

道路三维点云多特征卷积神经网络语义分割方法

张爱武^{1,2}, 刘路路^{1,2*}, 张希珍^{1,2}

¹首都师范大学资源环境与旅游学院三维信息获取与应用教育部重点实验室, 北京 100048;

²首都师范大学资源环境与旅游学院空间信息技术教育部工程研究中心, 北京 100048

摘要 针对道路场景下三维激光点云语义分割精度低的问题, 提出了一种基于卷积神经网络并结合几何点云多特征的端到端的语义分割方法。首先, 通过球面投影构造出点云距离、相邻夹角及表面曲率等特征图像, 以便于应用卷积神经网络; 接着, 利用卷积神经网络对多特征图像进行语义分割, 得到像素级的分割结果。所提方法将传统点云特征融入到卷积神经网络中, 提升了语义分割效果。使用 KITTI 点云数据集进行测试, 结果表明: 所提三维点云多特征卷积神经网络语义分割方法的效果优于 SqueezeSeg V2 等没有结合点云特征的语义分割方法; 与 SqueezeSeg V2 网络相比, 所提方法对车辆、自行车和行人分割的精确率分别提高了 0.3、21.4、14.5 个百分点。

关键词 遥感; 神经网络; 激光点云; 语义分割; 多特征; 点云投影

中图分类号 P237

文献标志码 A

doi: 10.3788/CJL202047.0410001

Multi-Feature 3D Road Point Cloud Semantic Segmentation Method Based on Convolutional Neural Network

Zhang Aiwu^{1,2}, Liu Lulu^{1,2*}, Zhang Xizhen^{1,2}

¹Key Laboratory of 3D Information Acquisition and Application, Ministry of Education, College of Resource Environment and Tourism, Capital Normal University, Beijing, 100048 China;

²Engineering Research Center of Space Information Technology, Ministry of Education, College of Resource Environment and Tourism, Capital Normal University, Beijing 100048 China

Abstract Aiming at the problem of low accuracy in semantic segmentation of three-dimensional laser point clouds in road scene, an end-to-end multi-feature point clouds semantic segmentation method based on convolutional neural network is proposed. Firstly, the feature images such as point cloud distance, adjacent angle and surface curvature are calculated based on spherical projection to apply to convolutional neural network; then, a convolutional neural network is adopted to process multi-band depth images to obtain pixel-level instance segmentation results. The proposed method combines traditional point cloud features with the deep learning method to improve the result of point cloud semantic segmentation. Using KITTI point cloud data set test, simulation results show that the multi-feature convolutional neural network semantic segmentation method has better performance than other semantic segmentation methods without combining with point cloud features such as SqueezeSeg V2. The precision obtained with proposed method for car, bicycle and pedestrian segmentation is 0.3, 21.4, 14.5 percentage points higher in comparison with the SqueezeSeg V2 network.

Key words remote sensing; neural network; laser point cloud; semantic segmentation; multi-feature; point cloud projection

OCIS codes 280.3640; 100.6890; 100.4996; 100.4999

收稿日期: 2019-09-25; 修回日期: 2019-11-08; 录用日期: 2019-11-26

基金项目: 国家自然科学基金(41571369)、国家重点研发计划(2016YFB0502500)、北京市自然科学基金(4162034); 青海省科技计划(2016-NK-138)

* E-mail: liululu@cnu.edu.cn

1 引 言

三维激光雷达已被广泛应用于无人驾驶汽车和智能机器人领域^[1]。三维点云分割是激光雷达应用的重要课题,是人们研究的热点之一,研究人员提出了很多较好的分割方法,如: Douillard 等^[2]先直接建立复杂的点云 3D 特征,再利用聚类算法将点云分割为一个个子集,实现了对点云数据的有效分割;贺亦峰等^[3]提出了一种结合 RANSAC 和欧氏聚类的点云分割方法,该方法可以很好地分割出建筑物立面点云;黄凯等^[4]利用基于法向量条件约束的随机采样一致性算法实现了密集圆形管道点云数据的自动分割方法。然而,上述这些方法的分割精度在很大程度上取决于各种人为设置的经验阈值,鲁棒性较差,对于复杂多变的道路场景难以做到有效分割。

近年来,随着深度学习方法的发展,出现了很多图像分割分类方法,其中,2015 年发布的 ResNet 网络^[5-6]在 ImageNet 图像识别竞赛中的错误率已经达到了低于人眼的错误率。有研究者曾将深度学习应用到三维激光点云中,但是受到点云稀疏性和无序性的限制,同一目标在不同位置采集的点云会有很大差别,无法直接利用现有的二维图像网络模型进行点云语义分割。现有的深度学习三维点云分割方法大致可分为两类:基于端到端的点云分割方法和基于投影的点云分割方法。基于端到端的点云分割方法以斯坦福大学的 PointNet 算法^[7]为代表,该方法直接对无序点云进行处理,利用对称函数解决了三维激光点云无序性的难题。之后,斯坦福大学的研究人员又对 PointNet 进行了改进,提出了 PointNet++^[8]算法,该算法将点云的全局特征和局部特征进行融合,增强了网络对点云局部特征的描述能力。基于投影的点云分割方法将三维点云直接投影到一个平面或者曲面上,利用投影得到的栅格图像对点云进行目标检测或者语义分割等处理。Complex-YOLO 网络^[9]是基于投影的三维点云目标识别的经典网络之一,该网络可以直接在三维激光点云的俯视投影图上进行操作,从而实现目标位置及种类信息的精确估计和定位; SqueezeSeg 网络是基于投影的三维点云分割网络的另一个代表,由 Wu 等^[10]在 2018 年首次提出。不同于 Complex-YOLO 网络^[9], SqueezeSeg 网络^[10]利用点云前视投影得到的深度图实现了三维点云物体的实时语义分割。同年, Wang 等在 SqueezeSeg 网络^[10]的基础

