第45卷第2期
2024年4月

河北科技大学学报
Journal of Hebei University of Science and Technology

Vol. 45, No. 2
Apr. 2024

文章编号:1008-1542(2024)02-0122-09

基于改进YOLOv5s和传感器融合的目标检测定位

郑宇宏,曾庆喜,冀徐芳,王荣琛,宋雨昕

(南京航空航天大学自动化学院,江苏南京 211106)

摘　要:针对无人车环境感知过程中相机无法提供道路目标的位置信息,激光雷达点云稀疏以致检测方面难以达到很好效果的问题,提出一种通过融合两者信息进行目标检测和定位的方法。采用深度学习中YOLOv5s算法进行目标检测,通过联合标定进行相机与激光雷达外参的获取以转换传感器之间的坐标,使雷达点云数据能投影到相机图像数据中,得到检测目标的位置信息,最后进行实车验证。结果表明,所提算法能在搭载TX2嵌入式计算平台的无人车自动驾驶平台上拥有27.2 Hz的检测速度,并且在一段时间的检测环境中保持12.50%的漏检率和35.32 m的最远识别距离以及0.18 m的平均定位精度。将激光雷达和相机融合,可实现嵌入式系统下的道路目标检测定位,为嵌入式平台下环境感知系统的搭建提供了参考。

关键词:传感器技术;深度学习;目标检测与定位;无人车环境感知;相机与LiDAR融合

中图分类号:TP242.6　　**文献标识码**:A　　**DOI**:10.7535/hbkd.2024yx02002

Target detection and localization based on improved YOLOv5s and sensor fusion

ZHENG Yuhong, ZENG Qingxi, JI Xufang, WANG Rongchen, SONG Yuxin

(College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu 211106, China)

Abstract: As two important sensors in the process of unmanned vehicle environment perception, the camera cannot provide the position information of the road target, and the LiDAR point cloud is sparse, which makes it difficult to achieve good results in detection, so that a method was proposed which fuses the information of the two sensors for target detection and localization. YOLOv5s algorithm in deep learning was adopted for target detection, and the external parameters of camera and LIDAR were acquired through joint calibration to convert the coordinates between the sensors, so that the radar point cloud data can be projected into the camera image data, and finally the position information of the detected target was obtained. The real vehicle experiments were conducted. The results show that the algorithm can achieve a detection speed of 27.2 Hz on the unmanned vehicle autopilot platform equipped with TX2 embedded computing platform, and maintain a leakage rate of

收稿日期:2023-08-01;修回日期:2023-12-05;责任编辑:冯民

基金项目:国家自然科学基金(51505221);南京航空航天大学研究与实践创新项目(xcxjh20220337)

第一作者简介:郑宇宏(1997—),男,湖南邵阳人,硕士研究生,主要从事机器人环境感知方面的研究。

通信作者:曾庆喜,副教授。E-mail:Jslyzqx@163.com

郑宇宏,曾庆喜,冀徐芳,等.基于改进YOLOv5s和传感器融合的目标检测定位[J].河北科技大学学报,2024,45(2):122-130.

ZHENG Yuhong, ZENG Qingxi, JI Xufang, et al. Target detection and localization based on improved YOLOv5s and sensor fusion[J]. Journal of Hebei University of Science and Technology, 2024, 45(2): 122-130.

12.50%, a maximum recognition distance of 35.32 m, and an average localization accuracy of 0.18 m over a period of time in the detection environment. The fusion of LiDAR and camera can achieve road target detection and localization in embedded system, providing a reference for the construction of environment perception systems on embedded platforms.

Keywords: sensor technology; deep learning; target detection and localization; unmanned vehicle environment perception; camera and LiDAR fusion

对无人车行驶过程中的障碍物(如行人)进行检测和定位是保障其行驶安全非常重要的一部分。相机、毫米波雷达、激光雷达等是无人驾驶领域最常用的感知传感器[1-2]，其中相机很容易受到光线、遮挡等环境因素的影响，而激光雷达在高线数下成本昂贵，低线数下点云稀疏[3]。如何结合不同传感器的优缺点，进行准确有效的检测定位，成为众多学者研究的一个重要方向。

随着大数据、神经网络等相关技术的发展，深度学习的概念首次由加拿大学者 HINTON 等提出[4]，由于其检测精度和速度都优于传统方法，成为当前的主流目标检测算法。而基于深度学习的目标检测算法可以分成 2 类：基于区域提取的 Two-stage 目标检测算法，以及直接进行位置回归的 One-stage 目标检测模型。Two-stage 目标检测算法有 R-CNN[5]，SPP-Net[6]，FastRCNN[7]，FasterRCNN[8]，Mask R-CNN[9] 等，One-stage 目标检测算法有 YOLO[10]，YOLOv2[11]，YOLOv3[12]，SSD 等。

