

基于改进YOLOv8s的双目视觉道路障碍物检测与测距方法研究

樊佳鑫, 郑雨悦, 张丁骅, 张正军

(南京理工大学数学与统计学院, 南京 210094)

摘要: 为了提升汽车辅助系统对前方障碍物的检测效果并进一步获取精确的距离信息, 文章提出了一种基于改进YOLOv8s的交通场景障碍物检测与双目测距方法。该方法以YOLOv8s(You Only Look Once v8s)网络为基础, 首先在Backbone中引入EMA注意力机制, 以提高目标检测精度; 其次将Neck中的PAN-FPN网络替换为ASF(Attentional Scale Sequence Fusion)网络, 并采用DIoU优化损失函数; 在特征匹配算法ORB的基础上, 利用RANSAC算法剔除误匹配的点对。通过在KITTI数据集和实际交通场景中的实验, 结果表明, 在20 m的距离范围内, 改进后的YOLOv8s网络对汽车、行人和非机动车3类障碍物的检测mAP(mean average precision)达到了91.1%, 提高了4.8%, 同时测距的平均误差仅为1.55%。

关键词: YOLOv8s; 道路障碍检测; ASF网络; 特征匹配; 双目视觉测距

中图分类号: TP391 **文献标识码:** A

Research on binocular vision road obstacle detection and distance measurement method based on improved YOLOv8s

FAN Jiaxin, ZHENG Yuyue, ZHANG Dinghua, ZHANG Zhengjun

(School of Mathematics and Statistics, Nanjing University of Science and Technology, Nanjing 210094, China)

Abstract: In order to improve the detection effect of car assistance systems on obstacles ahead and further obtain accurate distance information, this article proposes a traffic scene obstacle detection and binocular ranging method based on improved YOLOv8s. This method is based on YOLOv8s (You Only Look Once v8s) network, and first introduces EMA attention mechanism in Backbone to improve the accuracy of object detection. Next, replace the PAN-FPN network in Neck with the ASF (Attention Scale Sequence Fusion) network and use DIoU to optimize the loss function. On the basis of the feature matching algorithm ORB, the RANSAC algorithm is used to eliminate mismatched point pairs. Through experiments on the KITTI dataset and actual traffic scenarios, the results show that within a distance range of 20 meters, the improved YOLOv8s network achieves a mean average precision (mAP) of 91.1% for detecting three types of obstacles: cars, pedestrians, and non motorized vehicles, an increase of 4.8%. At the same time, the average distance measurement error is only 1.55%.

Key words: YOLOv8s, road obstacle detection, ASF network, feature matching, stereo vision distance measurement

1 引言

双目测距通过分析物体在2个相机拍摄的不同视角图像, 获取物体在三维空间中的深度信息。该技术广泛应用于工业检测、机器人导航、自动驾驶以及虚拟现实等领域。目前, 双目相机也在尝试与多种新兴技术相结合^[1]。本文建立了一种新型双目测距系统, 旨在满足实时性要求。为此, 采用改进YOLOv8深度学习模型对道路目标障碍物进行识别, 利用ORB算法进行特征点匹

配, 并引入RANSAC算法剔除误匹配点, 最后通过三角测距原理获取距离信息。

2 硬件平台

实验所使用的硬件主要包括计算机和双目相机。计算机硬件配置如下: CPU为Xeon Gold 5218R, 32 GB内存, GPU为NVIDIA Quadro RTX 6000。双目相机配置为MSK-SM型号, 支持30 fps, 500万像素, 最大图像分辨率为3 840×1 080, 基线为6 cm, 焦距为3 mm, 如图1所示。

软件环境为Linux, Python 3.8, OpenCV。



图1 双目相机

3 目标障碍物的检测

3.1 YOLOv8算法检测过程

本文采用YOLOv8网络来检测汽车、行人和非机动车在相机视野中的相对位置。YOLOv8有5种不同大小的递增模型:n, s, m, l, x, 以深度和宽度为标准^[2]。为了比较不同模型的表现, 本文在KITTI数据集上进行了实

验, 结果如表1所列。n网络的精度较低, 因此未予考虑; s网络在参数量明显低于m, l, x网络的情况下, 精度较高。因此, 本文选择了在精度和速度上相对平衡的YOLOv8s模型进行研究, 其网络结构如图2所示。