上引入了 squeeze reweighting 层和 enlargement 层,提出了 PointSeg 网络^[11],该网络增大了卷积神经网络的感受野,进而提高了分割精度。之后, Wu 等^[12]又提出了 SqueezeSeg V2 网络,该网络在原有网络的基础上加入了上下文聚合模块,进一步提高了网络的分割效果。在基于投影的点云分割算法中,除了基于深度学习的分割方法外,还有基于点云特征的分割方法。Bogoslavskyi 等^[13]利用建立点云深度图的方法,结合点云深度图的角度特征信息对非地面点云进行聚类分割,在 i7 处理器上实现了 667 frame/s 的实时分割速度。范小辉等^[14]利用自适应参数改进的 DBSCAN 算法结合点云深度图的角度特征信息实现了更高精度的聚类分割。

基于端到端的点云分割方法的计算复杂度比较高,很难实现点云的快速分割和精确识别;而基于投影的点云分割方法,将三维点云数据转化到了二维空间,大大降低了算法的复杂度^[13]。因此,基于投影的点云分割方法相较于基于端到端的点云分割方法更加高效,能够实现更快的分割速度,更加适合于道路场景下的点云分割。在基于投影的点云分割方法中,基于点云特征的分割方法虽然能够很好地利用点云的角度特征信息,实现很高的分割速度,但是需要设置经验阈值才能实现良好的分割效果,鲁棒性较差。基于深度学习的点云分割方法能够利用卷积神经网络在不设置经验阈值的情况下实现三维点云物体的实时语义分割,但该方法只利用了点云本身的坐标信息,且对自行车和行人的分割效果较差。因此,本文将点云特征信息与卷积神经网络结合起来,基于前视投影的深度图,结合点云特征建立了点云多特征图像矩阵,以 SqueezeSeg V2 网络^[12]为基础提出了一种基于卷积神经网络的道路多特征三维点云语义分割方法,该方法相较于没有结合点云特征的原始网络,实现了高精度的三维点云目标语义分割。

2 点云多特征图像构造

三维激光点云数据具有散乱无序的特性^[15],无法像图像一样使用 $H \times W \times N$ (H 代表图像高度, W 代表图像宽度, N 代表图像通道数)的三维张量来表示点云数据,需要利用投影的方法将无序点云数据转化为有序的二维图像,以适应传统卷积神经网络模型。现有的基于卷积神经网络的三维点云语义分割模型的输入层大都只包括点云 x 、 y 、 z 坐标和强度信息,鉴于此,本文通过建立多种点云特征构

造出了点云多特征三维矩阵,使得卷积神经网络能够直接获得具有先验知识的高级语义信息,从而提高了分割精度。

2.1 投影方式

点云的投影方式一般分为两种:一种是从上向下的俯视投影,得到的结果称为鸟瞰图;另一种则是利用球面或者外接圆柱进行前视投影,得到的结果称为深度图。根据俯视投影的投影方式可知,基于俯视投影的鸟瞰图在描述垂直立面物体时,立面物体下面部分的点云可能会被遮挡,从而丢失信息。与此同时,俯视投影后的结果一般会包含很多空白像素,在利用效率上也会弱于基于前视投影得到的深度图。因此,本文选择使用前视投影方法对点云进行投影。

前视投影的公式为

$$\theta = \arcsin \frac{z}{\sqrt{x^2 + y^2 + z^2}}, H = \theta / \Delta\theta, \quad (1)$$

$$\varphi = \arcsin \frac{y}{\sqrt{x^2 + y^2}}, W = \varphi / \Delta\varphi, \quad (2)$$

式中: θ 为天顶角; $\Delta\theta$ 和 $\Delta\varphi$ 分别为垂直方向和水平方向上投影的角分辨率; φ 为方位角; H 代表投影后影响垂直方向的像素数; W 代表投影后影像水平方向的像素数。前视投影方式示意图及结果如图 1 所示。