基于激光雷达的方法在近距离情况下提供准确的深度信息，通过点云学习到几何空间特征，但是由于远距离情况下点云的稀疏特性，探测远距离目标、小目标和形状相似目标的性能还有待加强。相较于能够提供丰富外观特性但缺乏良好的深度估计信息的相机，激光雷达在三维任务中的表现显然是优越的。这些互补特性促使基于激光雷达和相机信息融合的高精度三维目标检测方法被越来越多的学者关注，旨在克服单一传感器易受环境因素影响、抗干扰能力差的问题，从而提升整个感知网络的可靠性、准确性。学者们开发了多种传感器融合方案，如 F-PointNet[13]，ConvNet[14]，MV3D[15]，AVOD[16]，MVFusion[17] 等，但是这些传感器方案在嵌入式系统中却因为计算量大面临实时性和检测性能不能兼备的问题。

1　神经网络算法设计

采用深度学习 YOLOv5s 模型作为目标检测模块。该算法采用 CSP-DarkNet53 结构作为主干，CSP-DarkNet 作为一种 CSPNet，通过将浅层的特征图一分为二，分别通过特征提取模块向后传播以及跨过阶级层次直接与特征提取模块进行合并，能一定程度上解决主干网络中信息重复的问题和缓解梯度消失的问题，有效减少模型 FLOPs 值和参数数量，加快推理速度。其架构示意图如图 1 所示。

其中 Neck 部分采用 FPN＋PAN 的结构思路进行多尺度信息融合，并使用 SPPF 替代 SPP 来提高检测速度和应对多尺度问题。在检测头中，采用 CIOU LOSS 作为 Bounding box 的损失函数。

注意力机制作为一种资源分配机制，核心思想是根据原有的数据找到其关联性，然后突出重要的特征。YOLOv5s 作为轻量化模型具有较少的参数量。为了更好地满足在低光照环境下对道路目标检测的要求，本文采用 SENet 模块以增强网络的表示能力[18]，其分别通过 Squeeze 操作和 Exciation 操作实现所述功能。

首先是 Squeeze，输入一个维度为 $W\times H\times C$ 的特征图，其中 W 是 width，H 是 height，C 是 channel，通过全局平均池化操作，获得一个 $1\times 1\times C$ 的向量，见式(1)。

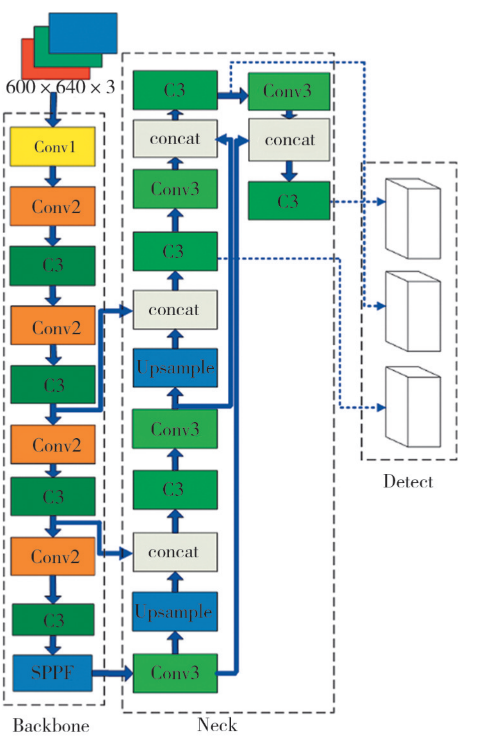

图 1　YOLOv5s 网络架构

Fig. 1　YOLOv5s network architecture

$$z=\boldsymbol{F}_{sq}(\boldsymbol{u}_C)=\frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}u(i,j)。\tag{1}$$

其次是激励(Exciation),它由 2 个全连接层组成。第 1 个全连接层有 $C\times S$ 个神经元,S 为一个缩放参数,输入为 $1\times1\times C$,输出为 $1\times1\times C\times S$;第 2 个全连接层有 C 个神经元,输入为 $1\times1\times C\times S$,输出为 $1\times1\times C$。最后通过 sigmoid 激活限制到[0,1]的范围得到权重矩阵,将权重矩阵和特征图的 C 个通道进行相乘计算以赋予权重,作为下一级的输入。其公式为

$$S=F_{ex}(Z)=\sigma(W_2\delta(W_1Z)),\tag{2}$$

式中:W_1 和 W_2 作为门控制机制,$W_1\in R^{\frac{c}{r}\times c}$,$W_2\in R^{c\times\frac{c}{r}}$;$\sigma$ 表示 sigmoid 函数;r 为一个固定参数。

YOLOv5s 主干特征提取网络采用 C3 结构,带来较大的参数量,当运用在嵌入式系统时,大量的参数量带来的将是更长的处理时间。本文将主干特征提取网络替换为更轻量的 Ghost 网络[19],以实现网络模型的轻量化,平衡速度和精度。Ghost Module 主要包括常规卷积、Ghost 生成和特征图拼接 3 步。首先用常规卷积 Y_{w*h*m} 得到本征特征图 $Y_{w'*h'*m}$:

$$Y'=X*f',\tag{3}$$

然后将 Y' 每一个通道的特征图 y'_i,用 Depthwise 卷积实现的线性操作 $\Phi_{i,j}$ 来产生 Ghost 特征图 y_{ij}:

$$y_{ij}=\Phi_{i,j}(y'_i),\quad \forall i=1,2,\cdots,m,\quad j=1,2,\cdots,s,\tag{4}$$