YOLOv8s模型包括4个主要部分: Input, Backbone, Neck, Head^[3]。Input部分采用了mosaic数据增强和自适应图片缩放, 以提升模型的鲁棒性。Backbone负责提取图像中的特征, 采用由多个卷积(Conv)模块、C2f模块以及空间金字塔池化(SPPF)模块组成的Darknet-53结构。Neck部分对提取的特征进行处理, 采用PAN-FPN(Path Aggregation Network 和 Feature Pyramid Networks)结构, 通过跨层连接实现特征融合^[4]。Head部分根据处理后的特征进行目标检测, 采用解耦头结构, 将分类头和检测头分开, 同时去除锚框, 加快了目标检测的速度。

表1 YOLOv8检测结果对比

Network	Params/MB	GFLOPs	Precision ^{val}	Recall ^{val}	mAP ₅₀ ^{val}	mAP ₅₀₋₉₅ ^{val}	Speed/(ms/piece)
YOLOv8n	12.0	8.1	0.835	0.695	0.779	0.513	2.3
YOLOv8s	44.5	28.4	0.894	0.752	0.842	0.583	3.3
YOLOv8m	103.4	78.7	0.921	0.765	0.860	0.613	6.1
YOLOv8l	174.4	164.8	0.927	0.785	0.874	0.637	12.1
YOLOv8x	272.5	257.4	0.909	0.811	0.877	0.645	32.7

