第 59 卷 第 4 期/2022 年 2 月/激光与光电子学进展 研究论文

激光与光电子学进展 先进成像

基于改进PFPN的语义地图构建方法研究

任丽军[1]，刘元盛[2*]，钟科娣[1]

[1]北京联合大学北京市信息服务工程重点实验室，北京 100101；

[2]北京市智能机械创新设计服务工程技术研究中心，北京 100101

摘要 针对园区等环境结构性强而全球导航卫星系统(GNSS)信号不稳定的应用场景及无人车应用激光雷达同时定位与建图(SLAM)技术缺乏对于场景的语义理解能力而造成的定位误差问题，提出了一种单目相机与激光雷达融合构建三维语义地图的方法。该方法以园区环境结构性强、车辆行人动态变化高的特征为依据，通过改进的全景特征金字塔网络(PFPN)进行场景视觉语义分割，然后采用像素级融合方法为激光点云提供语义信息，从而有效地去除了正态分布变换(NDT)建图过程中动态目标的干扰，进而提高了无人车SLAM技术在动态环境中的鲁棒性和精确性。在旋风智能无人驾驶平台上进行实验验证，并与原始NDT方法进行了比较，实验结果表明：所提方法能够全面提升建图精度，其中位姿精度最为显著，提高了34.34%；除此之外，建图的点云数量也降低了39.78%，大大提高了建图速度。

关键词 遥感；激光雷达；语义地图，语义分割，数据融合

中图分类号 **文献标志码** **doi**：10.3788/LOP202259.0428002

Building Method of Semantic Map Based on Improved PFPN

Ren Lijun[1], Liu Yuansheng[2*], Zhong Kedi[1]

[1]*Beijing Key Laboratory of Information Service Engineering, Beijing Union University, Beijing* 100101, *China*;

[2]*Beijing Engineering Research Center of Smart Mechanical Innovation Design Service, Beijing* 100101, *China*

Abstract Aiming at the parks and other similar environments application scenarios where the scene feature is unique and the global navigation satellite system (GNSS) signal is unstable, as well as the problem that the lack of semantic understanding in Lidar simultaneous localization and mapping (SLAM) results in a localization error of unmanned vehicle, a building method of three-dimensional semantic map with the data fusion of monocular camera and Lidar is proposed. This method is based on the characteristics of strong structural park environment and high dynamic variation of pedestrian in vehicles. The improved panoramic feature pyramid network (PFPN) is used for scene visual semantic segmentation, and then the pixel level fusion method is used to provide semantic information for the laser point cloud, so as to effectively remove the interference of dynamic targets in the process of normal distribution transformation (NDT) mapping, and then improve the robustness and accuracy of unmanned vehicle SLAM technology in the dynamic environment. Experimental validation is carried out on the cyclone intelligent self-driving platform and compared with the original NDT method, the experimental results show that the proposed method is able to improve the construction accuracy comprehensively, with the most significant improvement of 34.34% in positional accuracy; besides that, the number of point clouds for building the diagram is also reduced by 39.78%, which greatly improves the construction speed.

Key words remote sensing; Lidar; semantic map; semantic segmentation; data fusion

收稿日期：2021-03-08；**修回日期**：2021-03-30；**录用日期**：2021-04-02

通信作者：*yuansheng@buu.edu.cn

1 引　　言

随着无人驾驶技术的兴起与计算机视觉领域的发展，蕴含丰富信息的三维场景地图为自主导航、环境勘探等应用提供了更多的可能性。传统地图仅仅包含空间几何信息，使得机器人在执行高级任务或要更好地理解周围场景含义时存在限制。为此，科研人员开展了大量的研究工作，其中语义地图的研究是自主机器人[1]及无人车等各种机器人应用中的一个重要领域。语义地图意味着将真实环境重构到空间里，并在地图中嵌入语义信息。

语义分割是生成语义地图的一个基本步骤。根据算法所处的空间领域，关于语义分割的研究可以分为二维(2D)语义分割和三维(3D)语义分割。2D语义分割为图像中的每个像素赋予语义标签。自出现以来，该技术得到了广大研究人员的关注。文献[2-3]中使用马尔可夫随机场(MRF)和条件随机场(CRF)等概率模型进行语义分割。另外，随着卷积神经网络(CNN)的发展，已有多项研究利用CNN解决语义分割问题，并取得了显著的性能提升。Long等[4]提出了一个有代表性的全卷积网络(FCN)，该网络通过引入一种转置卷积层的上采样层，可以在保留原始输入图像空间信息的同时进行逐像素分类。在此基础上，ENET[5]、SegNet[6]、DilatedNet[7]和RefineNet[8]等几种优秀的2D语义分割网络体系结构也被开发出来。不同于2D语义分割，3D语义分割直接将点云作为输入，并为每个点分配语义标签。文献[9-11]的研究集中在使用同时定位与建图(SLAM)重建几何信息，而且已经应用于各种机器人中，但这种使用纯点云信息所建的地图不带有语义信息。近年来，一些研究人员改进了2DCNN的结构，并将其应用于3D语义分割，如PointNet[12]和SEGCloud[13]。文献[14]中将点云几何多特征与CNN结合起来进行端到端的语义分割。然而，这些方法都是基于高线束激光雷达的，需要大量的计算时间。文献[15]使用YOLOv3结合视觉SLAM构建语义地图，但是时效性不能够满足无人驾驶的需求。

