer Vision – ECCV 2018. Cham: Springer International Publishing, 2018: 3-19.
- [22] Hou Q, Zhou D, Feng J. Coordinate Attention for Efficient Mobile Network Design[C/OL]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2021: 13708-13717.
- [23] Zhang F, Cao W, Wang S, et al. Improved YOLOv4 recognition algorithm for pitaya based on coordinate attention and combinational convolution[J/OL]. Frontiers in Plant Science, 2022, 13[2023-02-05]. <https://www.frontiersin.org/articles/10.3389/fpls.2022.1030021>.
- [24] Xuan W, Gao Jian-zhuang, Hou Bo-jie, et al.A Lightweight Modified YOLOX Network Using Coordinate Attention Mechanism for PCB Surface Defect Detection[J/OL]. IEEE Sensors Journal, 2022, 22(21): 20910-20920.
- [25] Song Z, Huang X, Ji C, et al. Deformable YOLOX: Detection and Rust Warning Method of Transmission Line Connection Fittings Based on Image Processing Technology[J/OL]. IEEE Transactions on Instrumentation and Measurement, 2023, 72: 1-21.
- [26] Song Chao-yu, Zhang F, Li Jian-sheng, et al. Detection of maize tassels for UAV remote sensing image with an improved YOLOX model[J/OL]. Journal of Integrative

Agriculture, 2022: S2095311922002465.

- [27] Qu Z, Shang X, Xia Shu-fang, et al. A method of single-shot target detection with multi-scale feature fusion and feature enhancement[J/OL]. IET Image Processing, 2022, 16(6): 1752-1763.
- [28] Wang G, Liu Z, Sun H, et al. YoloX-BTFPN: An anchor-free conveyor belt damage detector with a biased feature extraction network[J/OL]. Measurement, 2022, 200: 111675.
- [29] Kumar S, Jain A, Rani S, et al. Deep Neural Network Based Vehicle Detection and Classification of Aerial Images[J/OL]. Intelligent Automation & Soft Computing, 2022, 34(1): 119-131.