最后将第 1 步得到的本征特征图和第 2 步得到的 Ghost 特征图拼接(identity 连接)得到最终结果 Output。

由于 YOLOv5s 主干特征提取网络采用 C3 结构,其参数量较大,在嵌入式设备中运用较为局限,因此采用 Ghost 瓶颈结构对 YOLOv5s 网络中的 C3 模块进行改进,将 C3 模块中添加轻量化的 Ghost 瓶颈网络,这样就能有效减少计算参数数量,使得整个网络的速度和效率大大增加。最后修改得到的网络架构如图 2 所示。

模型训练采用公开数据集 KITTI[20]。从检测精度、速度、计算复杂度等方面对最新的检测算法和本文提出的检测模型进行比较。KITTI 数据集是由德国卡尔斯鲁厄理工学院和丰田工业大学芝加哥分校联合赞助的用于自动驾驶领域研究的数据集,包含市区、乡村和高速公路,采集过程是在德国的卡尔斯鲁厄进行的,检测目标包括 Car,Van,Truck,Tram,Pedestrian,Person_sitting,Cyclist 和 misc 8 类,本文将 Car,Van,Truck 和 Tram 合为一类 Car,将 Person 和 Person_sitting 合为一类 Pedestrian,忽略 Misc 类,最后检测 Car,Pedestrian 和 Cyclist 3 类,一共 7 481 个训练图像和 7 518 个测试图像。由于标签的格式和 YOLOv5s 模型的格式不一致,通过脚本将其转换成 YOLOv5s 所需要的.txt 格式以进行模型的训练。

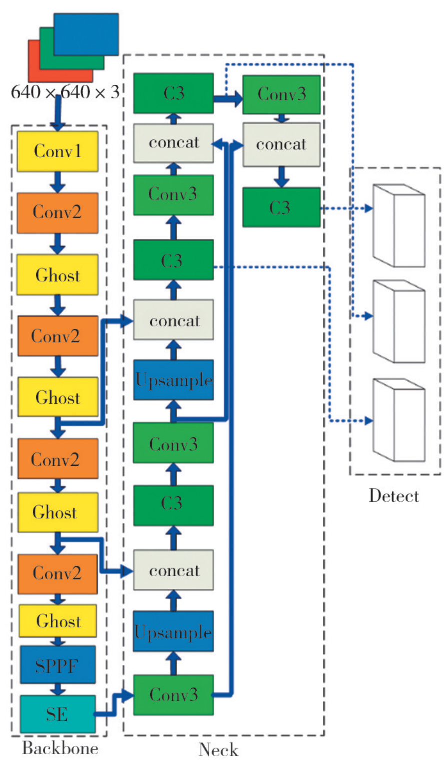

图 2 改进 YOLOv5s 网络架构

Fig. 2 Improved YOLOv5s network architecture

为了验证本实验所用模型的性能,选取精准率、召回率、平均精度和帧率作为模型的评价指标。精准度评价指标为

$$\text{Precision}=\frac{\text{TP}}{\text{TP}+\text{FP}},\tag{5}$$

式中:FP 为负类预测为正类;TP 为正类预测为正类。召回率评价指标为

$$\text{Recall}=\frac{\text{TP}}{\text{TP}+\text{FN}},\tag{6}$$

平均精度指标为

$$\text{mAP}=\int_0^1P(R)\mathrm{d}R,\tag{7}$$

对不同的模型设置相同的训练参数，将 Batchsize 设置成 16，初始学习率大小为 0.01，训练 50 个 epoch，训练和测试结果如表 1 所示。

表 1　目标检测模型对比

Tab. 1　Comparison of target detection models

模型	精准度	召回率	平均精度	帧率/Hz
YOLOv5s	0.944	0.884	0.914	13.0
SSD	0.837	0.745	0.778	9.2
本文模型	0.942	0.893	0.882	27.2

2　联合标定实验

2.1　联合标定模型

假设点云在以激光雷达为原点的坐标系下的具体坐标为$(X_l, Y_l, Z_l)^T$，在相机坐标系下空间坐标为$(X_c, Y_c, Z_c)^T$；$\boldsymbol{K}$ 是相机内参，$\boldsymbol{K} \in R^{3\times 3}$；$(u, v, 1)^T$ 是点云在像素平面上的投影点。

将点云由激光雷达坐标系变换到相机坐标系：

$$\begin{pmatrix} X_c \\ Y_c \\ Z_c \end{pmatrix} = \boldsymbol{R} \begin{pmatrix} X_l \\ Y_l \\ Z_l \end{pmatrix} + \boldsymbol{t} \ , \tag{8}$$

式中 $\boldsymbol{R}$ 和 $\boldsymbol{t}$ 分别为激光雷达和相机之间的旋转矩阵和平移向量，即激光雷达和相机的外参。再由相机坐标系变换到归一化平面坐标系，并投影到像素平面上：

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \boldsymbol{K} \begin{pmatrix} \dfrac{X_c}{Z_c} \\ \dfrac{Y_c}{Z_c} \\ 1 \end{pmatrix} 。 \tag{9}$$