文献[16]中提出的网络模型大多被用于解决室内场景的语义分割，鲜有涉及室外场景的分割。故而在众多的2D语义分割网络体系结构中，本文选择文献[17]中的一种简单、灵活、有效的全景特征金字塔网络(PFPN)结构，它可以同时使用基于区域的输出和密集像素输出的单个网络来提高实例分割和语义分割的准确性。针对园区场景特点，利用改进的PFPN作为语义分割的基本方法，对视觉语义分割和16线激光雷达点云进行数据融合，使得激光雷达点云获得语义信息，最后通过正态分布变换(NDT)方法进行3D语义地图的建立。

2 语义地图构建方法

所提语义地图构建整体框架如图1所示，包含语义分割与点云获取、多模态数据融合和NDT建图与优化3个模块。首先，通过改进的PFPN实现园区场景视觉语义分割；然后，对语义分割数据与点云数据通过空间匹配和时间匹配的方式进行数据融合，使得点云具有语义信息。在建图过程中，通过语义信息去除其中的动态障碍物进行语义地图的优化，从而达到语义的建立并提高建图精度与速度。

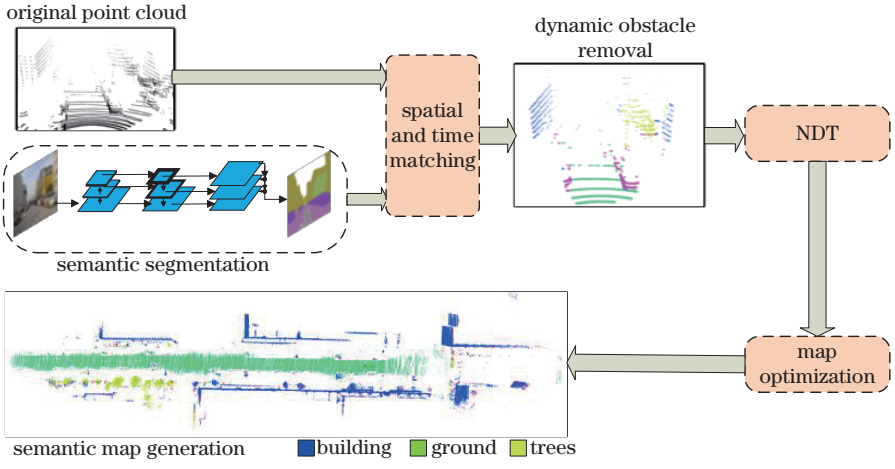

图1　语义地图构建框图

Fig. 1　Block diagram of semantic map construction

2.1　视觉语义分割

构建无人驾驶3D语义地图前要进行基于视觉的2D语义分割，所提方法通过深度学习的方式实现对园区自动驾驶过程中与定位强相关的建筑物、地面和树木等大型静态目标的视觉语义分割。通过对DANet、HRNet、PSPNet模型及基于3种不同主干网络的PFPN模型进行对比筛选，在考虑实时性和准确度为主要目标的情况下以主干网络为ResNet-50-1×的PFPN作为所提方法的视觉语义分割核心，结构如图2所示。PFPN由于具有高分辨率的特征图可以捕捉细节信息的结构，对丰富语义信息进行编码的同时能够准确地预测类别标签，可在多分辨率下预测填充区域捕获多尺度信息。在PFPN框架中，输入为一张图片，对该图片进行下采样之后送入主干网络FPN中完成语义分割特征结果的输出，并通过每一层的卷积操作和上采样操作将网络层恢复到原图像的1/4分辨率，最后使用加法器将每一层相加后输出最后的语义分割结果。PFPN是一种简单、灵活、有效的体系结构，它可以同时使用基于区域的输出和密集像素输出的单个网络来提高语义分割的准确性，但是在园区场景的语义分割中更注重大型静态结构化物体的分割及识别实时性，故而对PFPN进行了优化裁剪，将针对小目标分割的部分裁减掉，借此提升语义分割的速度。主干网络FPN中使用的残差网络(ResNet)分为5个阶段ResNet1～ResNet5，前一阶段1×1卷积操作结果与下一层的上采样输出进行加法运算得到P3～P5阶段的特征图，每一层再经过3×3的卷积操作和2倍上采样得到语义特征点进行语义分割，如图3所示。由于无人驾驶环境定位中主要关注识别的是大型结构化物体如建筑物、树

