

doi: 10.3969/j.issn.1002-0268.2025.03.002

# 基于深度学习与数据回归的智能车 单目测距方法

郭贺<sup>1,2</sup>, 张蕊<sup>\*1,2</sup>, 成英<sup>1,2</sup>, 彭涛<sup>1,2</sup>, 夏鹏<sup>3</sup>, 张杨<sup>1</sup>

(1. 天津职业技术师范大学 汽车与交通学院, 天津 300222;

2. 智能车路协同与安全技术国家地方联合工程研究中心, 天津 300222;

3. 天津海豚智行科技有限公司, 天津 300380)

**摘要:**【目标】针对自动驾驶中单目测距精度不高与泛用性不强的问题, 提出了一种基于深度学习与数据回归的单目测距方法。【方法】提出了一种改进的YOLOv5目标检测算法。首先, 采用K-means++算法对先验框进行重新聚类, 优化初始聚类中心点。其次, 在主干网络中引入注意力机制, 有效提升特征提取能力。同时, 将原损失函数替换为E-IOU损失函数, 显著降低了目标定位误差。然后, 基于改良的目标检测算法, 构建了单目测距模型。该模型依托于不同检测对象的斜角长度与摄像头的真实距离建立了若干回归方程, 扩充了单位图像内的像素点数目, 提高了检测精度。最后, 该模型利用卡尔曼滤波融合了汽车的运动数据, 实现了对检测距离的双重数据融合校正。【结果】改进后的算法检测精度较原始YOLOv5提升0.68%, 误报率得到明显改善。本研究算法在10~100 m内的测量平均相对误差为3.67%, 在30~100 m内的测量平均相对误差为3.24%。通过英伟达加速推理器TensorRT对整套算法进行加速, 单张图片推理速度达到12 ms, 比未加速的算法处理速度提升了5倍, 比双目算法处理速度提升了4倍, 且在大于30 m的测距场景中可以取得最优效果。【结论】本研究提出算法可以满足一般精度要求的智能车辆与非智能车辆的测距任务。

**关键词:** 智能交通; 单目测距; 深度学习; 注意力机制; 数据回归

中图分类号: U463.6

文献标识码: A

文章编号: 1002-0268 (2025) 03-0011-10

## Intelligent vehicle monocular ranging method based on deep learning and data regression

GUO He<sup>1,2</sup>, ZHANG Rui<sup>\*1,2</sup>, CHENG Ying<sup>1,2</sup>, PENG Tao<sup>1,2</sup>, XIA Peng<sup>3</sup>, ZHANG Yang<sup>1</sup>

(1. School of Automobile and Transportation, Tianjin University of Technology and Education, Tianjin 300222, China;

2. National and Local Joint Engineering Research Center for Intelligent Vehicle Infrastructure Cooperation and Safety Technology, Tianjin 300222, China;

3. Tianjin Dolphin Zhixing Technology Co., Ltd., Tianjin 300380, China)

**Abstract:** [Objective] In view of the problems of low accuracy and poor generality of monocular ranging in autonomous driving, a monocular ranging method based on deep learning and regression was proposed. [Method] An improved YOLOv5 object detection algorithm was proposed. First, the K-means++ algorithm

收稿日期: 2022-12-21 修改日期: 2025-02-02

基金项目: 国家自然科学基金项目 (52172350); 天津市教委科研项目 (2020KJ120, 2020KJ122); 天津市应用基础研究多元投入基金项目 (21JCZDJC00700)

作者简介: 郭贺 (1996-), 男, 天津人, 硕士, 研究方向为智能网联汽车。(936515285@qq.com)

\* 通讯作者: 张蕊 (1984-), 女, 天津人, 博士, 副教授, 研究方向为智能网联汽车、车联网、智能交通系统。(zhangrui318@163.com)

引用格式: 郭贺, 张蕊, 成英, 等. 基于深度学习与数据回归的智能车单目测距方法 [J]. 公路交通科技, 2025, 42 (3): 11-20. GUO He, ZHANG Rui, CHENG Ying, et al. Intelligent vehicle monocular ranging method based on deep learning and data regression [J]. Journal of Highway and Transportation Research and Development, 2025, 42 (3): 11-20.

© The Author(s) 2025. This is an open access article under the CC-BY 4.0 License. (<https://creativecommons.org/licenses/by/4.0/>)

was used to re-cluster the prior boxes, optimizing the initial cluster centers. Second, an attention mechanism was introduced into the backbone network, effectively enhancing the feature extraction capabilities. Meanwhile, the original loss function was replaced with the E-IOU loss function, significantly reducing the target localization errors. Then, the monocular ranging model was constructed based on the improved object detection algorithm. The model established several regression equations based on the diagonal lengths of different detected objects and their real distances from camera, increasing the number of pixels per unit image, and enhancing the detection precision. Finally, the model used Kalman filtering to integrate vehicle motion data, achieving the dual-data fusion correction for ranging. [ **Result** ] The detection accuracy of improved algorithm increases by 0.68% compared with the original YOLOv5 with noticeable reduction in false positives. The average relative error of proposed algorithm is 3.67% within the range of 10–100 m, and 3.24% within the range of 30–100 m. The entire algorithm is accelerated by using NVIDIA's TensorRT. The inference speed for a single image reaches 12 ms. That is 5 times faster than that with the unaccelerated version, and 4 times faster than that with the binocular algorithm. The proposed method achieves optimal performance in ranging scenarios beyond 30 m. [ **Conclusion** ] The proposed algorithm can meet the ranging requirements for both intelligent and non-intelligent vehicles with general accuracy needs.