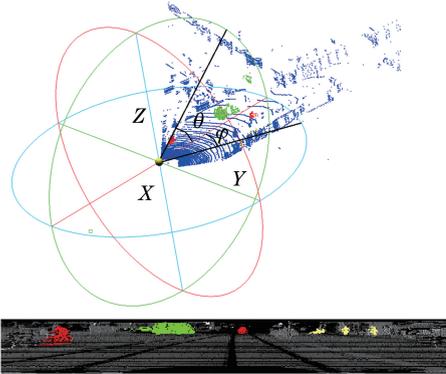


图 1 投影方式及深度图

Fig. 1 Projection type and the depth map

由于点云采集设备 Velodyne HDL-64E 激光扫描仪有 64 条扫描线^[16],因此投影后影像垂直方向的像素数 $H = 64$ 。与此同时,在投影过程中,本文选取了传感器正前方 90° 范围内的点云数据进行投影,并将其分到 512 个像素中,即投影后影像水平方向的像素数 $W = 512$ 。

2.2 特征计算

本文选择了点云原始信息(x, y, z)和几种点云

特征来构造点云多特征图像,所选点云特征如下:强度信息(intensity)、距离特征(d)、垂直夹角特征(α)、水平夹角特征(β)、点云表面曲率特征(curvature)。其中,距离特征是点云分割中一个常用的重要特征,代表每个点到激光雷达的距离。相邻点云之间连线的倾角所形成的角度值可以作为三维点云分割的一个重要特征,Bogoslavskyi 等^[13]通过计算相邻点云之间连线的倾角并设置阈值实现了三维点云在深度图中的快速分割,证明了相邻点云之间连线的倾角特征是三维点云场景分割中一个十分重要的特征。由于相邻点云连线的倾角存在垂直和水平两种连接方式,因此可以根据这两种方式计算出不同的夹角特征:垂直夹角特征和水平夹角特征。垂直夹角特征可以有效地帮助分割地面点与非地面点,水平夹角特征则可以用来帮助分割地面点上相邻的不同物体。点云曲率特征是点云曲面识别的重要特征,对于三维点云的分割有着十分重要的作用,是一个十分重要的点云局部几何属性^[17-18]。曾碧等^[18]利用曲率以及其他点云特征完成了复杂室内场景的点云分割,证明了点云曲率特征可以更好地区分不同曲率的物体,为点云分割提供了关键信息,显著提高了点云场景分割的正确率。

综上所述,本文选择距离特征、垂直夹角特征、水平夹角特征和点云表面曲率特征来构造点云多特征图像。下面介绍特征的计算过程。

距离特征是指每个点到激光扫描仪中心的距离,计算公式为 $d = \sqrt{x^2 + y^2 + z^2}$ 。距离特征投影图如图 2 所示。



图 2 距离特征投影图

Fig. 2 Distance feature projection image

垂直夹角特征是通过计算相同水平夹角上的两个垂直相邻点连线的倾角得到的。如图 3(a)所示,现有水平夹角相同的两垂直方向相邻的点 a 和 b ,通过计算这两点之间的连线与 X - Y 面的夹角即可得到角度 α 。而在投影后的深度图上计算垂直夹角特征则是通过 a, b 两点在投影后图像中对应的第 w 列第 $h-1$ 行和第 h 行的两个像素点的 x, y, z 坐标值来计算的,计算公式为

$$\alpha_{h,w} = \arctan(\Delta z_{h,w} / \Delta x_{h,w}), \quad (3)$$

$(h \in H, h > 1, w \in W)$ 。

利用第 $h-1$ 行第 w 列的 $z_{h-1,w}$ 值和第 h 行第

w 列的 $z_{h,w}$ 值计算出 $\Delta z_{h,w}$, 即 $\Delta z_{h,w} = |z_{h-1,w} - z_{h,w}|$, 其中, $h \in H, h > 1, w \in W$; 利用第 $h-1$ 行第 w 列的 $x_{h-1,w}$ 值和第 h 行第 w 列的 $x_{h,w}$ 值计算出 $\Delta x_{h,w}$, 即 $\Delta x_{h,w} = |x_{h-1,w} - x_{h,w}|$, 其中, $h \in H, h > 1, w \in W$; 之后利用 $\arctan(\Delta z_{h,w} / \Delta x_{h,w})$ 就可以计算出 α 。每一个 α 都需要两个点的坐标值才能计算, 所以实际计算出的 α 角度特征图的大小会少一组像素, 实际上只能得到大小为 63×512 的特征图。为了方便后期计算, 需要补全 α 角特征图, 使得到的特征图和其他特征图的大小保持一致。因此, 本文将 $h=2$ 时的 α 赋值给首行 $h=1$ 时的 α , 即

$$\alpha_{1,w} = \alpha_{2,w}, (h \in H, w \in W), \quad (4)$$

最终可以得到大小为 64×512 的 α 角特征图。

与垂直夹角特征类似, 水平夹角特征是通过计算相同垂直夹角上的两个水平相邻点之间连线的倾角得到的。如图 3(b) 所示, 通过计算相同垂直夹角上水平夹角相邻的 a, b 两点之间的连线与 $X-Z$ 平面的夹角得到水平夹角特征。在投影后的图像上计算水平夹角特征也是通过 a, b 两点在投影后图像中对应的第 h 行第 $w-1$ 列和第 w 列的两个像素点的 x, y, z 值来计算的, 计算公式为

$$\beta_{h,w} = \arctan(\Delta x_{h,w} / \Delta y_{h,w}), \quad (h \in H, w \in W, w > 1). \quad (5)$$