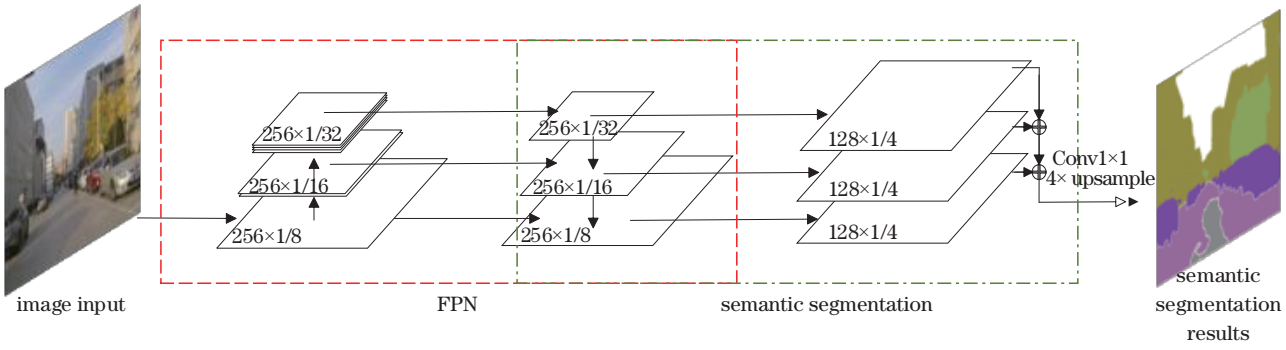

图2　PFPN语义分割网络框架

Fig. 2　Framework of PFPN semantic segmentation network

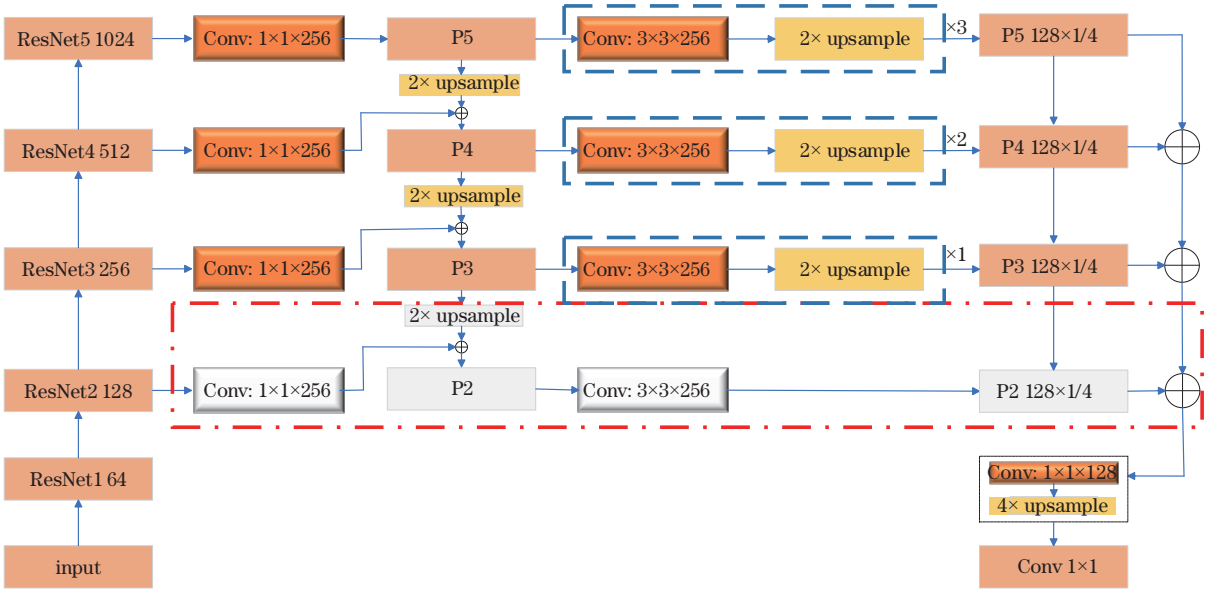

图3　裁减后的PFPN结构

Fig. 3　Structure of reduced PFPN

木、道路等，同时还要保证分割识别的强实时性，故所提方法将PFPN提取特征的主干网络ResNet第2阶段的卷积操作剪掉，图3中点线虚框是被裁剪掉的网络，从第3阶段开始，保留3个阶段特征提取层分别进行1×1的卷积操作得到P3～P5的特征，然后再对P3～P5特征的每一层都进行2倍上采样操作，同时与上一步的卷积操作输出进行加法运算得到每一层丰富的语义特征信息。

为了从FPN特征生成语义分割输出，所提方法采用了一种简单的设计将来自FPN金字塔的所有层的信息合并为一个输出。图2语义分割模块中详细说明了这一点。从最深的FPN(1/32)开始，执行3个上采样，生成了128×1/4尺度的特征图，其中每个上采样级由1×1卷积和2倍线性上采样组成。对于FPN1/16和1/8尺度重复执行上述上采样操作，得到一组相同的1/4比例的特征图，然后按元素求和。为了减小上采样的影响，最后使用1×1卷积、4倍双线性上采样和Softmax生成原始图像分辨率下的每像素类标签，从而达到语义分割的目的。