**Key words:** intelligent transport; monocular ranging; deep learning; attention mechanism; data regression

## 0 引言

高级驾驶员辅助系统(ADAS)已广泛应用于高端车辆,对前方目标物进行测距是ADAS的重要组成部分。实现目标物距离测量的方法有多种途径<sup>[1]</sup>,而使用单目摄像头进行测距是性价比最高的选择。相比使用毫米波雷达和双目摄像头等方案,单目测距具有价格低廉、计算开销小等特点。

大量学者与机构对单目测距技术进行了研究,取得了许多进展<sup>[2-5]</sup>。尽管不同测距模型的底层原理各不相同,但无论使用哪种方法,都需要在测距之前实现对测距目标的种类及位置信息的准确检测。现阶段对于交通目标物的检测方法有2条路线。一是基于传统的人工提取图像特征的识别方法,利用车辆边缘、颜色、阴影等特征进行识别。这种方法的运行速度虽快,但由人工提取交通目标物特征通常十分困难,且该方法面对复杂对象(如卡车、摩托车、自行车)时检测效果较差。二是基于深度学习的目标检测。随着深度神经网络技术的发展,基于卷积神经网络的检测在很多视觉检测的任务中都能取得不错的成果。然而,在自动驾驶领域中,无人驾驶对检测算法的运行速度有着极高的要求。由于深度学习的算法在检测任务的计算中通常需要花费大量时间,所以如何对网络模型进行优化以同时提高检测精度与速度是现在研究的主要方向。

现阶段的单目测距技术分为3类。一是利用相机成像的几何关系测距<sup>[6]</sup>,这种算法适用场景较广,

且在平坦路面环境下精度较高,但是测距精度严重依赖于汽车俯仰角的变化,当汽车进行俯仰运动时测距精度会产生较大偏差。二是利用相似三角形成像原理,将针孔摄像头中拍摄到的物体图像根据实际物体的长度进行测距,这样可以避免对相机俯仰角的依赖<sup>[7]</sup>。但是在实际应用中,由于该方法是对图片中物体所占像素点和预置的物体长度进行比例换算,因此在面对被测物体具有视差角度的情形时并不适用。三是基于数据回归的建模方法<sup>[8]</sup>。该方法通过非线性函数拟合距离曲线,隐含地消除了镜头畸变与俯仰角的误差,直接映射出被测物体的高度与距离的关系,但是测距精度依赖于数据集的精度,且当更换数据集时泛化能力较差。

为了解决现有车载单目测距中目标识别准确度不高测距误差较大,结果不稳定等问题。本研究提出了一种由深度学习与数据回归测距算法相结合的车载单目摄像头测距方法。该方法以YOLOv5为核心构建高精度目标检测框架,通过在YOLOv5主干网络中加入注意力机制,提升目标检测算法检测框的稳定性。同时,利用K-means++均值聚类算法对目标先验框进行分类,并利用E-IOU损失函数减少误报率,从而有效提升检测框输出数据的稳定性。此外,作为改良的数据回归测距方法,该方法通过采集目标物体的斜角长度和实际距离关系的数据集样本,建立非线性回归方程来预测目标物体的距离,有效解决现有测距方法泛化性差和精度差的问题。最终,将测距结果通过卡尔曼滤波融合算法与车辆

其他传感器采集到的车辆速度信息进行数据融合,进一步降低了误差和噪声,从而获得更加精确的测量结果。

## 1 基于YOLOv5算法的车辆目标检测方法

### 1.1 YOLOv5检测算法结构

YOLOv5检测算法包括 $S$ 、 $M$ 、 $L$ 、 $X$ 这4个不同尺度的模型,模型检测精度依次提高,但检测速度递减,对于不同的检测任务可以选择不同尺度的模

型<sup>[9]</sup>。以下仅以 $S$ 尺度的模型进行说明,即YOLOv5s网络模型结构,如图1所示。YOLOv5s模型主要由Input、Backbone、Neck、Output这4部分组成。

在YOLOv5s的Input端,采用Mosaic数据增强、自适应图像填充、自适应锚框对输入图像进行处理<sup>[10]</sup>。Mosaic技术通过图像处理方法(如裁剪、缩放等)将4张或更多图片拼接为一张图片,并对原始对象添加少量黑边,减少运算量,提高计算速度。

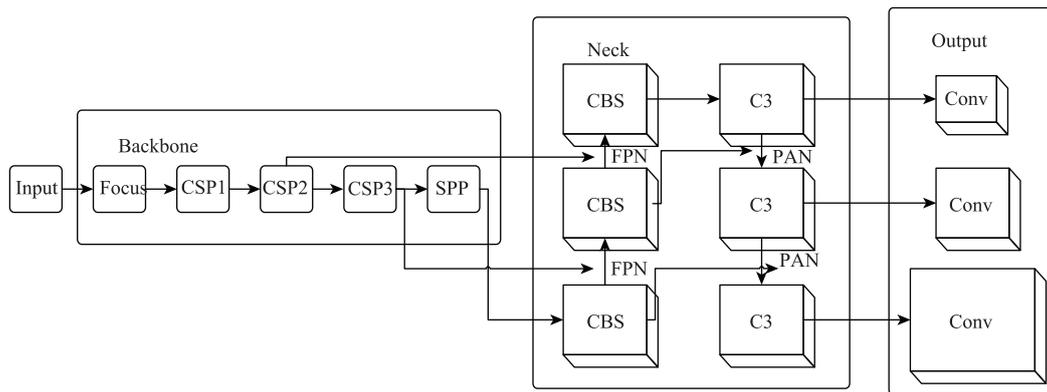


图1 YOLOv5s网络模型结构

Fig. 1 YOLOv5s network model structure

Backbone为CSPDarknet53主干网络,主要包含Focus、CSP、SPP结构。首先利用Focus层对输入图像进行切片操作,在不丢失图像信息的前提下大幅减少模型深度,提高计算运行速度。CSP网络结构由CBL模块、Resuit模块、Concat通道融合结构、Leaky\_ReLU组成。CBL模块由卷积模块(Conv)、批标准化模块(BN)、激活函数(Leaky\_ReLU)构成。SPP由大小不同的池化层组成,可以增大特征图的感受野,提高模型的特征提取效果。