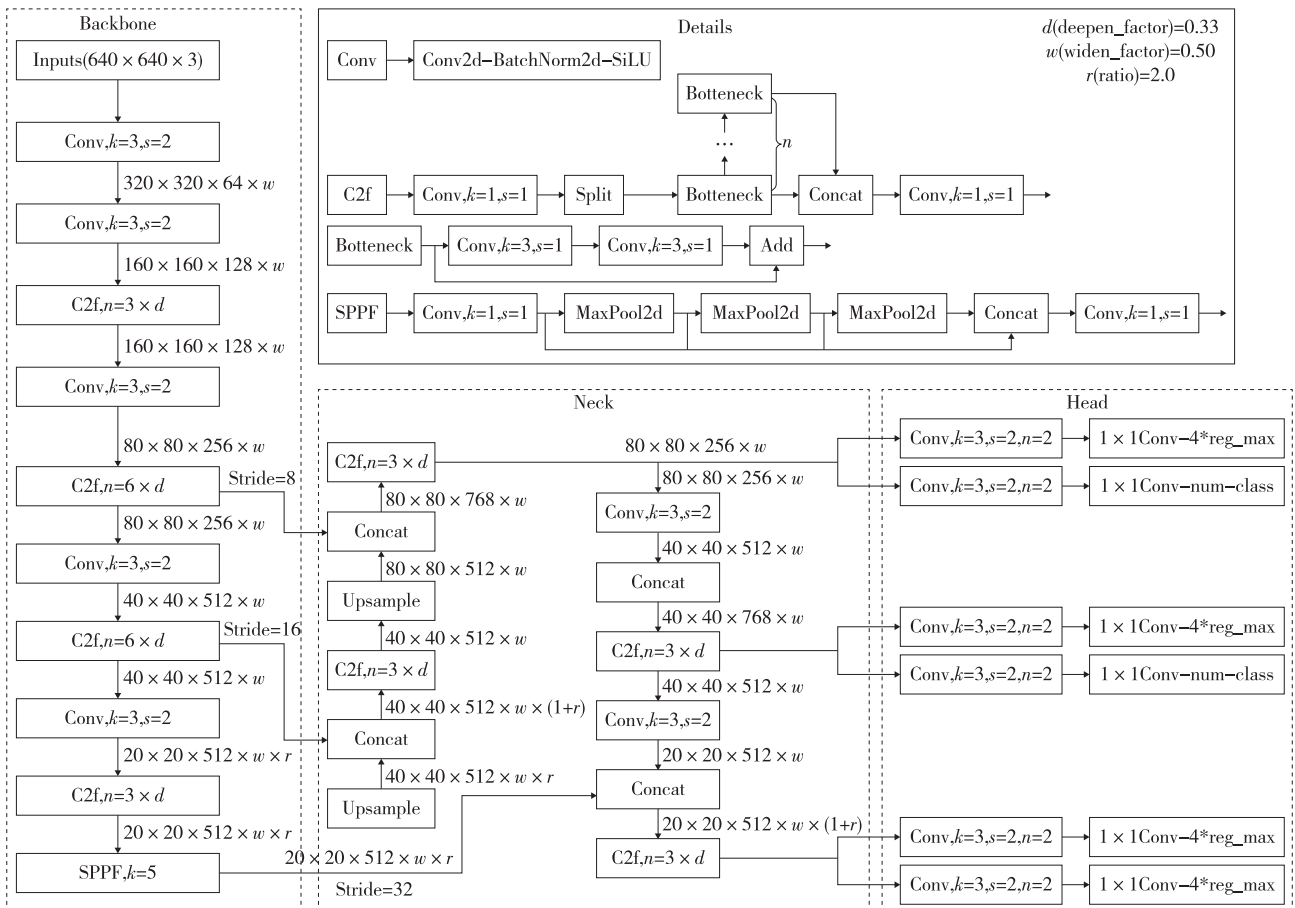


图2 YOLOv8s网络结构

3.2 模型改进

3.2.1 Neck部分的改进

ASF网络^[5]对PANet结构中提取的多尺度特征映射进行融合,该网络由2个主要模块组成:(1)SSFF模块通过标准化、上采样和将多尺度特征连接到一个3D卷积中,结合了不同尺度上的全局语义信息;(2)TFE模块能够捕捉小目标的细微空间信息。然后,通过PANet结构将TFE模块的信息整合到各特征分支中,再将其与SSFF模块得到的多尺度信息结合,最终合并到P3分支中。修改后的网络结构如图3所示。

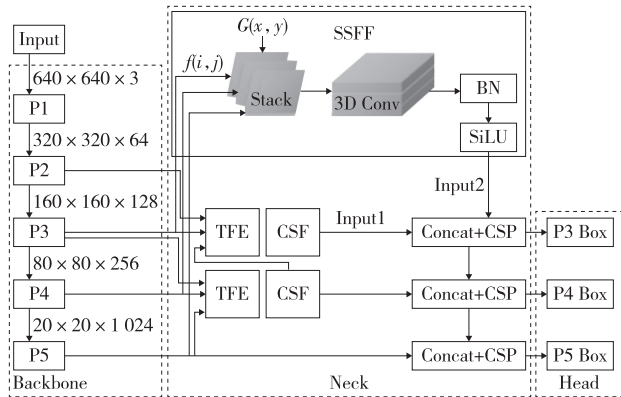


图3 ASF-YOLO模型结构

TFE模块的结构如图4所示。在进行特征编码之前,应调整特征图的通道数,使其与主尺度特征图保持一致。大尺度特征图经过卷积处理后,通道数调整为1C,并通过最大池化和平均池化的组合结构进行下采样,以保留高分辨率特征。对于小尺度特征图,通过卷积调整其通道数后,使用最近邻插值法进行上采样,从而避免小目标特征信息的丢失。最后,将大、中、小3个尺度的特征图调整到相同维度后,再进行一次卷积,并在通道维度上进行拼接。

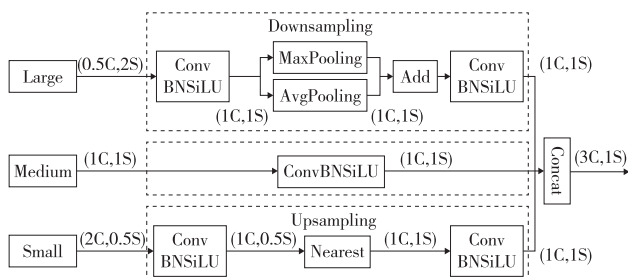


图4 TFE模块结构

3.2.2 添加注意力机制

由于光照、距离目标障碍物远近等因素的影响,检测结果可能会出现偏差。因此,本文在YOLOv8s的Backbone网络中的SPPF模块之前引入了跨空间学习的多尺度注意力机制EMA(Efficient Multi-scale Attention)^[6],如图5所示。该机制在不增加大量计算量的情况下,能够使空间语义特征在每个特征组中均匀分布,从而使网络更加关注数