2.2　多模态数据融合

视觉语义分割完成之后需要和激光点云进行融合，其中最主要的两个步骤为空间匹配和时间匹配。由于不同类型的传感器坐标系不同，为了使单目相机和激光雷达对目标的描述一致，需要将两种传感器的坐标转换到统一的坐标系下。两种传感器之间的坐标转换涉及的坐标系有世界坐标系、激光雷达坐标系、相机坐标系、图像坐标系及像素坐标系，坐标转换原理如图4所示，其中O_L-$X_LY_LZ_L$为世界坐标系，用来描述相机位置，单位为m；O_c-$X_cY_cZ_c$为相机坐标系，光心为原点，单位为m；o-xy为图像坐标系，光心为图像中点，单位为mm；uv为像素坐标系，原点为图像左上角，单位为pixel；P为

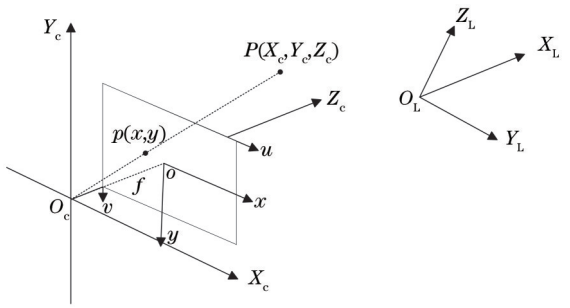

图4　坐标系示意图

Fig. 4　Schematic diagram of coordinate system

世界坐标系中的一点，即真实生活中的一点；p为点P在图像中的成像点，在图像坐标系中的坐标为(x, y)，在像素坐标系中的坐标为(u, v)。摄像头焦距f等于o与O_c的距离，$f=\left|\left|\overrightarrow{oO_c}\right|\right|$。

因为激光雷达和单目相机与实验车辆是刚性连接的，所以激光雷达和单目相机的相对位姿和距离是固定不变的。所提方法关心激光雷达和单目相机两种坐标系之间的转化关系，这里将激光雷达坐标系作为世界坐标系。在世界坐标系下，像素坐标可以利用旋转矩阵$\boldsymbol{R}$和平移矩阵$\boldsymbol{T}$转换到世界坐标系下，激光雷达和像素坐标的坐标转换关系为

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix}=\begin{bmatrix} \frac{1}{\mathrm{d}x} & 0 & u_0 \\ 0 & \frac{1}{\mathrm{d}y} & v_0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} \boldsymbol{R} & \boldsymbol{T} \\ 0 & 1 \end{bmatrix}\begin{bmatrix} X_L \\ Y_L \\ Z_L \\ 1 \end{bmatrix}。\tag{1}$$

由于不同类型传感器数据的采样频率不同，在进行传感器融合时需要对不同传感器采集的时间进行对齐。所使用的16线激光雷达的采样频率大约为10 Hz，单目相机的采样频率为30 Hz。图5中，上排的数据帧为激光雷达数据帧，下排的数据帧为

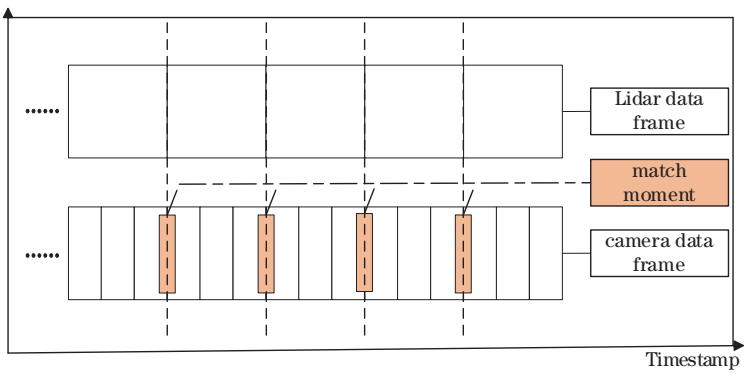

图5　激光雷达和相机时间匹配方法

Fig. 5　Time matching method of Lidar and camera

相机数据帧，阴影部分为激光雷达数据帧和相机数据帧匹配时刻的数据帧，竖直虚线为最佳匹配时刻。因此，当激光雷达完成一次采样后，通过寻找与激光雷达当前时刻最近邻时刻的图像完成两种传感器数据的时间维度匹配。

对相机和激光雷达在时间维度和空间维度的匹配，使得这两种传感器可以在同一时刻正确表示同一物体。在实际操作中，首先需要使用开源平台Autoware进行相机内参的标定，在标定过程中使用10×7的黑白格通过程序自动获取大约40组9×6的交叉点进行计算得出相机内参。再通过相机内参来进行相机与激光雷达的外参联合标定。在联合标定的时候，通过人工标注图像和点云图中一一对应的9组数据进行计算即可得到相机和激光雷达的联合标定参数。最终的标定结果如图6左下角的投影结果所示。