Neck层由FPN和PAN组成,此层的作用主要是通过信息融合,将各个层次的信息进行整合与分配,进而实现信息的传递。

Output通过Conv卷积块输出不同尺度的检测结果,包括非极大值抑制(NMS)和损失函数。

### 1.2 基于K-means++的锚框机制

在目标检测中,先验框是预先在输入的图像上设置出大小形状各不相同的框来确定框中是否存在检测物,并确定实际框与检测框的交并比问题。由于YOLOv5s网络中原来的锚框是针对COCO数据集所设定的,并不完全适用于交通对象的检测,因此需要对YOLOv5网络中的锚框进行重新定位。相比

使用K-mean算法重定位锚框,选用K-means++聚类算法的优点是对聚类中心进行随机分配,这种随机生成初始聚类中心的方法确保了初始聚类中心的离散性,提升了锚框的生成效果。

使用可以随机选择样本中心的K-means++聚类方法针对交通目标物图像的锚框进行定位。当数据集样本数为 $D$ 、聚类中心数为7时,K-means++算法的聚类步骤为:

- (1) 从样本集 $X$ 随机选择1个样本点作为第1个聚类中心;
- (2) 计算其他样本点 $x$ 到最近的聚类中心的距离 $d(x)$ ;
- (3) 以概率 $\frac{d(x)^2}{\sum_{x \in X} d(x)^2}$ 选择1个新样本点 $c_i$ 加入聚类中心点集合中,其中距离值 $d(x)$ 越大,被选中的可能性越高;
- (4) 重复步骤(2)和步骤(3)选定7个聚类中心;
- (5) 基于这7个聚类中心进行 $k$ 均值运算。

### 1.3 基于注意力机制的优化

注意力机制最先被应用于自然语言处理(NLP)

中,可以使神经网络聚焦于那些更容易包含重要信息的位置进行学习,提升网络学习效果。YOLOv5 的网络模型中并没有注意力机制,因此本研究将 CBAM 注意力机制添加到 YOLOv5s 的主干网络 Backbone 中,C3 模块与 CBAM 注意力机制相结合,从而实现在 YOLOv5s 模型中加入注意力机制。

CBAM 注意力机制分别从通道和空间维度提供注意力地图<sup>[11]</sup>,主要分为通道注意模块(CAM)和空间注意模块(SAM),可以细化模型提取的特征,有效提高模型的分类效果。CBAM 注意力模块结构如图 2 所示。

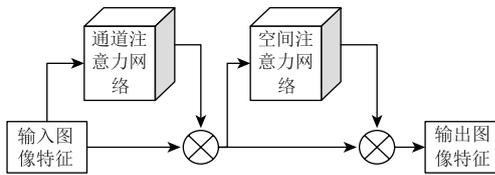


图 2 CBAM 模块结构

Fig. 2 CBAM structure

通道注意力模块(CAM)通过为输入图像的通道特征赋予不同权重,增强对不同图像特征的学习能力,可以在不改变通道维度的情况下压缩空间维度,其结构如图 3 所示。结构中包括并联的最大值池化与平均值池化,并将它们送入到多层感知机从而得到最后的特征图。

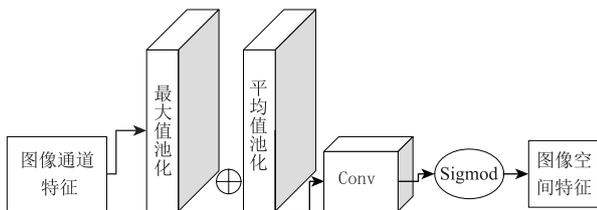


图 3 通道注意力模块结构

Fig. 3 Channel attention module structure

空间注意力模块(SAM)通过对输入图像的空间位置特征进行学习,从而提取出有助于提高图像分类能力的信息,可以在不改变空间维度的情况下压缩信道维度,其结构如图 4 所示。采用 2 个池化方法后将生成的特征图拼接起来,随后在拼接起的特征图上进行池化操作来生成最后的空间注意力特征。将 CAM 与 SAM 串联起来可以得到轻量级的注意力模块 CBAM。

#### 1.4 改进的损失函数

YOLOv5s 所采用的 C-IoU 损失函数考虑了边界框回归的重叠面积、中心点距离和纵横比,但是其

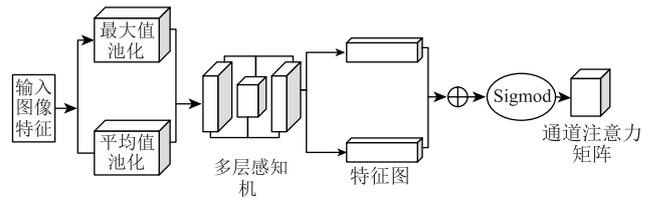


图 4 空间注意力模块结构

Fig. 4 Spatial attention module structure

公式中反映的是纵横比差异,而不是宽度和高度之间的实际差异。因此,本研究采用 E-IoU 方法替换 YOLOv5s 中所使用的 C-IoU 损失函数。E-IoU 在 C-IoU 的基础上分解了纵横比,并使其聚焦于高质量锚框,将其添加到 YOLOv5 框架中提高回归精度<sup>[12]</sup>。

E-IoU 的惩罚项基于 C-IoU 的处罚项来分割纵横比的影响因子,分别计算目标框和锚框的长度和宽度。E-IoU 由 3 部分组成,即  $L_{IoU}$  重叠损失,  $L_{dis}$  中心距离损失,  $L_{asp}$  宽度、高度损失。其主要的改进部分为宽度和高度损失,在此部分中对目标框和锚框的宽度和高度之间的差异进行了最小化,使收敛速度更快。惩罚公式为:

$$L_{EIoU} = L_{IoU} + L_{dis} + L_{asp} = 1 - I + \frac{\rho^2(b, b^{gt})}{c^2} + \frac{\rho^2(\omega, \omega^{gt})}{c_\omega^2} + \frac{\rho^2(h, h^{gt})}{c_h^2}, \quad (1)$$