根据图像大小判断投影点$(u, v, 1)^T$ 是否在图像上，保留投影在图像上的投影点并在图像上标出，建立图像投影点和立体点云的对应关系[21]。

利用机器人操作系统（ROS）进行数据通信，然后进行外参 $\boldsymbol{R}$，$\boldsymbol{t}$ 的调整，滤除距离远处的点云，就能找到点云投影点和图像像素的真实对应关系，使点云投影点和图像像素尽可能重合。如图 3 所示，$O_wX_wY_wZ_w$ 是标定板坐标系，$O_cX_cY_cZ_c$ 是相机坐标系，$O_lX_lY_lZ_l$ 为激光雷达坐标系。$\boldsymbol{P}_{c_i}$ 是相机坐标系下标定板的单位法向量，$\boldsymbol{P}_{l_i}$ 是激光雷达坐标系下标定板的单位法向量。$\boldsymbol{t}_{c_i}$ 是相机坐标系原点 O_c 到标定板坐标系原点 O_w 的平移向量，$\boldsymbol{t}_{l_i}$ 是激光雷达坐标系 O_l 到标定板坐标系原点 O_w 的平移向量[22]。

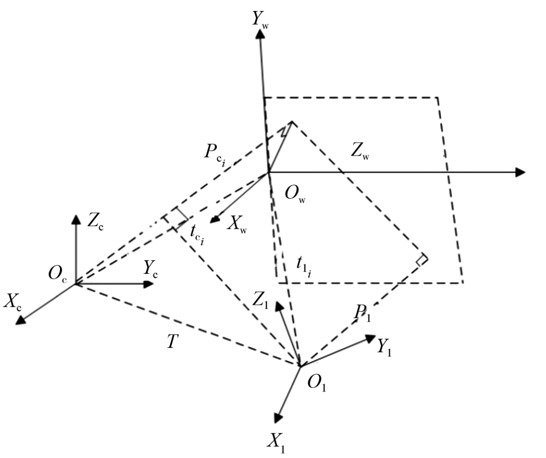

图 3　三坐标系模型

Fig. 3　Three coordinate system model

假设：

$$\boldsymbol{P}_c = [\boldsymbol{P}_{c_1} \quad \boldsymbol{P}_{c_2} \quad \cdots \quad \boldsymbol{P}_{c_{(n-1)}} \quad \boldsymbol{P}_{c_n}], \tag{10}$$

$$\boldsymbol{P}_l = [\boldsymbol{P}_{l_1} \quad \boldsymbol{P}_{l_2} \quad \cdots \quad \boldsymbol{P}_{l_{(n-1)}} \quad \boldsymbol{P}_{l_n}], \tag{11}$$

分别为 2 个坐标系下 n 个标定位置的 n 个法向量组成的矩阵，由夹角余弦之和最大可求得：

$$\boldsymbol{R} = \underset{R}{\operatorname{argmax}} \sum_i^n \boldsymbol{P}_{c_i}^T \boldsymbol{R} \boldsymbol{P}_{l_i} = \underset{R}{\operatorname{argmax}} \operatorname{tr}(\boldsymbol{R}\boldsymbol{P}_l \boldsymbol{P}_c^T), \tag{12}$$

解得 $\boldsymbol{R}' = \boldsymbol{V}\boldsymbol{U}^T$，其中 $\boldsymbol{V}$ 和 $\boldsymbol{U}$ 由矩阵奇异值分解 $\boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T = \boldsymbol{P}_l \boldsymbol{P}_c^T$ 得到。

由坐标系关系可知，激光雷达坐标系和相机坐标系的原点到标定平面的距离差为

$$\boldsymbol{P}_{c_i} \boldsymbol{T} = \| \boldsymbol{P}_{c_i} \boldsymbol{t}_{c_i} \| - \| \boldsymbol{P}_{l_i} \boldsymbol{t}_{l_i} \| 。 \tag{13}$$

对于 n 个标定位置，取目标函数为

$$\boldsymbol{T}=\min_{T}\sum\nolimits_{i=1}^{n}[\boldsymbol{P}_{c_i}^{\mathrm{T}}\boldsymbol{T}-(\|\boldsymbol{P}_{c_i}\boldsymbol{t}_{c_i}\|-\|\boldsymbol{P}_{l_i}\boldsymbol{t}_{l_i}\|)]^2, \tag{14}$$

其等价为求使 $\|\boldsymbol{P}_{c}^{\mathrm{T}}\boldsymbol{T}-(\boldsymbol{P}_{l}^{\mathrm{T}}\boldsymbol{T}_1-\boldsymbol{P}_{c}^{\mathrm{T}}\boldsymbol{T}_2)\|^2$ 最小的列向量 $\boldsymbol{T}$,其中 $\boldsymbol{T}_1=[\boldsymbol{t}_{c_1}\boldsymbol{t}_{c_2}\cdots\boldsymbol{t}_{c_n}]$,$\boldsymbol{T}_2=[\boldsymbol{t}_{l_1}\boldsymbol{t}_{l_2}\cdots\boldsymbol{t}_{l_n}]$,通过最小二乘法求解得到:

$$\boldsymbol{T}=(\boldsymbol{P}_{c}\boldsymbol{P}_{c}^{\mathrm{T}})^{-1}\boldsymbol{P}_{c}(\boldsymbol{P}_{l}^{\mathrm{T}}\boldsymbol{T}_1-\boldsymbol{P}_{c}^{\mathrm{T}}\boldsymbol{T}_2)。 \tag{15}$$

