

# 基于 3D 激光雷达的水面目标检测算法研究

周治国\*, 李怡瑶, 曹江微, 邸顺帆

北京理工大学信息与电子学院, 北京 100081

**摘要** 3D 激光雷达是水面无人艇(USV)的关键感知模块,水域杂波的干扰会降低模块的目标检测能效,影响自主导航避障功能。提出一种基于 3D 激光雷达的水面目标 DBSCAN-VoxelNet 联合检测算法。该算法采用具有噪声的密度聚类方法(DBSCAN)滤除水面杂波干扰;基于 VoxelNet 深度神经网络将水面稀疏点云数据划分为体素形式,并将结果输入哈希表进行高效查询;经特征学习层提取特征张量,并将张量输入卷积层获取全局目标信息,实现高精度目标检测。实验结果表明,所提联合检测算法对水域杂波有较好的抑制性能,平均精度均值(mAP)为 82.4%,有效提高了对水面目标的检测精度。

**关键词** 水面无人艇; 目标检测; 3D 激光雷达; 密度聚类方法; VoxelNet

中图分类号 TP183

文献标志码 A

DOI: 10.3788/LOP202259.1815006

## Surface Target Detection Algorithm Based on 3D Lidar

Zhou Zhiguo\*, Li Yiyao, Cao Jiangwei, Di Shunfan

School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China

**Abstract** A three-dimensional (3D) lidar is the main sensing module of an unmanned surface vehicle (USV). The interference of water clutter will reduce the energy efficiency of target detection and affect autonomous navigation's obstacle avoidance function. Based on 3D lidar, this study proposes a surface target DBSCAN-VoxelNet joint detection algorithm. The proposed algorithm employs a noise density clustering approach (DBSCAN) to filter surface clutter interference; a depth neural network VoxelNet is employed to divide surface sparse point cloud data into voxels, and the results are input into a Hash table for efficient query; the feature tensor is extracted through the feature learning layer, and the tensor is input into the convolution layer to obtain the global target information, resulting in high-precision target detection. The experimental results reveal that the proposed joint detection algorithm performs well in suppressing clutter in the water area, with a mean average precision (mAP) of 82.4%, which effectively enhances surface target detection accuracy.

**Key words** USV; target detection; 3D lidar; density clustering method; VoxelNet

## 1 引言

水面无人艇(USV)作为一种自主的水面作业航行器,凭借着智能化、无人化、灵活性的特点,应用与发展已初具规模。USV 在高动态环境中自主航行并完成任务,需要具备感知环境、发现潜在威胁及执行合理路径的能力<sup>[1]</sup>。其中,环境感知是 USV 避障的先决条件,而能够自主准确实时地检测环境中的障碍物又是其主要任务之一。在 USV 的环境感知领域发展中,主要出现了基于雷达、声呐、视觉、激光雷达等多种不同传感器的目标检测算法。但水面环境存在目标难以采

样、形变、光照变化、尺度变化及摄像头抖动等多种情况,加大了视觉等传感器的检测跟踪难度。激光雷达常用于对目标物的探测和跟踪,环境中目标物与雷达距离、反射率、散射形状不同,导致反射的激光脉冲参数不同,通过接收并测量反射回波,可以得到目标物的位置和反射率等信息<sup>[2]</sup>。与基于图像的检测相比,激光雷达提供了可靠的深度信息,可用于精确定位目标并表征其形状。

相比于 2D 激光雷达缺乏高度信息无法成像,3D 激光雷达能够还原物体的形状大小和空间的三维信息,可以实现动态实时成像。得益于无人平台领域的

收稿日期: 2021-06-23; 修回日期: 2021-07-02; 录用日期: 2021-07-27

基金项目: “十三五”装备预研领域基金(61403120109)