式中,  $L_{EIoU}$  为损失函数;  $I$  为预测框和真实框的交并比;  $\omega^{gt}$  和  $h^{gt}$  为真实框的宽度和高度;  $c_\omega$  和  $c_h$  为覆盖 2 个框的包围框的最小宽度和高度;  $\omega$  和  $h$  为覆盖 2 个框的最小边界框的宽度和高度;  $b$  为预测框中心点坐标;  $b^{gt}$  为真实框中心点坐标;  $c$  为常数,取对角线长度的一半。

#### 1.5 目标检测算法的改进效果验证

本研究使用 KITTI 交通目标数据集作为算法测试的数据来源<sup>[13]</sup>,选择其中的 2D 目标检测数据集,共包含 7 481 张各种类型的标注图像。为了便于数据整理,本研究将原本数据集中的 Person\_sitting 标签合并到 Pedestrian 中,标签种类由 7 类合并为 6 类。据此将 KITTI 数据集划分为训练验证集 6 121 张,测试集 1 360 张。训练批 (Batch Size) 为 128,迭代次数为 150 次,预设学习率为 0.1。本研究选用所有类别目标的平均精度均值作为算法精度评价指标,分别训练 YOLOv5 原版算法与改进的 YOLOv5 算法。改进 YOLOv5 算法的平均精度均值变化曲线如图 5 所示。

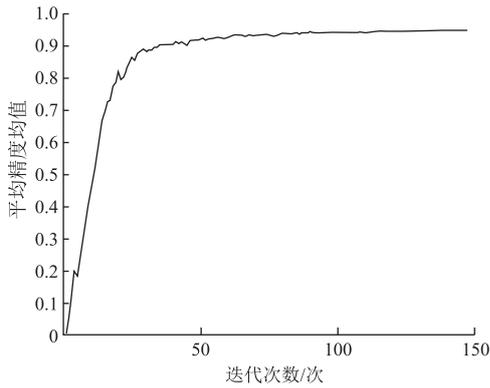


图5 改进 YOLOv5 算法的平均精度均值变化曲线

Fig. 5 Mean average precision variation curve of improved YOLOv5 algorithm

将本研究提出的算法与其他经典算法<sup>[14-15]</sup>在 KITTI 数据集下的精度对比(见表1),可见本研究提出的算法使检测精度达到了更高的水平。

表1 本研究提出的算法与其他经典算法在 KITTI 数据集下的精度对比

Table 1 Accuracy comparison between proposed algorithm and other classic algorithms on KITTI dataset

算法名称	平均精度均值
SSD512	79.70
Faster R-CNN <sup>[14]</sup>	76.90
YOLOv3 <sup>[15]</sup>	91.58
改进的 YOLOv3	93.61
YOLOv5	94.53
本研究提出的算法	95.04

## 2 改良的数据回归测距算法

数据回归测距算法是通过大量真实数据构建拟合函数来反映自变量(目标检测到的目标像素点位置关系)和因变量(距离目标的距离)的函数关系,以此来测量距离。

### 2.1 试验数据来源

构建数据回归算法所需要的数据集来源于 WIDC 算法数据集。该数据集利用超仿真虚拟驾驶平台所构架的虚拟仿真场景,包含各式车辆、红绿灯、交通标志、行人、自行车等共计约 15 070 余组图像及距离数据,保证了算法的实用性与真实可用性。数据集的类别与数量如表 2 所示。

### 2.2 基于斜角长度的图像回归

不同于前人采用的提取图像高度的做法<sup>[8]</sup>,本研究提取了图像斜角的长度作为参数观察函数映射关系。选用斜角长度作为映射参数而非选择图像高度的原因是为了提高像素,降低误差。在建立

表2 数据集的类别与数量(单位:组)

Table 2 Categories and quantities of dataset (unit: set)

类别	数量
轿车	2 226
行人	2 063
自行车	905
卡车	1 246
摩托车	1 706
红绿灯	1 554
各类交通标志牌	4 370
空标签	1 000
总计	15 070

像素的映射关系时,当检测框的精度不变,提取图片中更多的像素样本数作为参数将会使减少一定的误差。图像斜角的长度计算为:

$$H = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}, \quad (2)$$

式中  $x_1, x_2, y_1, y_2$  分别为 YOLOv5 检测模型输出的方框位置坐标。

真实距离与图像斜角长度的映射曲线如图 6 所示。根据所示曲线可以得出构建的非线性函数具有光滑的特性。为了推断该关系,使用多项式函数对其进行拟合。该关系曲线描述了交通标志牌的真实距离与图像像素斜角的对应关系,并且具有平整和圆滑的特征,因此采用多项式函数进行拟合是合理的。分别使用三次多项式、四次多项式、五次多项式对映射曲线进行拟合,对比出的试验结果如表 3 所示。

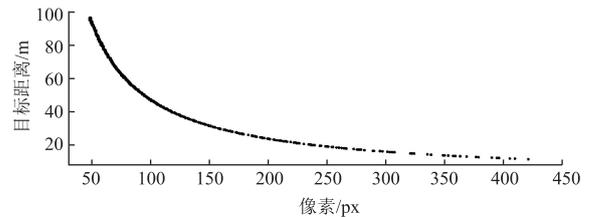


图6 真实距离与图像斜角长度的映射曲线

Fig. 6 Mapping curve of real distance and image oblique length

表3 不同多项式针对交通标志牌的测距拟合结果

Table 3 Ranging fitting results with different polynomials for traffic signs

真实距离/m	三次多项式	四次多项式	五次多项式
90	84.182	84.943	85.012
60	63.394	61.047	59.896
40	39.346	38.333	39.169
30	24.579	27.915	29.567
20	19.953	23.593	21.833
残差平方和	2 941.561	2624.108	2 609.742

分析拟合函数的精度,残差平方和是可以反映试验全部数据离散程度的一个指标,其值越小,则表明该模型的稳定程度越高。残差平方和计算为:

$$S_E = \sum_{i=1}^r \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2, \quad (3)$$