利用第 h 行第 $w-1$ 列的 $x_{h,w-1}$ 值和第 h 行第 w 列的 $x_{h,w}$ 值计算出 $\Delta x_{h,w}$, 即 $\Delta x_{h,w} = |x_{h,w-1} - x_{h,w}|$, 其中, $h \in H, w \in W, w > 1$; 利用第 h 行第 $w-1$ 列的 $y_{h,w-1}$ 值和第 h 行第 w 列的 $y_{h,w}$ 值计算出 $\Delta y_{h,w}$, 即 $\Delta y_{h,w} = |y_{h,w-1} - y_{h,w}|$, 其中 $h \in H, w \in W, w > 1$; 之后, 利用 $\arctan(\Delta x_{h,w} / \Delta y_{h,w})$ 就可以计算出 β 。与垂直夹角特征得到的特征图类似, 水平夹角特征图也需要进行补全操作, 本文将第二列的 β 赋值给第一列, 即

$$\beta_{h,1} = \beta_{h,2}, (h \in H, w \in W). \quad (6)$$

激光扫描仪在实际工作过程中会产生一些离群值和噪声, 同时由于误差的存在, 投影后的特征图中会出现空洞现象。为了解决这些问题对垂直夹角特征和水平夹角特征造成的影响, 本文使用 Savitzky-Golay 滤波器^[19]对每一列和每一行数据进行平滑滤波, 以减小离群值、噪声以及投影误差导致的空洞现象带来的影响。Savitzky-Golay 滤波器^[19]简称为 S-G 滤波器, 它的核心思想是对一定长度窗口内的数据点进行 k 阶多项式拟合, 从而得到更加平滑的数据。S-G 滤波器被广泛用于数据的平滑去噪, 公

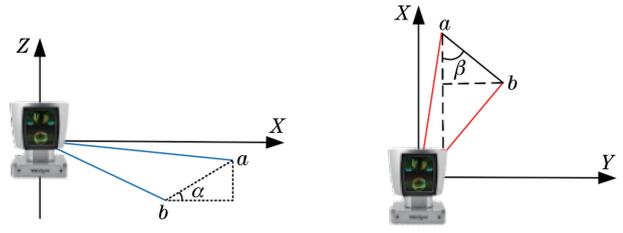


图 3 夹角特征计算示意图。(a)垂直夹角特征 α ;
(b)水平夹角特征 β

Fig. 3 Schematics of angle feature. (a) Vertical angle feature α ; (b) horizontal angle feature β

式为

$$Y_j^* = \frac{\sum_{i=-m}^m C_i B_{j+1}}{N}, \quad (7)$$

式中: Y_j^* 表示计算结果; B 代表原始数据; C_i 代表滤波系数; N 表示数据的滑动宽度 ($(2m+1)$)。S-G 滤波器处理前后的角度特征图对比如图 4 所示。

在曲率特征计算中, 本文利用滑动最小二乘曲面 (MLS)^[20]方法对点云中每个点的曲率进行计算。计算方法如下: 滑动最小二乘曲面是能量函数 $e(\mathbf{y}, \mathbf{a})$ 沿着向量场 $\mathbf{n}(x)$ 方向的局部最小值。其中 $e(\mathbf{y}, \mathbf{a})$ 和 $\mathbf{n}(x)$ 的计算公式为

$$\begin{cases} e(\mathbf{y}, \mathbf{a}) = \sum_{q_i \in Q} [(\mathbf{y} - \mathbf{q}_i) \mathbf{a}^\top]^2 \theta(x_i, \mathbf{q}_i) \\ \mathbf{n}(x) = \frac{\sum_{q_i \in Q} \mathbf{v}_i \theta(x, \mathbf{q}_i)}{\sum_{q_i \in Q} \theta(x, \mathbf{q}_i)} \end{cases}, \quad (8)$$

式中: \mathbf{y}, \mathbf{a} 分别为位置向量和方向向量; \mathbf{q}_i 属于点云集 Q 中的一个点向量; \mathbf{v}_i 为 \mathbf{q}_i 对应的法向量; $\theta(x_i, \mathbf{q}_i)$ 代表高斯加权函数。对函数 $e(\mathbf{y}, \mathbf{a})$ 求偏导数并寻找其极小值点, 即:

$$g(x) = \mathbf{n}(x)^\top \frac{\partial e[\mathbf{y}, \mathbf{n}(x)]}{\partial \mathbf{y}} \Big|_{\mathbf{y}=\mathbf{x}}. \quad (9)$$

由(9)式可以得到滑动最小二乘曲面的表达式, 则高斯曲率和平均曲率分别为

$$k_{\text{Gaussian}} = \frac{\text{Det} \begin{pmatrix} H(g(x)) & \nabla^\top g(x) \\ \nabla g(x) & 0 \end{pmatrix}}{\nabla g(x)^4}, \quad (10)$$

$$k_{\text{mean}} = \{\text{Det}[\nabla g^3(x)] \cdot \text{Trace}(H) - \nabla g(x) \cdot H[g(x)] \cdot \nabla^\top g(x)\} / \nabla g^3(x), \quad (11)$$