通信作者: zhiguo@bit.edu.cn

快速发展和标准数据集 KITTI 的助力,基于点云的 3D 目标检测方法近年来得到快速发展,按照对激光点云处理的网络输入格式,这些方法主要分为三类:点云(Pointcloud)、图像(Image)、体素(Voxel)。对于直接对点云数据进行操作的方法,文献[3]提出一种直接应用于点云数据的深度学习网络 PointNet,该网络可直接从点云中逐点学习特征,获得目标的分类和运动信息,在目标检测、目标分类和语义分割等场景中得到广泛应用。PointNet 对所有点云数据提取全局特征,和目前主流的卷积神经网络(CNN)逐层提取局部特征的方式不同,因此文献[4]提出了改进型网络 PointNet++,该网络能够在不同尺度提取局部特征,通过多层网络结构得到深层特征,使网络能够学习不同规模的数据。对于输入为点云投影后的图片格式的方法,文献[5-6]将激光雷达获取到的点云投影到透视图并应用基于图像的特征提取技术。文献[7]将点云栅格化后利用手工特征提取方法对其进行编码。文献[8-9]以 2D 点云检测网络为框架,融合图像信息,从而丰富点云特征。目前较先进的网络如 MVnet<sup>[10]</sup>,首先将点云栅格化并构建俯视图(Bird Eye View),之后采用成熟的图像网络对点云按图像的方式进行处理。对于输入为体素格式的方法,将给定的三维空间均匀分成很多小方格,用于激光雷达返回的数据为点云格式,要想转为体素格式需要预先确定体素大小。文献[11]在 2015 年提出了 VoxNet 网络结构,通过对点云数据体素化,再结合三维卷积将激光点云数据用于网络训练,为基于点云的目标识别开启了新思路。文献[12]用体素网络预测激光雷达点云数据的 3D 边界框。文献[13]结合体素网络和稀疏卷积<sup>[14]</sup>进一步提高了检测任务的实时性和有效性。近年来,随着自动

驾驶技术的蓬勃发展,基于 Voxel 的激光雷达目标检测受到广泛关注<sup>[15-20]</sup>。

USV 的速度较慢,在航行控制上机动性相对于车平台较差,在高海况条件下行动能力受到较大制约;在行驶稳定性上,风浪导致的摇晃引起俯仰角、侧倾角和偏航角变动,USV 的姿态改变会给激光雷达的检测带来困难。在检测水面目标时,除了会受到场景中其他光源影响或者其他物体反射产生的噪声点影响,还有水面风浪引起的杂波点影响。

由于水面点云具有高度的稀疏性和密度不均匀性,基于点云的检测方法的网络结构会有较高的计算和内存需求,实时性较差,难以满足 USV 自主航行的感知需求。基于图像的检测方法将三维点云投影后无法有效利用三维空间信息,且特征提取过程中存在信息瓶颈,即使融合视觉信息,但仍依赖于严格的时间同步和视觉传感器的校准,在实际系统中无法满足要求。本文针对 USV 在水面环境中的感知需求采集激光雷达实船数据,开展基于深度学习的 USV 水面目标检测算法研究,采用体素格式降低点云数据的计算量级,对比水面与陆地点云数据的异同,降低水面杂波影响,提出具有噪声的密度聚类方法(DBSCAN)-VoxelNet 联合解决方案,在保证检测速度的同时有效提高对水面目标的检测精度。

## 2 算法原理

相比于无人车地面环境,水面杂波较少,密集程度远低于地面目标,行驶约束与规则较少,如无特殊要求,水域范围内皆可作为可通行区域,更适于激光雷达检测。地面点云数据和水面点云数据如图 1 所示。