据中的重要特征,进而提升检测精度。

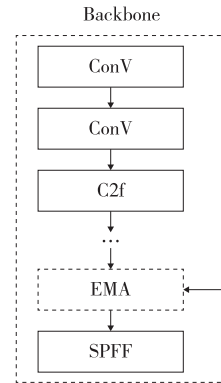


图5 EMA添加位置

EMA算法的工作原理是:首先,EMA将输入的特征图沿通道维度分成G个子特征,并通过3条并行路径来提取每个分组特征图的注意力权重描述符。2条并行路径由1x1卷积分支构成,分别对全局信息进行编码,通过一维全局平均池化在水平和垂直维度上进行信息提取,式(1)和式(2)表示了这一过程。第3条路径由3x3卷积分支构成,通过3x3卷积扩大特征空间。然后,分别对1x1和3x3卷积分支的全局空间信息进行二维全局平均池化(见式(3))。最后,通过聚合2个生成的空间注意力权重值并应用Sigmoid函数来计算每组输出特征图。EMA的结构如图6所示。

$$z_c^H(H) = \frac{1}{W} \sum_{0 \leq i \leq W} x_c(H, i) \quad (1)$$

$$z_c^W(W) = \frac{1}{H} \sum_{0 \leq j \leq H} x_c(j, W) \quad (2)$$

$$z_c = \frac{1}{H \times W} \sum_j \sum_i x_c(i, j) \quad (3)$$

3.2.3 损失函数的改进

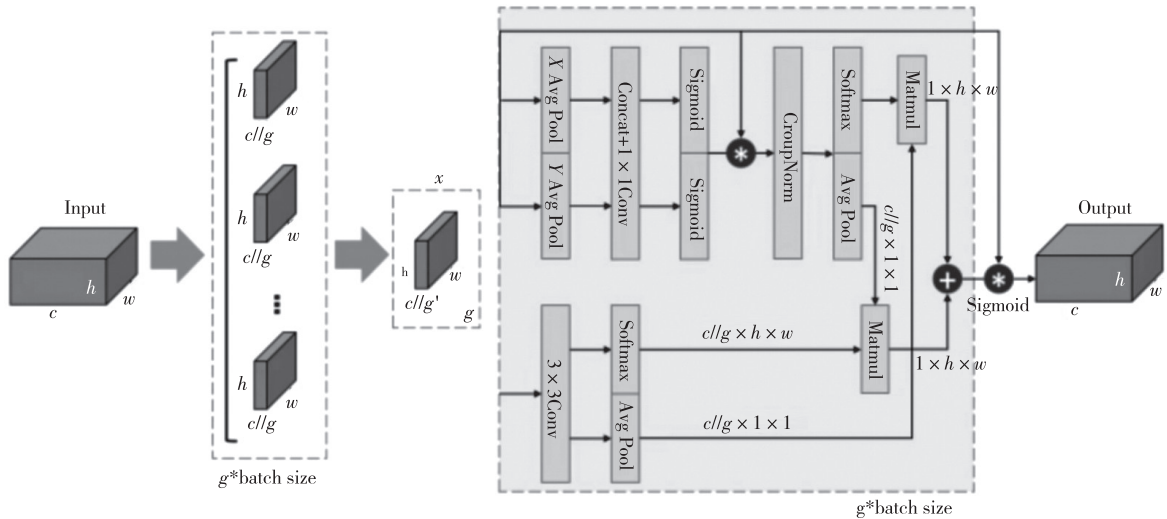
YOLOv8s默认将CIoU作为损失函数^[7],在实验中将其替换为DIoU后,模型在准确率和召回率等关键指标上表现出更优的性能。DIoU使得目标回归框更加稳定,如图7所示。

其中, b 和 b^{gt} 分别为预测框和真实框的中心点; $\rho(\cdot)$ 是欧氏距离; c 是同时包含真实框和预测框的最小包围区域的对角线距离; IoU 为2个候选框交集与并集的比值,计算式如式(4)所示。

$$DIoU = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} \quad (4)$$

4 障碍物的深度计算

本文采用张正友平面标定法^[8]进行相机标定。在实验中,利用Matlab软件的系统默认标定工具箱获取相机的内外参数、畸变参数、焦距以及右侧相机相对于左侧相机的位置参数,结果如表2所列。