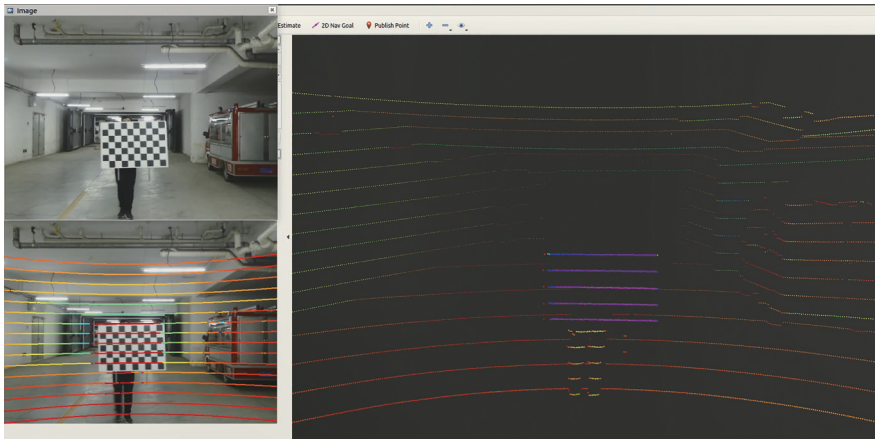

图6　融合结果

Fig. 6　Fusion result

2.3　语义地图构建

对图像语义信息与点云进行融合，使得激光雷达点云数据具有了相应的语义信息，可以开始构建语义地图。图7为语义地图的构建流程，将具有语义信息的点云作为语义地图建立的输入，经过点云滤波操作之后，使用NDT算法进行点云配准获取当前点云的位姿信息，再与原始具有语义信息的点云结合生成当前地图，同时将当前点云位姿与当前地图作为下一时刻的点云预测位姿并再一次进行点云配准，不断循环匹配叠加直至地图构建完成。

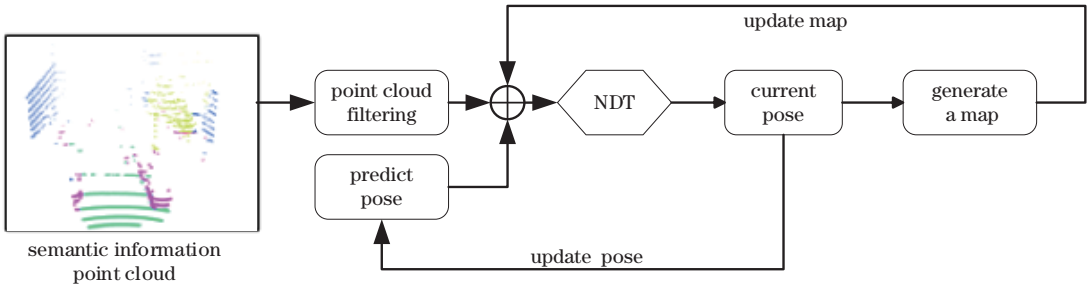

图7　语义地图的构建

Fig. 7　Construction of semantic map

NDT算法是一种应用于三维点的统计模型配准算法，使用标准最优化技术来确定两帧点云间的最优匹配，在配准过程中没有使用对应点的特征计算和匹配，因此NDT算法计算效率相对较高。NDT的基本思想为：先根据参考数据求出多维变量的正态分布，如果求出的两帧点云配准较好，那么变换后的点云数据在参考系中的概率密度最优。使用NDT算法进行位姿估计的步骤如下。

1）将参考点云所占的空间划分成指定大小的网格或体素；并计算每个网格的多维正态分布参数均值$\boldsymbol{q}$和协方差矩阵$\boldsymbol{\Sigma}$。

$$\boldsymbol{q}=\frac{1}{n}\boldsymbol{\Sigma}_i\boldsymbol{x}_i, \tag{2}$$

$$\sum=\frac{1}{n}\sum_i\left(\boldsymbol{x}_i-\boldsymbol{q}\right)\left(\boldsymbol{x}_i-\boldsymbol{q}\right)^{\mathrm{T}}, \tag{3}$$

式中：$\boldsymbol{x}_i$为每个网格内所有的扫描点。

2）根据正态分布参数计算每个转换点的概率密度。

$$p(\boldsymbol{x})\sim\exp\frac{(\boldsymbol{x}-\boldsymbol{q})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{q})}{2}。\tag{4}$$

3）通过对每个网络计算出的概率密度进行相加得到NDT配准得分（score）。

$$S_{\mathrm{score}}(p)=\sum_i\exp\left[-\frac{(\boldsymbol{x}_i'-\boldsymbol{q}_i)^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{x}_i'-\boldsymbol{q}_i)}{2}\right]。\tag{5}$$

2.4　语义地图优化

实验的园区环境中包含的障碍物有建筑物、树木、车辆、行人、地面5类物体，其中车辆与行人相对于其他3类物体来说不会长时间停留，故而将其作为同一类型动态障碍物。在采集实验数据时不可避免地会出现车辆与行人，而每次进行无人驾驶测试时由于环境的改变需要重新采集数据进行建图，这样大大降低了测试效率，同时建图过程中两帧数据的匹配也会因此而存在误差，使得语义地图建图精度降低，因此需要对语义地图进行优化。

通过前期语义分割、多传感器数据融合和语义地图的建立等步骤建立了一幅完整的语义地图。在构建完成的语义地图中可以通过不同标签来区分不同物体，而用于构建语义地图的每一帧点云数据已经具备了语义信息，语义地图的优化可以从每一帧的点云数据入手，去除每一帧点云数据中的行人、车辆数据，然后再使用NDT匹配方法进行语义地图的构建，从而达到优化语义地图的目的。