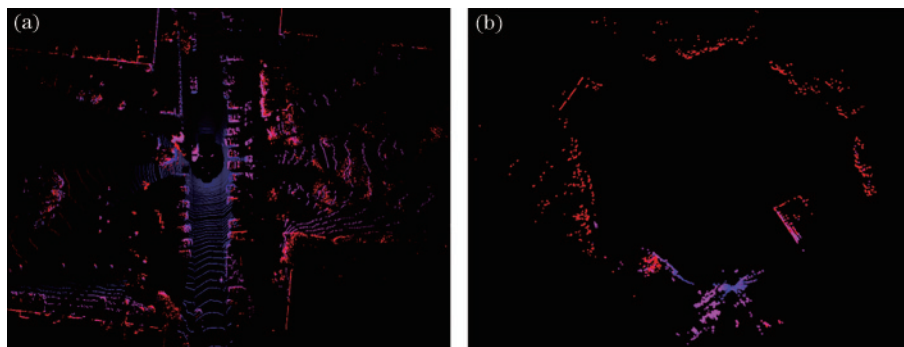


图 1 不同场景下的激光雷达点云数据。(a)地面点云;(b)平静水面点云

Fig. 1 Lidar point clouds in different scenes. (a) Ground point clouds; (b) calm water surface point clouds

对于面向风浪条件下的水面稀疏点云数据,检测算法框图如图 2 所示。通过 DBSCAN 滤除波浪噪声后将数据划分为体素形式,并对其进行分组和采样,降低数据量级,节省计算资源,提高对水面目标的检测速度。以 VoxelNet 深度学习网络为基本框架,采用体素特征编码(VFE)层实现预处理后点云的逐点特征和局部聚集特征的连接,经过多个 VFE 层后网络即可学