可知若进行联合标定,需要固定激光雷达和相机之间的位置关系,采集足够多点的数据进行计算。

2.2　相机内参标定

相机的内参标定首先要准备一块 9 m×6 m×0.024 m 的标定板,借助 OpenCV 的相机标定工具完成内参标定[23],图 4 为标定界面。

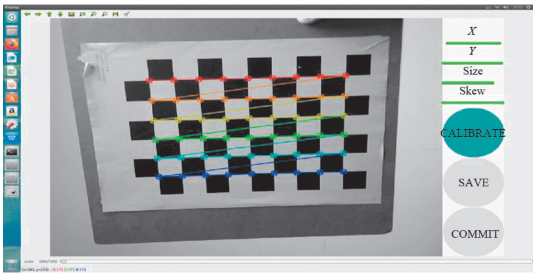

图 4　相机内参标定

Fig. 4　Camera internal reference calibration

通过改变标定板在画面中的位置、远近、角度等,收集到足够多的数据进行标定,输出一个.yaml 文件,其中就包含相机的内参矩阵 $\boldsymbol{R}$ 和畸变系数向量 $\boldsymbol{t}$。相机的内参包含 4 个参数,即 f_x,f_y,u_0 和 v_0。其中 f_x 也就是 f/d_x,f_y 也就是 f/d_y,f 代表的是焦距,d_x 和 d_y 代表的像元密度,也就是在像素坐标系下每一个像素分别在 x 和 y 方向分别占用了多少个单位,f_x 和 f_y 代表了焦距 f 在 x 方向和 y 方向上的投影,u_0 和 v_0 为感光板在坐标系 x 方向和 y 方向下的坐标。这 4 个参数就构成了相机内参,当经过标定获得了这些参数就能够把模型完整的表达出来。

内参标定结果如下:

$$\boldsymbol{R}=\begin{bmatrix}502.889\,76 & 0 & 303.387\,44\\ 0 & 504.454\,30 & 249.910\,56\\ 0 & 0 & 1\end{bmatrix},\quad \boldsymbol{t}=\begin{pmatrix}0.082\,64\\ -0.100\,47\\ 0.000\,71\\ -0.000\,61\\ 0\end{pmatrix}^{\mathrm{T}}。$$

2.3　相机和激光雷达联合标定

激光雷达配置成功后进行相机与激光雷达的联合标定,激光雷达和相机安装位置如图 5 所示。

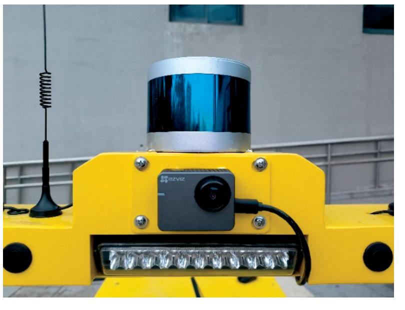

图 5　激光雷达和相机安装位置

Fig. 5　Installation location of LiDAR and camera

然后用相机和激光雷达进行录包操作,手持标定板在不同的位置停留 5 s 左右,在不同的位置都采集一定数量的数据,包括近点靠左、近点靠中、近点靠右、远点靠左等。在不同的位置时,手持标定板的倾斜幅度也需要进行调整,以便获得更多的数据。

包录制完成后,打开 Autoware 的 Camera Calibration Tools,通过 ROS 进行传感器数据的通信,然后先导入上文中标定获得的内参文件,然后对实现录制好的包进行播放,先调整点云显示区的角度、亮度等使后续标定板能够较清晰地呈现,对包中合适的图片和点云进行 Grab 收集,然后对每一张点云图进行手动标记,最后点击右上角 Calibrate 键获得标定结果。由于标定板较小,16 线雷达点云密度相对稀疏,因而需要多次进行标定以确保标定的准确性,标定界面如图 6 所示。点击右上角的 Project 键,外参文件即被输出到 home 文件夹下[24]。

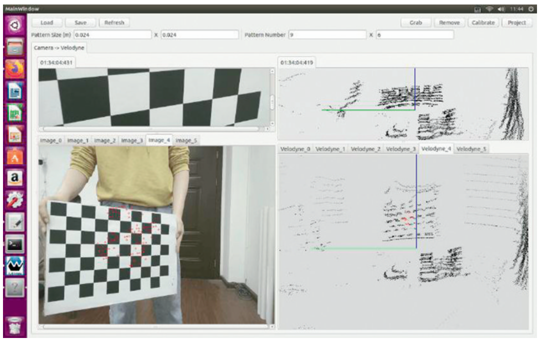

图 6　联合和标定界面

Fig. 6　Joint calibration interface

计算得到外参矩阵如下:

$$\boldsymbol{R}=\begin{bmatrix}499.63744 & 0 & 327.01053\\ 0 & 500.10249 & 235.77021\\ 0 & 0 & 1\end{bmatrix},\quad \boldsymbol{t}=\begin{pmatrix}0.10419\\ -0.20482\\ -0.00225\\ -0.00151\\ 0.16447\end{pmatrix}^{\mathrm{T}}。$$