则主曲率可由高斯曲率和平均曲率得到:

$$k_{\pm} = k_{\text{mean}} \pm \sqrt{k_{\text{mean}}^2 - k_{\text{Gaussian}}}. \quad (12)$$

对每个点云的曲率进行计算后进行前视投影,

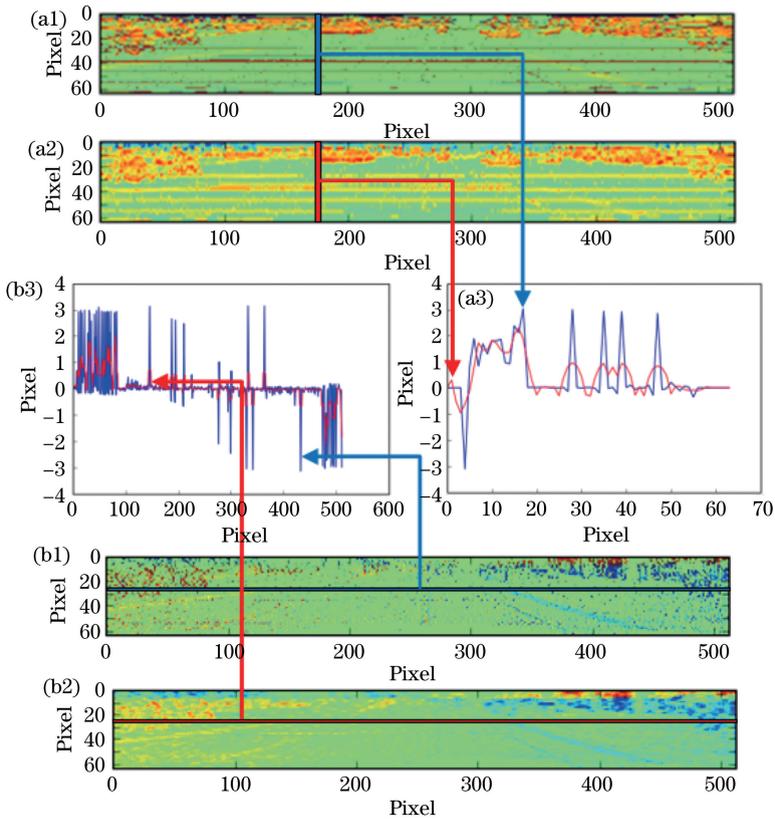


图 4 夹角特征图及滤波后的效果。(a1)垂直夹角特征图;(a2)滤波后的垂直夹角特征图;

(a3)一列像素滤波前后的对比图; (b1)水平夹角特征图;(b2)滤波后的水平夹角特征图;(b3)一行像素滤波前后的对比图
 Fig. 4 Angle feature images and their effect after filtering. (a1) Vertical angle feature image; (a2) vertical angle feature image after filtering; (a3) comparison of pixel values before and after filtering of a column of pixels; (b1) horizontal angle feature image; (b2) horizontal angle feature image after filtering; (b3) comparison of pixel values before and after filtering of a row of pixels

得到的特征图像如图 5 所示。



图 5 曲率特征图像

Fig. 5 Projection image of curvature feature

由点云曲率特征图像可以看出,点云曲率特征可以很好地区分平面和曲面信息。图中的地面和自行车、行人之间差异巨大,可以很好地帮助分割自行车和行人等非平面物体。

通过以上点云特征的计算,共得到了 4 种激光点云原始信息(x 、 y 、 z 、强度)和 4 种点云特征信息(距离特征、垂直夹角特征、水平夹角特征、点云表面曲率特征),将原始信息和特征信息合并起来即可得到一个大小为 $64 \times 512 \times 8$ 的含有原始信息和特征信息的三维图像矩阵,即神经网络的输入数据,如图 6 所示。

3 点云多特征卷积神经网络结构

点云多特征卷积神经网络以前视投影为基础,

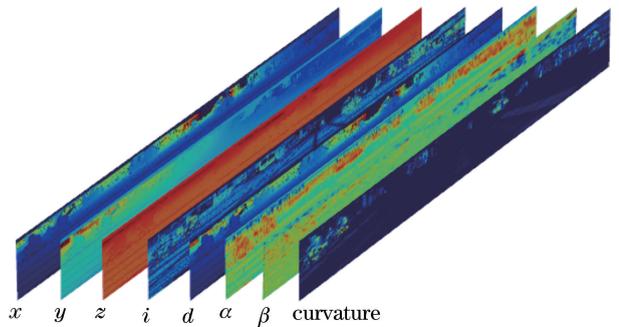


图 6 包含点云原始信息和特征信息的三维矩阵

Fig. 6 Three-dimensional matrix containing primitive information and feature information of point cloud

针对要分割的目标特征通过建立多种点云特征来生成点云多特征图像矩阵。点云多特征图像矩阵作为输入层被输入到卷积神经网络中,可直接为卷积神经网络提供具有先验知识的高级语义信息,使卷积神经网络在特征提取时能更加高效地获取信息,提升分割精度。