语义地图经过优化后，在每一次无人驾驶测试时就可以不用重新采集数据来构建点云地图，节省了测试时间，为真正需要测试的问题留出了更多解决时间。同时在去除每一帧数据中的动态障碍物后，点云数据也会大大降低，保留更多的有效数据，不会因为动态障碍物的去除使得建图过程出现误差，从而提高语义地图的构建精度。

3　实验设计与分析

3.1　实验设计与数据集介绍

为了验证所提方法的有效性，进行了大量的相关实验。所采用的实验平台为北京联合大学旋风智能车平台，实验测试场景为北京联合大学无人驾驶测试路线，全长约为180 m，如图8所示。实验硬件计算平台配置为Intel（R）Core（TM）i7-10875H CPU @2.30 GHz，32 GB RAM和RTX2070显卡，软件为Ubuntu 18.04。实验共分为两个阶段，实验1为视觉语义分割，通过对多种视觉语义分割方法进行对比实验，验证所提方法在实时性和准确性方面的性能。实验2为语义地图构建的实验，通过对比去除行人、车辆等动态障碍物前后的效果，验证所提方法的性能。

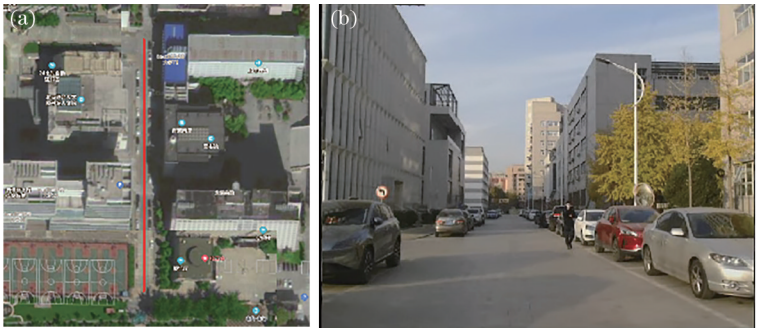

图8　实验测试路线与测试环境。(a)测试路线；(b)测试环境

Fig. 8　Experimental test route and test environment. (a) Test route; (b) test environment

为了使所训练出来的视觉语义分割模型可以适应各种复杂的园区环境实现对车辆、行人、建筑、树木等物体的分割，所提方法选择COCO数据集中的全景分割数据集对模型进行训练，该训练数据集有40000张图片，可输出92个Stuff类。实验使用的实验测试数据集为北京联合大学校区无人驾驶测试路线数据，其中图像数据集共有2736张图片，测试图片为随机选取的五分之一。

3.2　语义分割的实验结果与分析

语义分割模型在精度上常使用的评价指标是平均像素交并比（mIoU），mIoU可以有效衡量模型对场景中物体分割性能的好坏，mIoU值越高，说明性能越好。对应同样数据集，对7种不同方法进行了性能对比测试，结果如表1所示。从表中分析可以得知，虽然HRNet方法的mIoU值在表中最好，但是3.89 s的测试时间远大于无人驾驶实时性要求0.1 s

表 1　不同方法性能对比

Table 1　Comparison of different methods

Method	Backbone	mIoU /%	Time /s
DANet	ResNet-101	39.7	
HRNet	HRNet	42.7	3.89
PSPNet	ResNet-50	39.9	0.36
PFPN	ResNet-101-3×	42.1	0.28
PFPN	ResNet-50-3×	40.2	0.21
PFPN	ResNet-50-1×	41.2	0.18
Proposed method	ResNet-50-1×	41.0	0.13

的时间。均衡考虑性能及测试时间，所提方法的 mIoU 为 41.0%，测试时间 0.13 s 与 0.1 s 最接近，大大节省了计算时间，与原始主干网络为 ResNet-50-1× 的 PFPN 算法的 0.18 s 相比，减少了 27.78%。

3.3　语义地图的实验结果与分析

通过语义分割视觉信息与激光雷达的融合方法，点云数据具有了语义信息。通过 NDT 方法对具有语义信息的点云进行 SLAM 建图的实验结果如图 9 所示。图 9(a)为原始点云建立的 SLAM 地图，从该图中并不能够具体对环境中的物体信息进行分类，图 9(b)为具有语义信息的点云去除环境中车辆信息的语义地图。图 9(c)、(d)分别为图 9(a)、(b)中矩形区域的放大显示，图 9(c)方框中显示的是停放在道路两旁的车辆，从图 9(d)中明显可以看出，图 9(c)中相同位置的车辆已被移除，只保留了建筑、树木等信息。实验点云数据的总帧数约为 912，表 2 为平均每帧点云个数，从表 2 中可以看出，在大约 180 m 的直线路径中，原始点云数中平均每帧点云个数约为 2089，去除动态障碍物之后平均每帧点云个数为 1258，减少了 831，减少百分比为 39.78%，表明所提方法提高了点云有效数量，从而为点云配准提供了良好的有效点云数据基础。表 3 为平均单次配准时间对比，由表 3 可以看出，使用原始方法进行配准时单次平均时间为 53 ms，所提方法单次平均配准时间为 31 ms，比原始方法平均配准时间减少了 22 ms，时间减少百分比为 41.51%。