式中,  $r$  为组的数量;  $x_{ij}$  为多项式中的一个子项;  $\bar{x}_i$  为多项式的均值。外层表示对所有的组进行求和,内层表示对所有的观测值进行求和。

通过表(3)可以看出采用五次多项式所产生的误差最小,因此选用五次多项式对函数的映射关系进行拟合。对于数据集中不同类型的图像均给出拟合关系式(均保留3位有效数字)。式(4)~(7)分别为小型红绿灯、交通标志、行人和汽车的基于五次多项式的像距函数。建立模型的函数后,可以将目标检测得到的结果代入该函数,进行距离的估计。

$$f_1(z) = 4.98 \times 10^2 - 135z + 1.88 \times 10^{-1}z^2 - 1.43 \times 10^{-3}z^3 + 5.72 \times 10^{-6}z^4 - 8.79 \times 10^{-8}z^5, \quad (4)$$

$$f_2(z) = 2.26 \times 10^2 - 3.96z + 3.30 \times 10^{-2}z^2 - 1.40 \times 10^{-4}z^3 + 2.86 \times 10^{-7}z^4 - 2.27 \times 10^{-9}z^5, \quad (5)$$

$$f_3(z) = 1.29 \times 10^2 - 1.63z + 7.80 \times 10^{-3}z^2 + 5.65 \times 10^{-6}z^3 - 1.07 \times 10^{-7}z^4 + 2.20 \times 10^{-8}z^5, \quad (6)$$

$$f_4(z) = 1.12 \times 10^2 - 6.29 \times 10^{-1}z + 1.64 \times 10^{-3}z^2 - 2.15 \times 10^{-5}z^3 + 1.36 \times 10^{-7}z^4 - 3.32 \times 10^{-9}z^5, \quad (7)$$

式中  $z$  为目标检测获得的斜角线长度。

### 2.3 卡尔曼滤波优化测距数据

在图像测距任务中,由于目标距离物越远,单一像素所表达的距离值就越大,因此像素误差可能导致测距结果产生较大的误差。为了提高测距精度,采用卡尔曼滤波优化测距函数曲线,其原理是通过动态估计,将预测值和测量值加权融合,得到最优的状态估计值,从而生成更为准确的距离估计值。

首先,构建状态方程式(8)与测量方程式(9)对测距系统进行描述:

$$X(k+1) = \mathbf{A}X(k) + \mathbf{B}u(k+1) + \omega(k), \quad (8)$$

$$Z(k) = \mathbf{H}X(k) + v(k), \quad (9)$$

式中,  $X(k+1)$  为系统在  $k+1$  时刻的状态估计值;  $X(k)$  为系统在  $k$  时刻的状态估计值; 系统状态矩阵

$\mathbf{A}$  描述了系统状态从  $k$  时刻到  $k+1$  时刻的演变规律; 控制量  $u(k+1)$  和控制矩阵  $\mathbf{B}$  则描述了系统状态的外部控制情况; 过程噪声  $\omega(k)$  描述了系统状态的不确定性和外部干扰等随机因素;  $Z(k)$  为  $k$  时刻的真实测量值;  $\mathbf{H}$  为观测系统的测量矩阵;  $v(k)$  为了系统的测量误差。

卡尔曼滤波的应用分为2个阶段,第1个阶段为预测,利用式(10)对模型实现先验估计,利用式(11)得到当前系统的噪声协方差:

$$\hat{x}_t^- = \mathbf{F}\hat{x}_{t-1}^- + \mathbf{B}u_{t-1}, \quad (10)$$

$$P_t^- = \mathbf{F}P_{t-1}^-\mathbf{F}^T + \mathbf{Q}, \quad (11)$$

式中,  $\hat{x}_t^-$  和  $\hat{x}_{t-1}^-$  分别为当前时刻和前一时刻的最优估计;  $u_{t-1}$  为系统前一刻的系统控制量;  $\mathbf{F}$  和  $\mathbf{B}$  分别为从  $t-1$  时刻到  $t$  时刻的状态转移矩阵和状态转移矩阵系统控制矩阵;  $P_t^-$  和  $P_{t-1}^-$  分别为系统的先验估计协方差和上一时刻的最优估计值方差和;  $\mathbf{Q}$  为超参数,代表了过程噪声的方差。

卡尔曼滤波的第2个阶段为更新,当测距系统获得下一个阶段的观测数据时,利用式(12)完成对卡尔曼增益的更新:

$$K_t = P_t^- \mathbf{H}^T (\mathbf{H}P_t^- \mathbf{H}^T + R)^{-1}, \quad (12)$$

式中,  $K_t$  为  $t$  时刻的卡尔曼增益;  $R$  为测量误差协方差。最后根据式(13)得出  $t$  时刻的最优估计值,并根据式(14)更新协方差矩阵。

$$\hat{x}_t = \hat{x}_t^- + K_t(z_t - \mathbf{H}\hat{x}_t^-), \quad (13)$$

$$P_t = (\mathbf{I} - K_t \mathbf{H}) P_t^-, \quad (14)$$

式中  $z_t$  为观测系统观测值。

将行驶车辆的速度  $V$  引入到测距算法中。构建

系统的状态转移矩阵为  $\begin{bmatrix} 1 & t_0 \\ 0 & 1 \end{bmatrix}$ , 代入式(8), 得到式(15)。

$$X(k+1) = X(k) + V \times t_0, \quad (15)$$

式中  $t_0$  为了图片的采样间隔。

构建先验估计协方差。根据测试,基于斜角长度的数据回归测距算法的数值相对于真实距离数值误差约为10%。因此设定测量噪声值  $R$  为0.1,而超参数  $Q$  值同样设定为0.1,代入式(11)计算出卡尔曼增益系数为式(16)。

$$K_t = P_t^- (P_t^- + 0.1)^{-1}. \quad (16)$$

## 3 试验与结果分析

### 3.1 算法部署

本研究的算法训练平台选用 Intel(R) Xeon(R)