2.4　传感器时间同步

不同种类的传感器运行频率不相同，扫描频率和信号的传递速度也存在偏差，因此就需要数据帧同步的方式对数据进行对齐。常见的数据帧同步方法主要包括硬件同步方式以及软件同步方式[25]，本文采用的相机和激光雷达的融合方式并不存在能够同步触发的硬件同步方式，因此采用软件同步的方式。

本文采用 ROS 机器人操作系统来对不同传感器进行数据帧同步[26]，使用的是 message_filters::Synchronizer 时间同步器，通过设置阀门，只有当在同一时间戳同时接收到 2 个传感器的数据时才控制消息的输出，以达到数据对齐的目的。其工作原理大致是：首先相机和激光雷达分别发布话题/image_raw 和/point_raw，然后 message_filters::Synchronizer 时间同步器程序订阅 2 个话题，在回调函数中对同步结果进行输出，当时间同步对齐时，继续进行点云和图像数据的处理，否则就继续等待时间同步的话题数据到来。而判断 2 个传感器的数据是否对齐则是根据 2 个传感器的时间戳的差值是否小于某一阈值。本文所用的激光雷达发布话题的频率是 10 Hz，相机发布话题的频率是 30 Hz，当 2 个话题的时间戳差值在 0.1 s 以内时就可以将 2 个数据同步输出，认为其已经同步对齐，当时间戳的差值大于 0.1 s 或者有 1 个传感器没有数据传输到时间同步器的时候，都不能正常触发回调函数将数据输出，如图 7 所示。

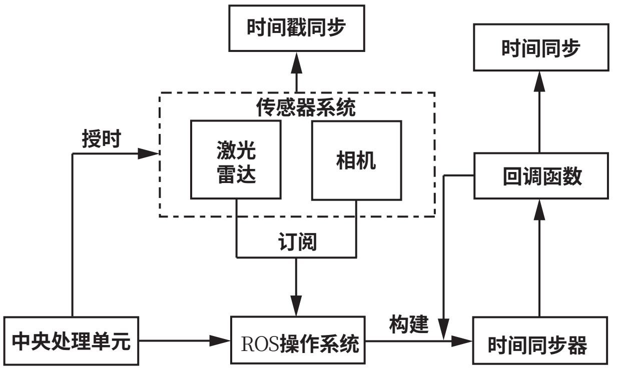

图 7　传感器同步系统框图

Fig. 7　Sensor synchronization system diagram

3　试验验证

无人车试验平台在设计上分为底层和上层 2 个部分。底层由 STM32 单片机作为主控制器，进行车体的运动控制；上层由 NVIDA Jetson TX2 嵌入式开发板控制，分别接入单目相机、激光雷达。

激光雷达选用镭神的 16 线激光雷达，相机采用感光元件为 200 万 1/2.8CMOS 的 60 帧 1080p 驰图运动相机摄像头，在检测的时候，会将相机实际拍摄获得的图像通过自适应图像缩放的方式变成 640×640 大小的图像送入网络模型中进行处理。试验平台为自主研发的智能自动驾驶无人车，如图 8 所示。

图 8　智能自动驾驶无人车平台

Fig. 8　Intelligent autonomous driving unmanned vehicle platform

所用软件平台为基于 ARM 架构的 ubuntu18.04 系统，算法框架主要在 ROS melodic 系统中开发运用，GPU 为 NVIDIAPascalTM，256 CUDA cores，CPU 为 dual-core NVIDIA Denver2 + quad-core ARM Cortex-A57，TensorRT 版本为 8.2.1，Cuda 版本为 10.2，系统内核为 LinuxUbuntu 4.9.253-tegra。Jetson TX2 运行 Autoware 自动驾驶框架、Pytorch 深度学习框架、ROS 机器人操作系统，分别负责传感器之间的标定配准和融合、深度学习模型的部署和训练，不同传感器和模块之间的数据通信。上层主要负责图像和激光数据的采集处理以及算法的实现，底层则主要负责无人车的运动控制，如图 9 所示。