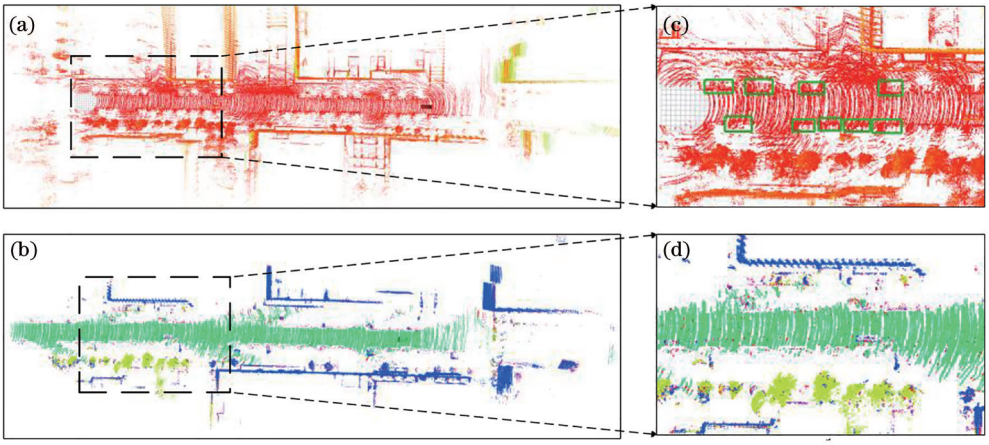

图 9　地图对比。(a)原始点云建立的 SLAM 地图；(b)去除环境中车辆信息后的语义地图；(c)(d)图 9(a)、(b)中矩形区域的放大显示

Fig. 9　Map comparison. (a) SLAM map established by original point cloud; (b) semantic map after removing vehicle information in environment; (c) (d) enlarged display of rectangular area in Fig. 9 (a), (b)

表 2　平均每帧点云个数

Table 2　Average number of point clouds per frame

Method	Nnumber of point clouds
Original method	2089
Proposed method	1258
Change	831

表 3　平均单次配准时间对比

Table 3　Comparison of average single registration time

Method	Time /ms
Original method	53
Proposed method	31
Change	22

通过点云数量的对比可知，所提方法很好地去除了移动障碍物，保留了有效数据。接下来从点云 x 轴位移误差、y 轴位移误差及整体位姿误差 3 方面来分析所提方法的有效性。图 10 中圆点表示的轨迹为全球导航卫星系统(GNSS)轨迹，由于实验环境下对空信号良好，在本实验中将 GNSS 作为轨迹真值与其他方法轨迹进行对比。通过图中的轨迹对比可以发现，所提方法的倒三角轨迹相较于原始

方法的上三角轨迹更接近GNSS真值轨迹。图11为x轴的平移误差和y轴的平移误差，从图中可以看出，所提方法的误差更小。图12为位姿误差对比图，实线为所提方法位姿误差，与原方法相比误差

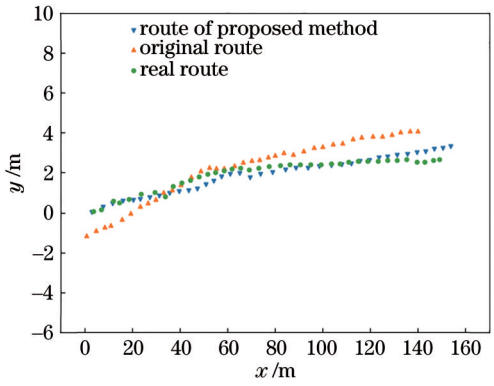

图10　建图轨迹

Fig. 10　Mapping track

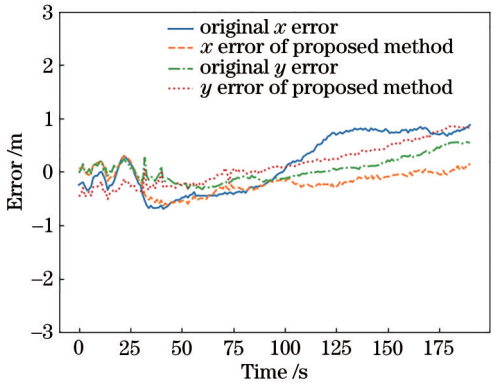

图11　x轴与y轴误差对比

Fig. 11　Error comparison between x-axis and y-axis

更小。

定性分析x轴、y轴的平移误差以及位姿误差后，对上述参数进行定量分析，结果如表4所示，其中x表示车辆x轴上的位移，y表示车辆y轴上的位移，P表示位姿。从表4中可以看出，所提方法在x轴上的平均误差降低了0.1090 m，降低百分比为24.84%，在y轴上的平均误差降低了0.1137 m，降低百分比为32.41%，位姿平均误差降低了0.0844 rad，精度提高百分比为34.34%，表明了所提方法的有效性。