习到每个体素内的特征表示;三维卷积进一步聚合体素特征,将点云数据转换为高维张量后输入卷积网络,最终区域推荐网络产生高精度检测结果。

### 2.1 水面杂波干扰抑制

如图 3 所示,在同一湖泊水域不同时刻进行实船测试,发现平静的水面几乎不会产生激光回波或只产生孤立的噪声点,但是风浪条件下浪花与泡沫会将激

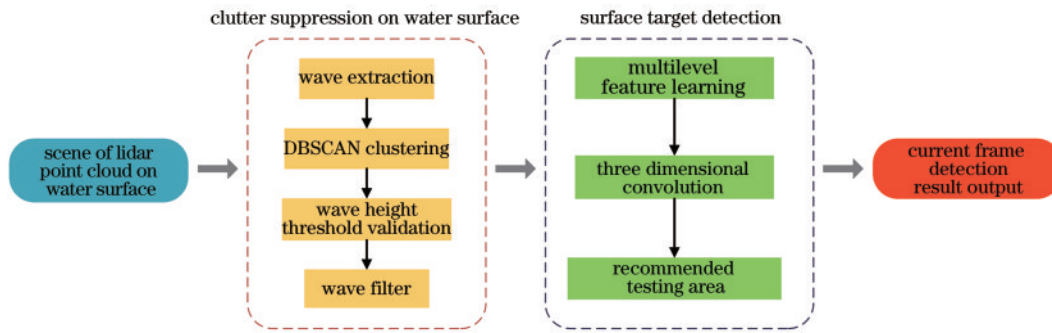


图 2 算法框图

Fig. 2 Algorithm block diagram

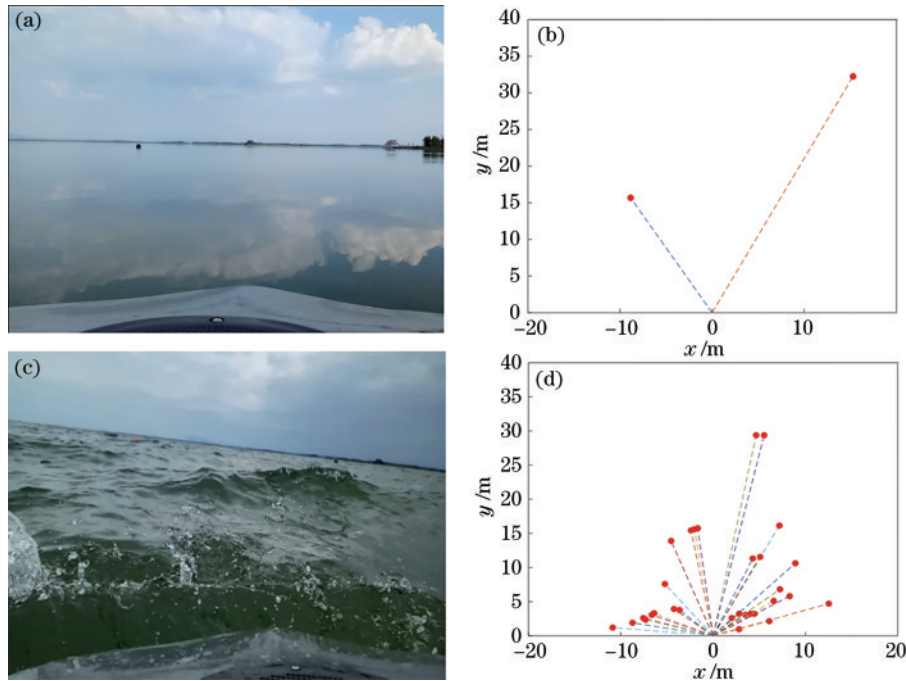


图 3 激光雷达水面回波数据。(a)(b)平静水面；(c)(d)风浪水面

Fig. 3 Lidar water surface echo data. (a) (b) Calm water surface; (c) (d) wave water surface

光束部分反射至接收器,而且在近岸、岛礁边缘等深度极浅的水域也存在着回波数据。这些来自水面的点云信息给目标检测带来了极大干扰,需要对无效点云进行滤除。

激光雷达风浪杂波点范围大、密度高,可以采用基于密度的聚类方法对其聚类整合后选定滤除。DBSCAN、OPTICS、DENCLUE 等算法属于基于密度的聚类方法,假设聚类结构可以通过样本分布的紧密程度确定,那么就主要考察样本密度之间的可连接性,然后根据可连接样本再扩大聚类簇获得聚类结果。其中,DBSCAN 算法的应用最广,由于对集中区域的效果较好,为了发现任意形状的簇,这类方法将簇看作是数据空间中被低密度区域分割开的稠密对象区域,将具有足够高密度的区域划分为簇,并在具有噪声的空间数据中能发现任意形状的簇。

在 USV 平台的激光雷达目标检测中,波浪噪声数量是未知的,而且算法的较高检测频率、距离测量精度

和水平分辨率保证了相邻点之间均匀的几何距离。DBSCAN 算法对密度相对均匀且相邻点之间距离较小的样本有着很好的效果,而且对于水面数据集中的异常值,聚类后可直接筛选出噪声点并进行滤除。与 K-means 等基于距离划分的聚类方法相比, DBSCAN 算法虽然在计算复杂度上略高,但是不需要对类别数量做出先行设置,可以对任意形状的样本分布进行聚类,所以适用于激光雷达波浪点云聚类。

DBSCAN 算法的具体实现流程为

- 1) 首先寻找任意一个点云数据样本点,将其作为核心点,根据设定的邻域范围  $\epsilon$  与邻域内样本数量  $N_{minpts}$  参数,确定与该点全部密度相连的点云;
- 2) 遍历此核心点  $\epsilon$  邻域内的其他点,寻找与这些点密度相连的样本,把非核心的点云数据作为边界点;
- 3) 重复步骤 1) 和 2),直到访问完所有核心点,把没有归为任何一簇的样本作为水面噪声点。

但是激光雷达的数据特点导致不同距离的目标点



云间隔不等,波浪距离较近时点云分布密集,相邻点间距小,而距离较远时点云数据稀疏,相邻点间距大。所以为避免波浪目标的过分割与欠分割,需要在实际的数据处理中对不同距离上的目标设置不同的分割阈值。

影响分割结果的参数是  $\epsilon$  与  $N_{\text{minpts}}$ , 以等距相邻点

的弧长  $l$  为参考, 设置分割参数:

$$\begin{cases} \epsilon = 2d\sigma \\ N_{\text{minpts}} = 3 \end{cases} \quad (1)$$