在基于深度图的点云语义分割网络中,

SqueezeSeg V2 网络^[12]无论是相较于前代 SqueezeSeg 网络^[10]还是 Wang 等^[11]提出的 PointSeg 网络,效果都更加优异。因此本文选择使用 SqueezeSeg V2 网络作为本文卷积神经网络部分的主要网络结构,并将这

一网络进行适当的改进,以适应多特征三维点云数据。本文多特征点云语义分割网络整体上是一个漏斗形状,主要分为 4 部分:输入层、特征提取层、上采样层和输出层。网络的总体架构如图 7 所示。

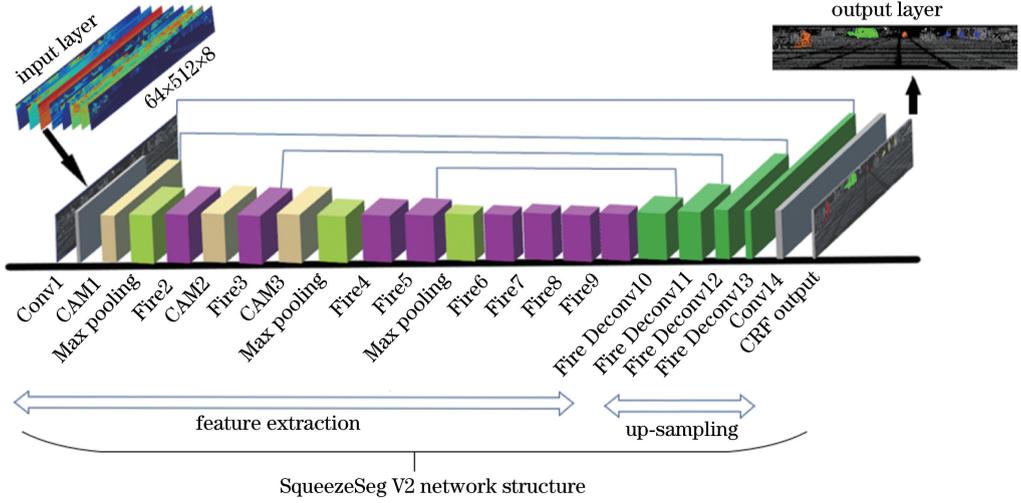


图 7 点云多特征卷积神经网络结构图

Fig. 7 Structure of multi feature point cloud convolution neural network

网络的输入层主要包括通过点云特征计算得到的 4 种激光点云的原始信息和由 4 种点云特征信息组成的大小为 $64 \times 512 \times 8$ 的三维图像矩阵。相较于 SqueezeSeg V2 原始网络的输入数据,本文的点云多特征三维图像矩阵可以在输入端提供高级语义信息。网络的特征提取层和上采样层基本采用了 SqueezeSeg V2 网络结构,在特征提取时主要利用 Fire 模块^[21]提取特征,上下文聚合模块(CAM)^[12]用以降低对噪声点的敏感度并扩大感受野。在上采样时利用 Fire Deconv 模块^[10]来恢复空间信息。最后输出层根据上采样层的数据输出一张结果图像,其长宽与输入层的三维矩阵相同,不同的目标用不同的颜色进行渲染,其余部分以强度值填充,最终输出分割后的结果图。

4 实验结果及对比分析

本文中使用的数据来自 KITTI 数据集^[22],以该数据集中的三维点云作为基础,计算出每个点的角度和曲率等特征,并将点云及特征进行前视球面投影得到的点云多特征图像作为本文的数据。KITTI 数据集只在训练数据集中提供了详细的标签信息,因此本文将训练数据集分为两个部分,将 6000 帧点云作为训练数据集,1481 帧点云作为测试数据集。本文的实验环境是单 gtx1080ti 显卡、ubuntu18.04,cuda9 和 CUDNN V7。Batch size 设

置为 24,经过 50000 次训练得到训练好的模型。本文方法及 SqueezeSeg V2^[12]的分割结果对比如图 8 所示,相较于 SqueezeSeg V2 网络^[12],本文方法对汽车的分割效果稍好一些,而对行人和自行车的分割效果则有较大优势。

在精度评价方面,本文选择交并比(IoU)、精确率(p)和召回率(r)作为点云分割性能的评价指标。交并比代表网络预测出的目标区域与真实目标区域的重叠度,是二者的交集部分与并集部分的比值;精确率是指在网络的预测结果中预测正确的比例;召回率是指所有目标中被正确监测到的比例。在理想情况下,精确率和召回率的值越大越好。但是实际情况下,二者的关系是此消彼长的,想要全面评估模型的精确率就不能只关注其中的一个值而忽略另一个值,必须同时关注精确率和召回率。因此,这里引入一个新指标 F_{score} ^[23]来平衡二者的关系。 $F_{score} = 1 + \beta^2 \frac{pr}{\beta^2 p + r}$,这里的 β 表示的是召回率相较于精确率更重要的一个倍率值,如果召回率相较于精确率更加重要的话,则将 β 的值设置为大于 1 的值,反之则设置为小于 1 的值。本文认为召回率和精确率同等重要,因此将 β 值设置为 1,即可得到 $F_{1-score} = 2 \frac{pr}{p+r}$ 。本文方法与 SqueezeSeg V2、PointSeg 的评价指标对比如表 1 所示。