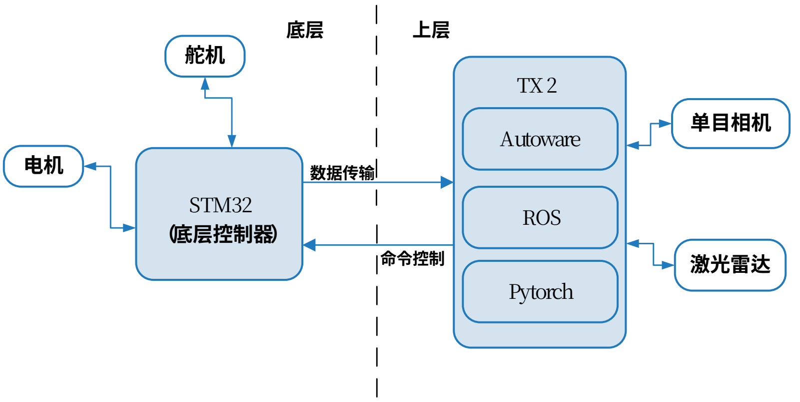

图 9 无人车系统软件平台

Fig. 9 Unmanned vehicle system software platform

将激光雷达和运动相机安装到阿克曼无人车上,并采用 TTL 转 USB 的转接线将无人车控制器和 TX2 相连接,在前方出现行人或者障碍物时无人车能顺利地进行避障或者制动。无人车沿着学校道路进行驾驶,运行过程中,激光雷达不断采集周围物体的点云信息并通过预处理进行特征采集,相机则是通过 YOLOv5s 神经网络对前方采集到的图像进行道路目标检测,对周围的目标进行检测以及定位,如图 10—图 12 所示。

图 10 图像采集数据

Fig. 10 Image data acquisition

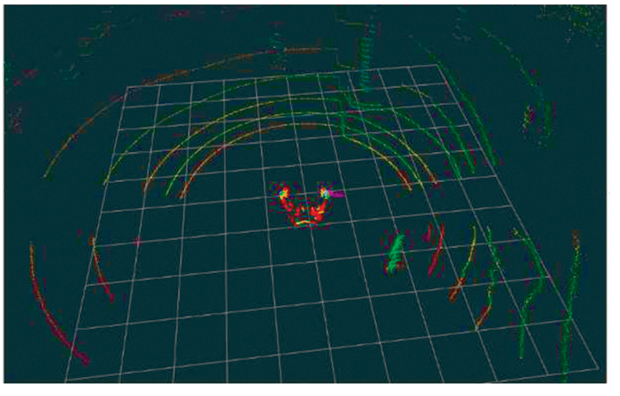

图 11 激光雷达采集数据

Fig. 11 Lidar data acquisition

对于采集到的数据包也分别采用 YOLOv5s, CenterPoint 和 Autoware 中的 LiDAR+SSD 算法进行处理,漏检率以图像的框选率为基准。当一个目标在 15 m 内还没有被正确检出,则判定其为漏检,用漏检的目标数量比总目标数量,得出的结果就为漏检率。最远检测距离为能检测到的最远的行人的距离,平均定位精度为所有目标被检出后激光雷达对其进行定位距离和其物体质心的定位距离差值的平均值。对一段时间内学校校内门口处和天桥底下路过的行人,车辆、自行车进行检测和定位,定位真值以事先用激光测距仪的测量数据为准,具体结果如表 2 所示。

图 12 目标检测定位结果

Fig. 12 Target detection and positioning results

表 2 道路目标检测定位结果

Tab. 2 Road target detection and positioning results

模型	漏检率/%	最远检测距离/m	平均定位精度/m
YOLOv5s	16.85	32.87	0
CenterPoint	25.31	15.28	0.35
LiDAR+SSD	31.43	20.11	0.49
本文模型	12.50	35.32	0.18

其中影响最远检测距离的主要因素是对小目标的检测性能，对小目标检测性能好的检测器在这个部分的表现性能较好。由于本文使用的激光雷达为 16 线，点云比较稀疏，所以 CenterPoint 点云检测算法表现性能较差。影响平均定位精度的主要是检测框对检测物体的拟合性能，欧氏聚类算法在这个部分表现性能差。

4 结　语

本文在 Jetson TX2 嵌入式系统中，采用激光雷达和相机融合技术和基于深度学习 YOLOv5s 的目标检测算法，实现了目标检测和定位系统的设计和开发。通过实车试验实现了道路目标的检测和定位，结果显示在一段时间的检测环境中实现了 12.50％的漏检率，最远识别距离达到了 35.32 m。

本文采用的传感器融合系统考虑到嵌入式系统计算能力有限，主要是进行决策级的融合，没有对异构传感器的原始数据进行融合处理，未来还需对融合算法进行进一步的改进，使得算法能够更好地利用不同传感器的优点，提高算法的鲁棒性。

参考文献/References：

[1] 蒋婷. 无人驾驶传感器系统的发展现状及未来展望[J]. 中国设备工程，2018(21)：180-181.
JIANG Ting. Development status and future prospects of autonomous driving sensor systems[J]. China Plant Engineering，2018(21)：180-181.

[2] 余世全. 无人驾驶汽车的发展现状和相关建议[J]. 专用汽车，2023(1)：11-14.
YU Shiquan. The development status of driverless vehicles and related suggestions[J]. Special Purpose Vehicle，2023(1)：11-14.

[3] 张保，张安思，梁国强，等. 激光雷达室内定位技术研究及应用综述[J]. 激光杂志，2023，44(3)：1-9.
ZHANG Bao，ZHANG Ansi，LIANG Guoqiang，et al. A review of LIDAR indoor positioning technology research and application[J]. Laser Journal，2023，44(3)：1-9.

[4] HINTON G E，OSINDERO S，TEH Y W. A fast learning algorithm for deep belief Nets[J]. Neural Computation，2006，18(7)：1527-1554.

[5] GIRSHICK R，DONAHUE J，DARRELL T，et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus：IEEE，2014：580-587.