式中:  $d$  为目标距激光雷达的直线距离;  $\sigma$  为激光雷达水平分辨率。具体的 DBSCAN 算法描述如图 4 所示。

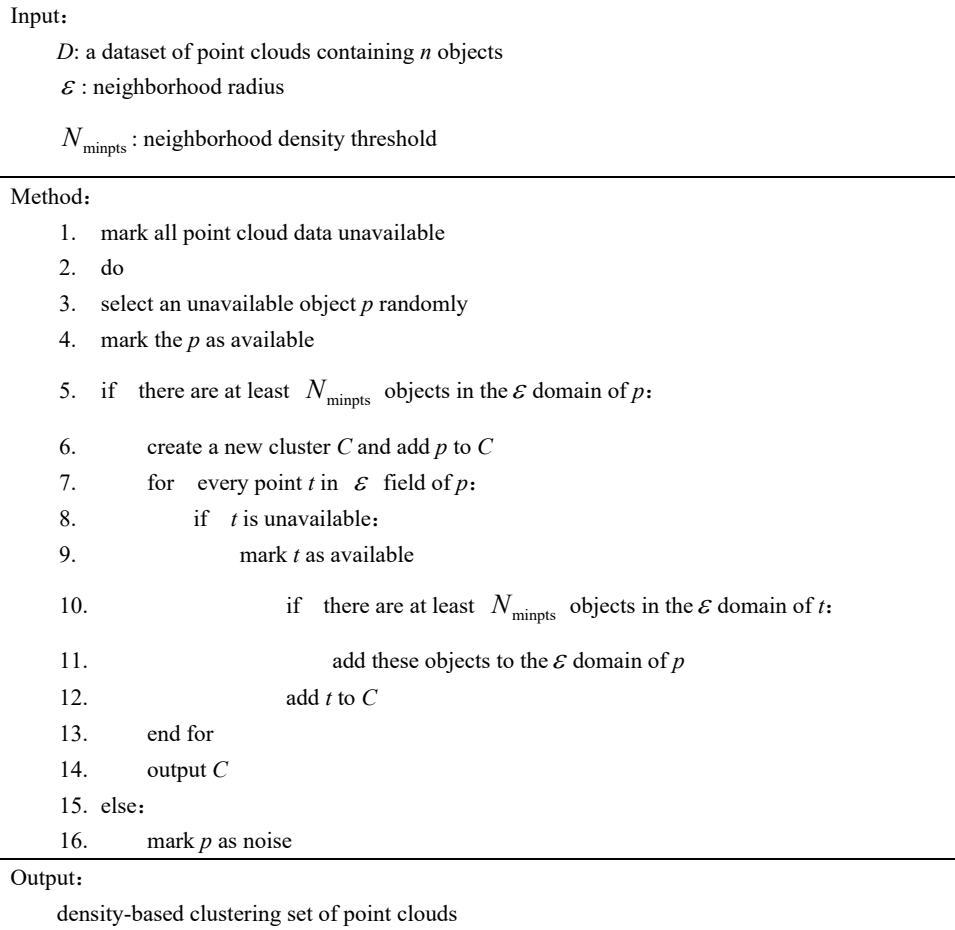


图 4 水面点云 DBSCAN 滤波算法

Fig. 4 DBSCAN filtering algorithm based on water surface point clouds

湖泊、水库中的波浪高度比海洋中的小,一般随着风的出现而立即产生,也能随着风的停止而迅速消失,其波高为 0~10 cm。因此对聚类结果范围进行计

算,设置波高阈值为 0.1 m,将该范围内的点云滤除即可去除波浪杂波。波浪水面及聚类结果如图 5 所示。

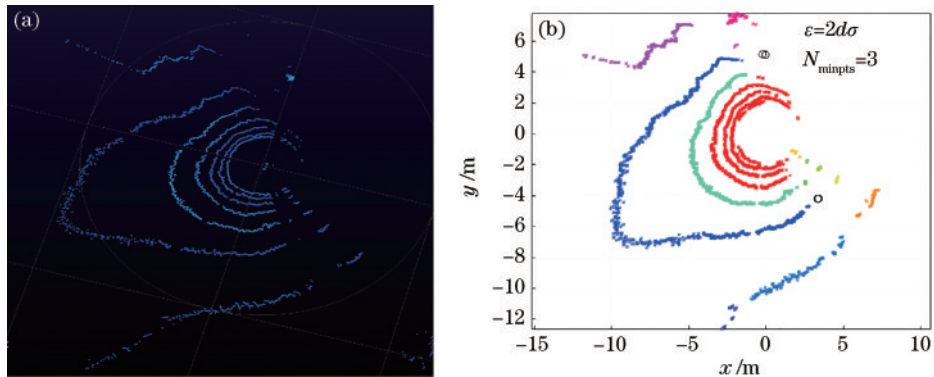


图 5 DBSCAN 聚类。(a)波浪水面点云数据;(b)聚类结果

Fig. 5 DBSCAN clustering. (a) Wave point clouds over water; (b) clustering result