Input-输入; c -通道数; h -高; w -宽; g -划分的组; Con-卷积层; batch size-批大小; X Avg Pool-水平方向全局平均池化; Y Ag Pool-垂直方向全局平均池化; Concat-拼接; GroupNorm-组归一化; Sigmoid-激活函数; Matmul-矩阵乘法; Softmax-激活的数; Output-输出

图6 EMA算法结构

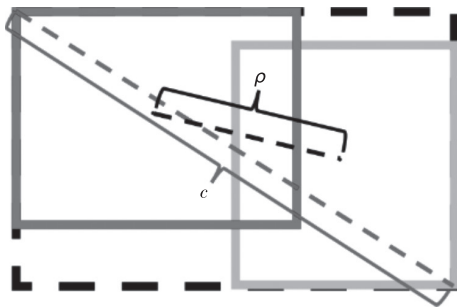


图7 DIoU示意图

表2 相机标定结果

参数名称	参数
左相机焦距	[1 086.0, 1 085.8]
左主点坐标	[1 025.3, 513.728 3]
右相机焦距	[1 086.4, 1 087.0]
右主点坐标	[1 011.2, 541.831 5]

4.1 图像校正

本文采用Bouguet算法进行立体校正。该算法通过最小化图像中的重投影误差,同时最大化观测面积来成立体校正。利用标定得到的旋转矩阵 R 与平移向量 T ,并将 R 分解成2个合成旋转矩阵 r_1 和 r_r ,使得左右2个图像经过这2个合成旋转矩阵变换后,恰好旋转一半角度并精确对齐。校正前后的效果对比如图8所示。

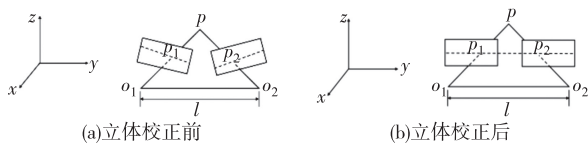


图8 立体校正前后对比

4.2 基于改进ORB的特征点提取与匹配算法

4.2.1 特征点提取

ORB算法的特征点提取部分采用了FAST算子,通过比较一个点与周围点的像素值,找出局部像素变化明

显的点并将其作为特征点。为了实现旋转不变性,ORB算法通过计算特征点灰度值与质心之间的偏移来确定特征点的方向。图像区域 B 的矩计算如式(5)所示。

$$m_{pq} = \sum_{x,y \in B} x^p y^q I(x,y) \quad (5)$$

通过矩计算得到质心如式(6)所示。

$$C = \left(\frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}} \right) \quad (6)$$

然后,连接几何中心 O 与质心 C ,得到方向向量 \overrightarrow{OC} 并将其作为特征点方向。特征描述部分定义BRIEF描述符,描述向量的计算如式(7)所示。

$$\tau(p; A, B) = \begin{cases} 1, I_A > I_B \\ 0, I_A \leq I_B \end{cases} \quad (7)$$

其中, I_A 和 I_B 分别为点 A 和点 B 的灰度值,随机选择 n 对点 (x_i, y_i) 生成二进制字符串,最终的特征向量如式(8)所示。

$$f_n(p) = \sum_{1 \leq i \leq n} 2^{i-1} \tau(p; x_i, y_i) \quad (8)$$

根据特征点的主方向,对 n 个点进行旋转生成新的点对,再通过式(8)生成最终具有旋转不变性的BRIEF描述符。

4.2.2 RANSAC剔除误匹配点对

为了去除误匹配的点对,本文采用了RANSAC算法,其主要步骤如下。

(1)在初始匹配的数据集中随机选取 k 对特征点作为样本数据,并估计当前模型的参数。

(2)设定误差阈值,根据模型估计值计算所有样本点的匹配误差,将误差小于阈值的样本点加入内点集。

(3)更新内点数的最大值,重复步骤(1)和(2),直到达到最大迭代次数或最大内点数大于指定阈值,跳出循环。

(4)对最大内点数所对应的内点集进行模型参数估计。

5 实验结果分析

5.1 改进的YOLOv8s障碍物检测结果

实验使用KITTI数据集进行测试,共7 481张图片,数据集按6:2:2的比例划分为训练集、验证集和测试集。评价指标包括Precision(精度)、Recall(召回率)、 mAP_{50} 和 mAP_{50-95} 。测试结果分为真阳性、真阴性、假阳性和假阴性,基于这些结果计算精度和召回率。 mAP 是所有召回率下准确率的均值。