[6] MSONDA P，UYMAZ S A，KARAAĞAÇ S S. Spatial pyramid pooling in deep convolutional networks for automatic tuberculosis diagnosis[J]. Traitement du Signal，2020，37(6)：1075-1084.

[7] GIRSHICK R. Fast R-CNN[C]//2015 IEEE International Conference on Computer Vision (ICCV). Santiago：IEEE，2015：1440-1448.

[8] REN Shaoqing，HE Kaiming，GIRSHICK R，et al. Faster R-CNN：Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2017，39(6)：1137-1149.

[9] HE Kaiming，GKIOXARI G，DOLLÁR P，et al. Mask R-CNN[C]//2017 IEEE International Conference on Computer Vision (ICCV). Venice：IEEE，2017：2980-2988.

[10] REDMON J，DIVVALA S，GIRSHICK R，et al. You only look once：Unified，real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas：IEEE，2016：779-788.

[11] REDMON J，FARHADI A. YOLO9000：Better，faster，stronger[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu：IEEE，2017：6517-6525.

[12] REDMON J，FARHADI A. YOLOv3：An Incremental Improvement[EB/OL]. https://export.arxiv.org/abs/1804.02767，2018-04-08.

[13] QI C R，LIU Wei，WU Chenxia，et al. Frustum PointNets for 3D object detection from RGB-D data[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City：IEEE，2018：918-927.

[14] WANG Zhixin，JIA Kui. Frustum ConvNet：Sliding frustums to aggregate local point-wise features for amodal 3D object detection[C]//2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Macau：IEEE，2019：1742-1749.

[15] CHEN Xiaozhi，MA Huimin，WAN Ji，et al. Multi-view 3D object detection network for autonomous driving[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu：IEEE，2017：6526-6534.

[16] KU J，MOZIFIAN M，LEE J，et al. Joint 3D proposal generation and object detection from view aggregation[C]//2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Madrid：IEEE，2018：1-8.

[17] FADADU S，PANDEY S，HEGDE D，et al. Multi-view fusion of sensor data for improved perception and prediction in autonomous driving[C]//2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Waikoloa：IEEE，2022：3292-3300.

[18] 王玲敏，段军，辛立伟. 引入注意力机制的 YOLOv5 安全帽佩戴检测方法[J]. 计算机工程与应用，2022，58(9)：303-312.
WANG Lingmin，DUAN Jun，XIN Liwei. YOLOv5 helmet wear detection method with introduction of attention mechanism[J]. Computer Engineering and Applications，2022，58(9)：303-312.

[19] 李北明，金荣璐，徐召飞，等. 基于特征蒸馏的改进 Ghost-YOLOv5 红外目标检测算法[J]. 郑州大学学报(工学版)，2022，43(1)：20-26.

LI Beiming,JIN Ronglu,XU Zhaofei,et al. An improved Ghost-YOLOv5 infrared target detection algorithm based on feature distillation [J]. Journal of Zhengzhou University(Engineering Science),2022,43(1):20-26.

[20] 王麒. 基于深度学习的自动驾驶感知算法[D]. 杭州:浙江大学,2022.
WANG Qi. Deep Learning-Based Algorithms for Autonomous Driving Perception[D]. Hangzhou:Zhejiang University,2022.

[21] 常昕,陈晓冬,张佳琛,等. 基于激光雷达和相机信息融合的目标检测及跟踪[J]. 光电工程,2019,46(7):85-95.
CHANG Xin,CHEN Xiaodong,ZHANG Jiachen,et al. An object detection and tracking algorithm based on LiDAR and camera information fusion[J]. Opto-Electronic Engineering,2019,46(7):85-95.

[22] 王俊. 无人驾驶车辆环境感知系统关键技术研究[D]. 合肥:中国科学技术大学,2016.
WANG Jun. Research on Key Technologies of Environment Awareness System for Driverless Vehicles[D]. Hefei:University of Science and Technology of China,2016.

[23] 李莉. OpenCV 耦合改进张正友算法的相机标定算法[J]. 轻工机械,2015,33(4):60-63.
LI Li. Camera calibration algorithm based on OpenCV and improved Zhang Zhengyou algorithm[J]. Light Industry Machinery,2015,33(4):60-63.

[24] 欧阳毅. 基于激光雷达与视觉融合的环境感知与自主定位系统[D]. 哈尔滨:哈尔滨工业大学,2019.
OUYANG Yi. The Environment Awareness and Autonomous Positing System Based on LiDAR and Vision[D]. Harbin:Harbin Institute of Technology,2019.

[25] 陈德江,江灏,高永彬,等. 基于 16 线激光雷达和相机融合的三维目标检测[J]. 智能计算机与应用,2023,13(3):117-122.
CHEN Dejiang,JIANG Hao,GAO Yongbin,et al. 3D target detection based on fusion of 16-line LiDAR and camera[J]. Intelligent Computer and Applications,2023,13(3):117-122.

[26] 曾彪. 基于多传感器信息融合的机器人建图研究与实现[D]. 成都:电子科技大学,2021.
ZENG Biao. Research and Implementation of Robot Mapping Based on Multi-sensor Information Fusion[D]. Chengdu:University of Electronic Science and Technology of China,2021.