### 2.2 点云目标检测网络

VoxelNet 是一个端到端的点云目标检测网络,可以直接处理稀疏的 3D 点云数据,自动学习到对应的目标特征,并利用哈希表实现点云的高效查询,从而避免

信息瓶颈,完成多尺度检测任务。VoxelNet 包含三个主要功能模块:特征学习层、卷积中间层以及区域推荐层。VoxelNet 算法描述如图 6 所示。

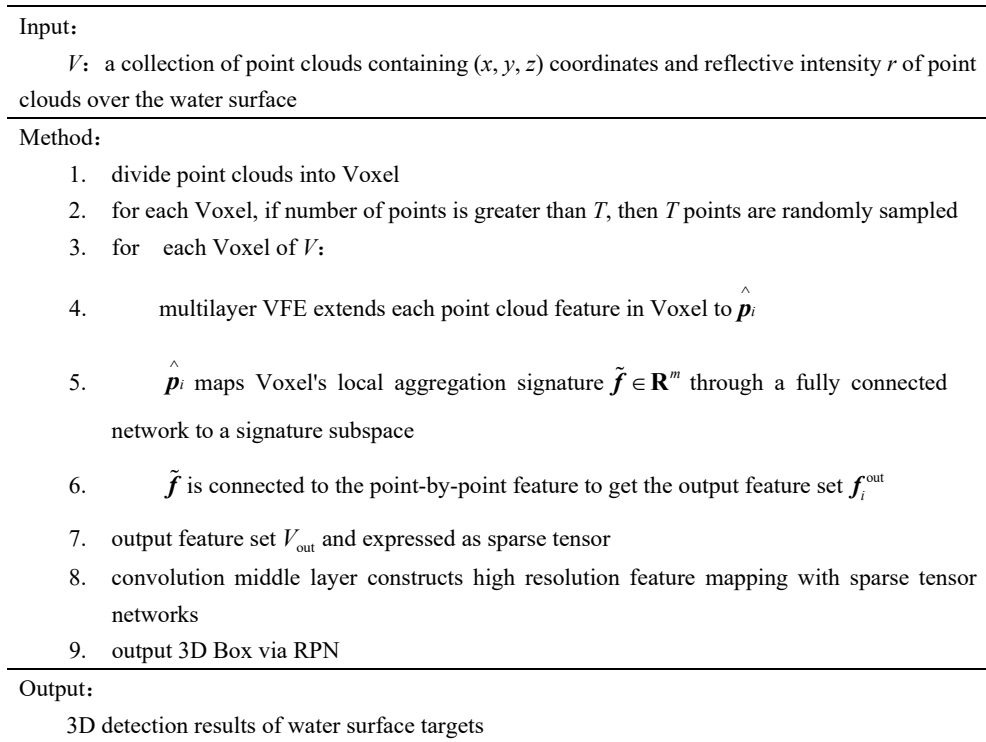


图 6 水面目标检测算法

Fig. 6 Water surface target detection algorithm

#### 2.2.1 特征学习层

特征学习层的结构如图 7 所示,包括体素分割、点

云分组、随机采样、多层体素特征编码(S-VFE)和稀疏张量表示等步骤。

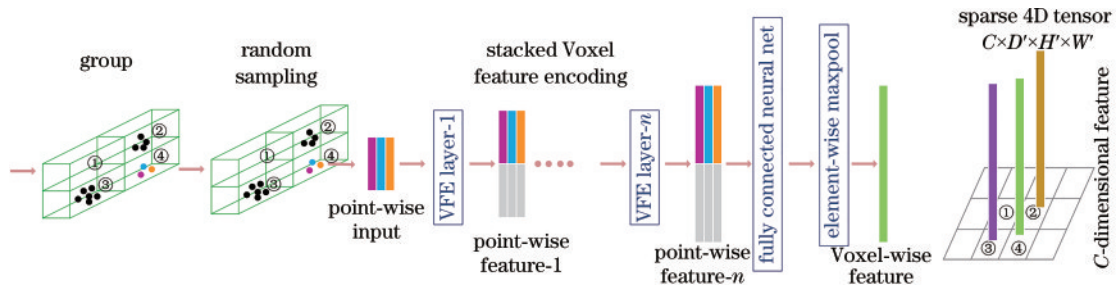


图 7 特征学习网络结构

Fig. 7 Structure of feature learning network

对于滤除波浪噪声后的点云,使用一个深度、高度、宽度分别为  $D, H, W$  的立方体表示整个 3D 检测空间,相应地用尺寸为  $v_D, v_H, v_W$  的立方体对空间进行分割,则点云数据三维体素化后分别在  $X, Y, Z$  轴上生成的体素网格个数为  $D' = \frac{D}{v_D}, H' = \frac{H}{v_H}, W' = \frac{W}{v_W}$ 。分割完成后对点云按照所在体素位置进行分组,受目标距离、相对姿态、遮挡和非均匀采样等因素的影响,体素点云数据稀疏且高度不均匀,目标所在位置的体素格数远大于水面体素格,为减小计算机内存消耗和检测