Gold 5320 CPU@2.20GHz 和 RTX4000 GPU, 并选用 Ubuntu 18.04 操作系统。验证平台选用 Core(TM) i7-12700 CPU 和 3060(laptop) GPU。

由于单目测距算法对检测速度的要求兼具高效与稳定。本研究在部署目标检测算法时利用 TensorRT 加速技术对改良的 YOLOv5s 进行了优化与重构。TensorRT 是高性能推理工具包, 可对主流深度学习框架 (Pytorch, TensorFlow, Caffe 等) 进行加速推理支持<sup>[16]</sup>。

本研究使用 TensorRT 技术在 2 个方面对算法进行优化重构。一方面是进行层间融合与张量融合。由于 YOLOv5s 网络中存在许多的卷积模块及张量, 可以将卷积模块中的多个步骤利用 TensorRT 融合一次性完成计算, 大幅减少了数据在网络中间的传输与计算的时间。另一方面是对数据精度进行校准, 由于在深度学习框架的部署阶段, 不需要进行反向传播, 所以可以适当降低数据精度, 将 FP32 类型的数据降低至 FP16 和 INT8 精度, 借此来大幅提升推理计算的速度。

为了更好体现出将目标检测算法部署到 TensorRT 中的性能优势, 将加速与未加速的 YOLOv5 算法的推理速度与其他学者提出的双目测距的车辆测距方法推理的速度<sup>[17]</sup>进行了比较。对比结果 (见表 4) 体现出了本研究提出算法速度的优越性, 单张图片推理时间为 12 ms, 相比于未加速的算法推理速度减少了 5 倍, 相比于双目算法推理速度提升了 4 倍。

表 4 单张图片推理速度对比

Table 4 Comparison of inference speed for single image

方法	帧数/(帧·s <sup>-1</sup> )	推理时间/ms
未加速算法	13.80	72.0
加速后算法	83.0	12.0
双目测距方法 <sup>[17]</sup>	19.92	50.18

### 3.2 仿真试验结果

测量交通标志距离时, 卡尔曼滤波优化后的测距效果与未经优化过的曲线及真实距离的曲线优化效果如图 7 所示。

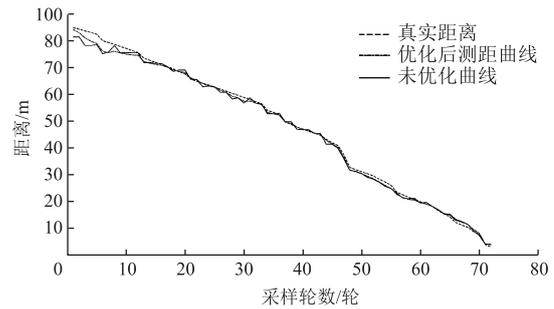


图 7 曲线优化效果

Fig. 7 Curve optimization effect

经过卡尔曼滤波优化后的算法相比未经优化的算法更加稳定, 更加平滑, 对于误差有着明显的抑制效果。试验检测场景均选择在平直的交通路口, 包括大型红绿灯、环道标志牌、人行横道线上的行人、道路旁边的卡车。

部分仿真场景下的测距结果如表 5 所示。结果

表 5 部分仿真场景下的测距结果

Table 5 Ranging results in partial simulation scenarios

测距/m	交通标志		红绿灯		行人		汽车	
	回归测距的 误差/%	改进回归测 距的误差/%	回归测距的 误差/%	改进回归测 距的误差/%	回归测距的 误差/%	改进回归测 距的误差/%	回归测距的 误差/%	改进回归测 距的误差/%
10	10.0	8.0	—	—	11.76	9.51	9.12	5.21
20	5.65	4.10	—	—	12.95	10.32	5.81	2.59
30	4.01	3.33	—	—	2.68	1.94	4.85	3.30
40	3.42	2.67	4.50	4.10	1.26	1.51	1.37	0.67
50	2.18	1.76	3.96	3.64	3.43	3.16	3.38	2.86
60	1.76	1.0	0.90	1.11	4.12	3.75	5.7	4.87
70	1.77	1.37	0.71	1.01	3.21	3.34	5.24	3.79
80	1.27	0.70	5.38	4.23	5.07	4.70	4.5	4.05
90	3.20	2.32	7.28	5.94	—	—	—	—
100	—	—	1.78	1.79	—	—	—	—
120	—	—	1.51	1.46	—	—	—	—
140	—	—	1.54	1.48	—	—	—	—
平均误差	3.70	2.80	3.06	2.02	5.56	4.77	4.99	3.41

表明,相较于传统的数据回归算法,本研究提出的改进测距算法精度具有明显的提升。在交通标志、红绿灯等具有统一规格尺寸的目标物中,使用普通的数据回归算法的测距方法可以取得相对较高的精度,平均误差为3.7%和3.06%。然而,对于个体数值差异较大的目标物(如行人、汽车等),单纯使用数据回归的测距方法通常会造成较大误差,平均误差高达5.56%和4.99%。造成这种情况的原因主要有2方面:一方面是行人与汽车等目标物因易于背景色混淆,造成目标检测算法定位出的检测框并不精确,从而造成测距误差;另一方面是由于交通标志牌与红绿灯尺寸大多统一,建立出的像距函数通用性高;而行人与汽车等尺寸差异较大,建立出的像距函数通用性相对较差。

本研究改进后的测距算法通过加入卡尔曼滤波进行数据优化,针对个体数值差异较大的目标物的误差得到了明显降低,平均误差分别为4.77%和3.41%。另外,以各个不同目标物的平均误差度为基线,并排除因仿真试验环境问题导致的误差的数据点,可以发现另一个现象:本研究所提出的算法在距离检测目标相对较近时,会产生测距范围的最大误差,但是当检测距离较远时,测距误差总体趋于减小。这一现象尤其在交通标志、红绿灯等具有统一规格尺寸的目标物上表现得非常明显。因此,本研究提出算法的准确测距范围应根据不同的目标物在10~140 m之间。