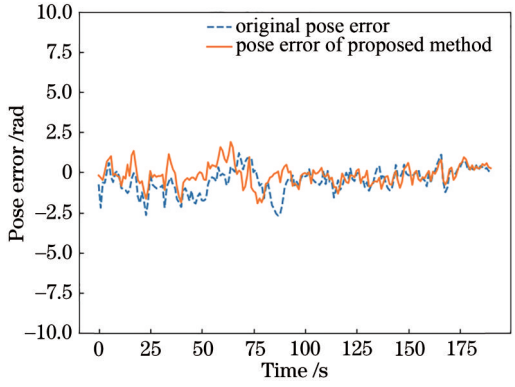

图12　整体位姿误差对比

Fig. 12　Comparison of overall pose error

表4　关键参数平均误差分析

Table 4　Average error analysis of key parameters

Parameter	Original method	Proposed method	Change
x /m	0.4388	0.3298	0.1090
y /m	0.3508	0.2371	0.1137
P /rad	0.2458	0.1614	0.0844

4　结　　论

基于单目相机与激光雷达数据融合，结合改进的PFPN进行园区场景语义分割，使得激光雷达点云数据获得相对应的语义信息，通过NDT方法构建了语义地图，从而有效滤除场景中的动态障碍物，提高建图精度，实现精确定位，为无人车的自主导航提供了更加精确先验地图，在无人驾驶领域具有实际的应用价值。接下来的研究是探索如何直接对点云进行语义分割，来减少多传感器融合过程中带来的误差，更进一步提高建图精度，将语义地图应用于更加复杂的园区场景中。

参　考　文　献

[1] Nüchter A, Hertzberg J. Towards semantic maps for mobile robots[J]. Robotics and Autonomous Systems, 2008, 56(11): 915-926.

[2] Floros G, Leibe B. Joint 2D-3D temporally consistent semantic segmentation of street scenes[C]//2012 IEEE Conference on Computer Vision and Pattern Recognition, June 16-21, 2012, Providence, RI, USA. New York: IEEE Press, 2012: 2823-2830.

[3] Larlus D, Jurie F. Combining appearance models and Markov Random Fields for category level object segmentation[C]//2008 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2008, Anchorage, AK, USA. New York: IEEE Press, 2008: 10139758.

[4] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New

York: IEEE Press, 2015: 3431-3440.

[5] Paszke A, Chaurasia A, Kim S, et al. ENet: a deep neural network architecture for real-time semantic segmentation[EB/OL]. (2016-06-07) [2021-03-01]. https://arxiv.org/abs/1606.02147.

[6] Badrinarayanan V, Kendall A, Cipolla R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12): 2481-2495.

[7] Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions[EB/OL]. (2015-11-23)[2021-03-01]. https://arxiv.org/abs/1511.07122.

[8] Lin G S, Milan A, Shen C H, et al. RefineNet: multi-path refinement networks for high-resolution semantic segmentation[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 5168-5177.

[9] Davison A J, Reid I D, Molton N D, et al. MonoSLAM: real-time single camera SLAM[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(6): 1052-1067.

[10] Whelan T, Salas-Moreno R F, Glocker B, et al. ElasticFusion: real-time dense SLAM and light source estimation[J]. The International Journal of Robotics Research, 2016, 35(14): 1697-1716.

[11] Mur-Artal R, Montiel J M M, Tardós J D. ORB-SLAM: a versatile and accurate monocular SLAM system[J]. IEEE Transactions on Robotics, 2015, 31(5): 1147-1163.

[12] Charles R Q, Hao S, Mo K C, et al. PointNet: deep learning on point sets for 3D classification and segmentation[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 77-85.

[13] Tchapmi L, Choy C, Armeni I, et al. SEGCloud: semantic segmentation of 3D point clouds[C]//2017 International Conference on 3D Vision (3DV), October 10-12, 2017. Qingdao. New York: IEEE Press, 2017: 537-547.

[14] Zou B, Lin S Y, Yin Z S. Semantic mapping based on YOLOv3 and visual SLAM[J]. Laser & Optoelectronics Progress, 2020, 57(20): 201012.
邹斌, 林思阳, 尹智帅. 基于 YOLOv3 和视觉 SLAM 的语义地图构建[J]. 激光与光电子学进展, 2020, 57(20): 201012.

[15] Zhang A W, Liu L L, Zhang X Z. Multi-feature 3D road point cloud semantic segmentation method based on convolutional neural network[J]. Chinese Journal of Lasers, 2020, 47(4): 0410001.
张爱武, 刘路路, 张希珍. 道路三维点云多特征卷积神经网络语义分割方法[J]. 中国激光, 2020, 47(4): 0410001.

[16] Zhang J Y, Zhao X L, Chen Z. Review of semantic segmentation of point cloud based on deep learning[J]. Laser & Optoelectronics Progress, 2020, 57(4): 040002.
张佳颖, 赵晓丽, 陈正. 基于深度学习的点云语义分割综述[J]. 激光与光电子学进展, 2020, 57(4): 040002.

[17] Kirillov A, Girshick R, He K M, et al. Panoptic feature pyramid networks[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 6392-6401.