由于文献[12]只提供了交并比这一评价指标,



图 8 本文方法与 SqueezeSeg V2、PointSeg 网络的语义分割效果图对比以及对应场景的真值和图像
Fig. 8 Comparison of semantic segmentation effect among the proposed method with SqueezeSeg V2 and PointSeg, as well as the ground truth and images of the corresponding scene images

表 1 本文方法与 SqueezeSeg V2、PointSeg 的评价指标对比

Table 1 Comparison of evaluation indices of the proposed method, SqueezeSeg V2 and PointSeg

Method	Car				Cyclist				Pedestrian			
	IoU	p	r	$F_{1-score}$	IoU	p	r	$F_{1-score}$	IoU	p	r	$F_{1-score}$
SqueezeSeg V2	72.6	81.8	86.7	84.2	30.5	45.6	48	46.8	22.7	59.5	26.9	37
PointSeg	67.4	74.8	92.3	82.6	32.7	41.4	59.7	48.3	19.2	41.4	29.3	36.5
Proposed method	76.3	82.1	91.5	86.5	57	67	79.1	72.5	65.9	74	85.8	79.5

无法直接与本文结果进行全方位对比,因此本文按照文献[12]中的方法计算出了除交并比外的其他指标。如表 1 所示,在本文的车辆语义分割评价指标中,交并比达到了 76.3%,精确率达到了 82.1%,召回率达到了 91.5%, $F_{1-score}$ 达到了 86.5%。在自行车语义分割评价指标中,交并比达到了 57%,精确率达到了 67%,召回率达到了 79.1%, $F_{1-score}$ 达到了 72.5%;在本文的行人语义分割评价指标中,交并比达到了 65.9%,精确率达到了 74%,召回率达到了 85.8%, $F_{1-score}$ 达到了 79.5%。

从表 1 可以看出:在车辆的语义分割中,本文方法除了在召回率上比 PointSeg 网络稍低外,其他各项都比 SqueezeSeg V2 和 PointSeg 网络方法略高;而在行人和自行车的检测方面,本文方法在每项评价指标上相较于 SqueezeSeg V2 和 PointSeg 网络都有着更好的检测效果。此外,经实测可知,本文方法可以在单 gtx1080ti 显卡上达到 14.16 ms/frame 的速度,比 SqueezeSeg V2 网络的 11.39 ms/frame

慢一些,但优于 PointSeg 网络的 27.44 ms/frame,如图 9 所示。

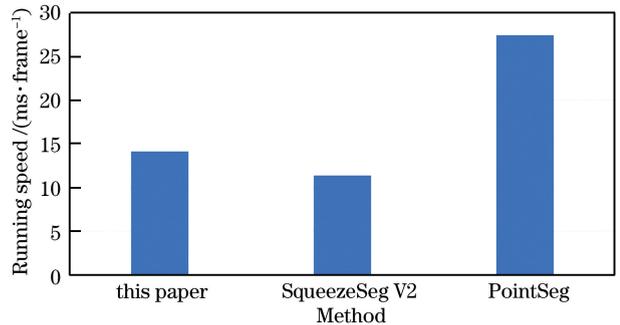


图 9 运行性能的对比

Fig. 9 Comparison of running performance

5 结 论

本文将点云数据进行前视投影得到三维点云多特征图像矩阵,解决了点云的无序性问题,使得神经网络模型可以像处理图像一样处理点云投影后的图

像矩阵。与此同时,本文计算出了距离、角度和曲率等点云特征数据,并将其生成的点云多特征图像矩阵直接输入到神经网络中,实现了传统点云处理中的特征信息和深度学习技术的有效结合,既解决了传统点云特征分割方法中经验阈值设置的问题,又提升了基于神经网络的点云语义分割网络的分割精度。结果显示:本文将角度和曲率等点云特征信息加到神经网络的输入层,可使神经网络在针对行人和自行车等复杂目标时能够提取到更多的特征信息,相较于其他无特征信息的语义分割神经网络模型(SqueezeSeg V2等)在点云语义分割的召回率和精确率等指标上有着更好的表现。