### 3.3 实车试验

除了在仿真环境中进行验证,本研究还在实车试验中进行了验证。在验证中,本研究选择对静态目标进行验证,与动态目标场景有一些差异。在静态场景下,被检测物体的位置与姿态相对固定,且不会发生移动,因此可以更加准确地获取其真实距离。而在动态场景下,被检测物体的位置、姿态、速度等参数会随着时间不断变化,这使得真实距离的获取更加复杂,容易受到多种因素的影响,例如目标的加速度、遮挡、动态物体的相互作用等。由于动态目标的测距效果偶然性过大,并不能很好地检测出算法的效果,且静态目标的检测在实际应用中占比很大,例如红绿灯、标志牌、及车辆行驶中的前方障碍物等都属于静态目标物,故本研究选用静态目标物作为测距对象。

本试验采用的设备为智能网联汽车,车辆图像来自于车内安装的车载摄像头,型号为MYNT EYE D1000-50/Color,目标真实距离数据使用车辆上方安

装的激光雷达及毫米波雷达进行辅助采集,试验整体流程如图8所示。

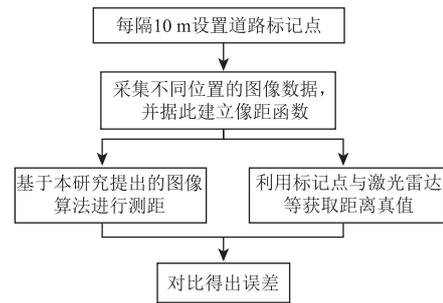


图8 实车试验流程

Fig. 8 Real vehicle test flow

试验场景选择室外平直封闭道路,光照条件良好。在开展试验前利用行车记录仪提供的视频采取一定时间间隔进行截取图像作为目标检测算法的训练数据,并依照2 m的间距选取待检目标距离与像素大小的比值。为验证测距算法的准确性,在道路上分别设置静态车辆和行人,以10 m为间隔在距离目标物10, 20, 30, 40, 50, 60, 70, 80, 90, 100 m的位置上设置标识点,以确认车辆距离检测目标的距离。除了上述目标物之外,同样在真实交通路口的位置对红绿灯与交通标志的距离进行建模测量,并以道路中清晰的车道线辅助判断红绿灯的距离信息。

车辆分别以20 km/h和40 km/h的速度向测量目标行驶。为保证试验车辆的安全性,在20 km/h的速度下,将测试区间设置为10~100 m,而在40 km/h的速度下,设置为30~100 m。另外,由于交通路口的位置的特殊性,对于红绿灯的距离采集截止为40 m。随后根据试验拍摄的视频选取5帧/s的采样间隔来验证算法。试验结果如表6所示。

表6 部分实车场景下的测距结果

Table 6 Ranging results in partial real vehicle scenarios

测距/ m	不同速度(km/h)下的误差/%							
	交通标志		红绿灯		行人		汽车	
	20	40	20	40	20	40	20	40
10	9.69	—	—	—	13.37	—	7.30	—
20	6.30	—	—	—	9.62	—	5.10	—
30	2.31	1.53	—	—	5.97	3.99	0.60	5.53
40	2.68	4.71	3.97	2.62	3.29	4.21	5.47	1.29
50	3.57	2.26	2.76	1.95	1.17	2.81	4.64	3.94
60	2.99	2.08	1.66	3.27	4.32	1.70	6.08	5.43
70	2.71	0.53	2.44	4.83	5.97	4.33	3.51	2.61
80	2.47	2.47	4.39	1.29	2.47	3.33	2.47	3.37
90	1.40	3.75	1.74	2.55	2.13	2.17	0.40	1.09
100	1.23	2.37	0.79	1.71	1.29	2.64	1.20	2.66
平均误差	3.54	2.46	2.22	2.29	4.96	3.15	3.67	3.24

通过比较实车环境下的测距算法数据和仿真试验数据(见表5和表6),发现测距算法的精度在实车环境下有所降低。这显然是由于实车环境下建立像距函数的精度不够所导致的,继续提高数据采集精度可以减小这种误差。

无论是实车还是仿真环境,都存在对近距离目标物测距误差过大的问题。这是由于目标检测物的视差角度对目标检测算法标记框的位置准确度产生影响,从而导致测距函数不能准确测出距离。距离越近,目标检测越容易受到视差角度的影响,进而影响测距精确度。

尽管本研究提出的算法在较近距离时的误差相对较大,但由于实际场景中车辆的速度相对较快,考虑到车辆制动问题,大多数场景对于测距准确距离的范围需求都在30 m以上,而这正处于本研究提出方法精确度较高的范围中。本研究提出算法对于真实车辆测距中在30 m以上的精度可以达到3.24%,而对标志牌与红绿灯的测距精度可以达到2.46%与2.29%,因此本研究提出的算法具有较高的工程实用价值。将本研究算法与文献[18]和文献[19]所提出的算法作比较,无论是在测量精度还是测距范围都表现出了明显的优势。

#### 4 结论

本研究提出的基于深度学习与数据回归的单目测距方法,相比于现有单目测距算法,在目标位置信息的准确度、测距稳定性、准确度、泛化性等方面均有显著的提升。经过试验验证,通过加入注意力机制、改进损失函数、优化锚框等方式改进YOLOv5算法,相比原算法可以提高检测精度约0.68%,而将数据回归算法与卡尔曼滤波算法结合使用可以将测距的准确度提高约30%。将整套算法部署到TensorRT图像推理加速引擎中,可以达到优秀的运行速度。对于前方车辆测距任务,在10~100 m范围内的测距误差为3.67%,而在30~100 m测距误差约为3.24%。本研究提出算法可以满足一般精度要求的智能车辆与非智能车辆的测距任务。

本研究发现在近距离的测距任务中精度会受视差角度影响从而降低。因此,在接下来的研究中,将着重研究存在视差角度影响下的智能车辆单目测距任务。

#### 参考文献:

#### References:

[1] 张照震. 基于单目视觉车距测量方法综述[J]. 汽车

实用技术, 2022, 47(4): 153-157.