为了进一步验证所提改进算法的检测性能,本文将改进的YOLOv8s算法与当前主流检测算法SSD, Faster-RCNN, YOLOv5s在相同数据集上进行对比,结果如表3所列。由表3可知,改进的YOLOv8s算法在4个评价指标上均优于其他算法。

表3 检测算法性能对比

Network	P	R	mAP_{50}	mAP_{50-95}
SSD	0.769	0.683	0.741	0.426
Faster-RCNN	0.556	0.717	0.698	0.346
YOLOv5s	0.915	0.767	0.860	0.612
YOLOv8s	0.900	0.787	0.869	0.623
Improved-YOLOv8s	0.927	0.832	0.911	0.662

利用改进的YOLOv8s进行室外目标检测的结果如图9所示,即使在复杂的外部环境下,检测效果依旧良好。



图9 目标检测结果

5.2 改进的ORB特征点匹配结果

在使用改进YOLOv8s进行目标检测后,进一步应用改进的ORB算法进行特征点匹配,结果如图10所示。



(a)提升前



(b)提升后

图10 特征点匹配结果

5.3 距离测量结果

为了验证改进的YOLOv8s目标检测算法与双目测距

算法结合的可行性,从初始距离6 m开始,每次增加2 m,进行了8组双目视觉测距实验,实验结果如表4所列。在10—16 m的范围内,测量误差较小,平均误差仅为0.24%。总体而言,测距的平均误差为1.55%。

表4 测距结果

实际距离/m	测量距离/m	绝对误差/m	相对误差/%
6	6.18	0.18	3.00
8	8.21	0.21	2.63
10	9.99	-0.01	0.1
12	12.01	0.01	0.08
14	14.03	0.03	0.21
16	15.91	-0.09	0.56
18	18.66	0.66	3.67
20	19.57	-0.43	2.15

6 结束语

本文通过改进YOLOv8s算法对道路障碍物进行检测,加入EMA注意力机制、ASF改进Neck网络,并将损失函数替换为DIOU。实验结果表明,在KITTI数据集上,改进后的YOLOv8s算法在精度上提升了3%,召回率提升了5.71%, mAP_{50} 提升了4.8%, mAP_{50-95} 提升了6.3%,综合性能超过了SSD, Faster-RCNN和YOLOv5s等主流目标检测算法。在ORB特征点匹配算法的基础上,添加了RANSAC算法剔除误匹配点,结合双目视觉进行测距。实验结果显示,目标障碍物识别精度达到92.7%,较之前提升了3%,在20 m范围内,平均测距误差仅为1.55%。

参考文献:

- [1] 刘永华.多脉冲远距离激光测距系统研究[J].南方农机, 2023,54(21):146-148.
- [2] 施克权,李祺,隋皓,等.IEMAYoloViT:基于改进YOLOv8的水下目标检测算法[J/OL].电讯技术,1-10[2024-02-22].<https://doi.org/10.20079/j.issn.1001-893x.231206005>.
- [3] 程焕新,乔庆元,骆晓玲,等.基于改进YOLOv8的无人机航拍图像目标检测算法[J/OL].无线电工程,1-10[2024-02-22].<http://kns.cnki.net/kcms/detail/13.1097.TN.20231127.1157.002.html>.
- [4] 赵继达,甄国涌,储成群.基于YOLOv8的无人机图像目标检测算法[J/OL].计算机工程,1-10[2024-02-22].<https://doi.org/10.19678/j.issn.1000-3428.0068268>.
- [5] KANG M, TING C M, TING F F, et al.ASF-YOLO: A novel YOLO model with attentional scale sequence fusion for cellinstance segmentation[J].Image and Vision Computing, 2024:105057.