参 考 文 献

- [1] Xie D S, Xu Y C, Wang R D, et al. Obstacle detection and tracking for unmanned vehicles based on 3D laser radar [J]. *Automotive Engineering*, 2018, 40(8): 952-959.
谢德胜, 徐友春, 王任栋, 等. 基于三维激光雷达的无人车障碍物检测与跟踪 [J]. *汽车工程*, 2018, 40(8): 952-959.
- [2] Douillard B, Underwood J, Kuntz N, et al. On the segmentation of 3D LIDAR point clouds [C] // 2011 IEEE International Conference on Robotics and Automation, May 9-13, 2011. Shanghai, China: IEEE, 2011: 2798-2805.
- [3] He Y F, Hu R, Zou J G. Analysis on point cloud of building facade segmentation algorithms based on 3D laser scanning data [J]. *Bulletin of Surveying and Mapping*, 2019(4): 26-31.
贺亦峰, 胡荣, 邹进贵. 利用三维激光扫描数据进行建筑物立面点云分割算法分析 [J]. *测绘通报*, 2019(4): 26-31.
- [4] Huang K, Cheng X J, Jia D F, et al. An automatic segmentation algorithm for dense pipeline point cloud data [J]. *Chinese Journal of Lasers*, 2018, 45(11): 1104004.
黄凯, 程效军, 贾东峰, 等. 一种密集管道点云数据自动分割算法 [J]. *中国激光*, 2018, 45(11): 1104004.
- [5] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [C] // *Proceedings of the 25th Conference and Workshop on Neural Information Processing Systems*. [S.l.:s.n.], 2012: 1097-1105.
- [6] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition [C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016. Las Vegas, NV, USA: IEEE, 2016: 770-778.
- [7] 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: IEEE, 2017: 652-660.
- [8] Qi C R, Yi L, Su H, et al. PointNet++: deep hierarchical feature learning on point sets in a metric space [C] // *Advances in Neural Information Processing Systems*. Long Beach, USA: [s. n.], 2017: 5099-5108.
- [9] Simon M, Milz S, Amende K, et al. ComplexYOLO: an Euler-region-proposal for real-time 3D object detection on point clouds [M] // *Lecture notes in computer science*. Cham: Springer International Publishing, 2019: 197-209.
- [10] Wu B C, Wan A, Yue X Y, et al. SqueezeSeg: convolutional neural nets with recurrent CRF for real-time road-object segmentation from 3D LiDAR point cloud [C] // 2018 IEEE International Conference on Robotics and Automation (ICRA), May 21-25, 2018. Brisbane, QLD: IEEE, 2018: 1887-1893.
- [11] Wang Y, Shi T Y, Yun P, et al. PointSeg: real-time semantic segmentation based on 3D LiDAR point cloud [J/OL]. <https://arxiv.org/abs/1807.06288>.
- [12] Wu B C, Zhou X, Zhao S C, et al. SqueezeSeg V2: improved model structure and unsupervised domain adaptation for road-object segmentation from a LiDAR point cloud [C] // 2019 International Conference on Robotics and Automation (ICRA), May 20-24, 2019. Montreal, QC, Canada: IEEE, 2019: 4376-4382.
- [13] Bogoslavskyi I, Stachniss C. Efficient online segmentation for sparse 3D laser scans [J]. *PFG-Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, 2017, 85(1): 41-52.
- [14] Fan X H, Xu G L, Li W L, et al. Target segmentation method for three-dimensional LiDAR point cloud based on depth image [J]. *Chinese Journal of Lasers*, 2019, 46(7): 0710002.
范小辉, 许国良, 李万林, 等. 基于深度图的三维激光雷达点云目标分割方法 [J]. *中国激光*, 2019, 46(7): 0710002
- [15] Li B, Zhang T L, Xia T. Vehicle detection from 3D lidar using fully convolutional network [J/OL]. <https://arxiv.org/pdf/1608.07916.pdf>.
- [16] Halterman R, Michael B. Velodyne HDL-64E lidar for unmanned surface vehicle obstacle detection [J]. *Proceedings of SPIE*, 2010, 7692: 76920D.
- [17] He M F, Zhou L S, Shen H C. Curvature estimation

- of scattered-point cloud data and its application[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2005, 37(4): 515-519.
- 贺美芳, 周来水, 神会存. 散乱点云数据的曲率估算及应用[J]. 南京航空航天大学学报, 2005, 37(4): 515-519.
- [18] Zeng B, Huang W. An indoor point cloud segmentation method fusing with multi-feature cluster ensemble[J]. Computer Engineering, 2018, 44(3): 281-286.
- 曾碧, 黄文. 一种融合多特征聚类集成的室内点云分割方法[J]. 计算机工程, 2018, 44(3): 281-286.
- [19] Cai T J, Tang H. Overview of the principle of least square fitting of Sabitzky-Golay smooting filters[J]. Digital Communication, 2011, 38(1): 63-68, 82.
- 蔡天净, 唐瀚. Savitzky-Golay 平滑滤波器的最小二乘拟合原理综述[J]. 数字通信, 2011, 38(1): 63-68, 82.
- [20] Yang P, Qian X. Direct computing of surface curvatures for point-set surfaces[C] // Symposium on Point Based Graphics. Prague, Czech Republic: [s. n.], 2007: 29-36.
- [21] Iandola F N, Han S, Moskewicz M, et al. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size[J/OL]. <https://arxiv.org/abs/1602.07360>.
- [22] Geiger A, Lenz P, Urtasun R. Are we ready for autonomous driving? The KITTI vision benchmark suite [C] // 2012 IEEE Conference on Computer Vision and Pattern Recognition, June 16-21, 2012, Providence, RI: IEEE, 2012: 3354-3361.
- [23] Evaluation of clustering [EB/OL]. (2009-04-07) [2019-11-20]. <https://nlp.stanford.edu/IR-book/html/htmledition/evaluation-of-clustering-1.html>.