ZHANG Zhaozhen. Review of vehicle distance measurement based on monocular vision [J]. *Automotive Applied Technology*, 2022, 47(4): 153-157.

[2] 余厚云, 张为公. 基于单目视觉的跟驰车辆车距测量方法[J]. *东南大学学报(自然科学版)*, 2012, 42(3): 542-546.

YU Houyun, ZHANG Weigong. Method of vehicle distance measurement for following car based on monocular vision [J]. *Journal of Southeast University (Natural Science Edition)*, 2012, 42(3): 542-546.

[3] TUOHY S, O' CUALAIN D, JONES E, et al. Distance determination for an automobile environment using inverse perspective mapping in OpenCV [C] // *Proceedings of IET Irish Signals and Systems Conference (ISSC 2010)*. London: IET, 2010: 100-105.

[4] MENG C C, BAO H, MA Y, et al. Visual meterstick: Preceding vehicle ranging using monocular vision based on the fitting method [J/OL]. *Symmetry*, 2019, 11(10): 1081. (2019-08-28) [2022-12-19]. <https://doi.org/10.3390/sym11091081>.

[5] TANG J, LI J. End-to-end monocular range estimation for forward collision warning [J/OL]. *Sensors*, 2020, 20(20): 5941. (2020-10-21) [2022-12-19]. <https://doi.org/10.3390/s20205941>.

[6] 高维岳, 陈宇拓, 刘洋, 等. 基于单目视觉的车辆前方障碍物测距方法[J]. *计算机工程与设计*, 2022, 43(4): 1016-1022.

GAO Weiye, CHEN Yutuo, LIU Yang, et al. Distance measurement method based on monocular vision for obstacles in front of vehicles [J]. *Computer Engineering and Design*, 2022, 43(4): 1016-1022.

[7] NAKAMURA K, ISHIGAKI K, OGATA T, et al. Real-time monocular ranging by Bayesian triangulation [C] // *Proceedings of 2013 IEEE Intelligent Vehicles Symposium (IV)*. New York: IEEE, 2013.

[8] 沈志熙, 黄席樾. 基于数据回归建模的单目视觉测距算法[J]. *计算机工程与应用*, 2007, 43(24): 15-18.

SHEN Zhixi, HUANG Xiyue. Monocular vision distance detection algorithm based on data regression modeling [J]. *Computer Engineering and Applications*, 2007, 43(24): 15-18.

[9] 何廷全, 俞山川, 张生鹏, 等. 基于目标检测的驾驶人分神驾驶行为检测方法研究[J]. *公路交通科技*, 2022(10): 153-161.

HE Tingquan, YU Shanchuan, ZHANG Shengpeng, et al. Study on driver's distracted driving behavior detection

- method based on object detection [J]. *Journal of Highway and Transportation Research and Development*, 2022, 39 (10): 153-161.
- [10] 马军, 刘大海, 薛梦婕, 等. 基于车辆跟踪的高速公路全天候违停检测算法 [J]. *公路交通科技*, 2024, 41 (11): 78-85.
- MA Jun, LIU Dahai, XUE Mengjie, et al. Expressway all-weather illegal parking detection algorithm based on vehicle tracking [J]. *Journal of Highway and Transportation Research and Development*, 2024, 41 (11): 78-85.
- [11] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module [C] // *Proceedings of the European Conference on Computer Vision (ECCV)*. Berlin: Springer, 2018: 3-19.
- [12] ZHANG Y F, REN W Q, ZHANG Z, et al. Focal and efficient IOU loss for accurate bounding box regression [J] *Neurocomputing*, 2022, 506: 146-157.
- [13] GEIGER A, LENZ P, URTASUNR, et al. Are we ready for autonomous driving? The KITTI vision benchmark suite [C] // *IEEE Conference on Computer Vision and Pattern Recognition*. New York: IEEE, 2012: 3354-3361.
- [14] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks [J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2017, 39 (6): 1137-1149.
- [15] REDMON J, FARHADI A. YOLOv3: An incremental improvement [J/OL]. *arXiv (2018-04-08)* [2022-10-19]. <https://doi.org/10.48550/arXiv.1804.02767>.
- [16] 郭奕裕, 周箩鱼. 安全帽佩戴检测网络模型的轻量化设计 [J]. *计算机工程*, 2023, 49 (4): 312-320.
- GUO Yiyu, ZHOU Luoyu. Lightweight design of safety helmet wearing detection network model [J]. *Computer Engineering*, 2023, 49 (4): 312-320.
- [17] 颜佳桂, 李宏胜, 任飞. 基于SSD和改进双目测距模型的车辆测距方法研究 [J]. *激光杂志*, 2020, 41 (11): 42-47.
- YAN Jiagui, LI Hongsheng, REN Fei. Research on vehicle ranging method based on SSD algorithm and improved binocular ranging model [J]. *Laser Journal*, 2020, 41 (11): 42-47.
- [18] 张云飞, 王桂丽, 周旭廷, 等. 基于单目视觉的测距算法 [J]. *计算机与数字工程*, 2020, 48 (2): 356-360.
- ZHANG Yunfei, WANG Guili, ZHOU Xuting, et al. A range algorithm based on monocular vision [J]. *Computer & Digital Engineering*, 2020, 48 (2): 356-360.
- [19] 张煜, 赵奉奎, 张涌. 基于单目视觉的智能车障碍物检测及测距算法研究 [J]. *智能计算机与应用*, 2022, 12 (4): 41-46, 53.
- ZHANG Yu, ZHAO Fengkui, ZHANG Yong. Research on obstacle detection and distance measurement for intelligent vehicle based on monocular vision [J]. *Intelligent Computer and Applications*, 2022, 12 (4): 41-46, 53.