偏差,对点云密集的体素随机采样  $T$  个数据点。

S-VFE 是特征学习层的主要网络结构,可以自动学习到每个体素内的特征表示。以第一个 VFE 为例,结构如图 8 所示。

非空体素输入  $V = \{p_i = [x_i, y_i, z_i, r_i]^T \in \mathbf{R}^4\}_{i=1, \dots, T}$  是一个体素格内随机采样的点云集,其中  $(x_i, y_i, z_i, r_i)$  分别表示点云的  $X, Y, Z$  轴坐标和激光束的反射强度。首先计算体素内所有点的平均值  $(v_x, v_y, v_z)$  作为体素格的形心,随后将体素格内所有点云的特征形式扩展为

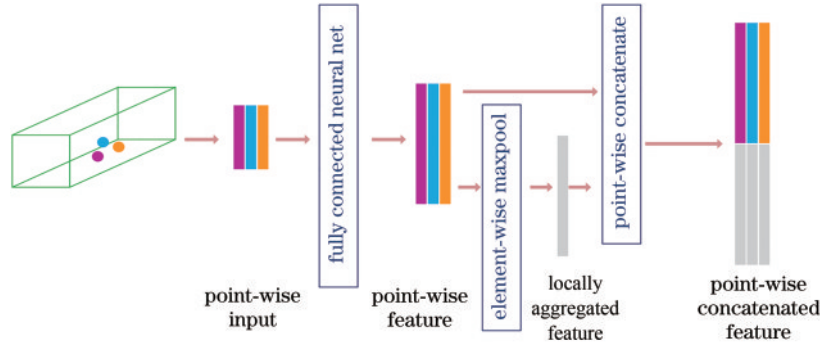


图 8 VFE 层结构

Fig. 8 VFE layer structure

$$V = \{ \hat{\mathbf{p}}_i = [x_i, y_i, z_i, r_i, x_i - v_x, y_i - v_y, z_i - v_z]^T \in \mathbf{R}^7 \}_{i=1, \dots, t \leq T} \quad (2)$$

体素内每一个  $\hat{\mathbf{p}}_i$  都将通过一个包含线性映射、批标准化(BN)以及线性整流(ReLU)的全连接网络(FC),得到逐点的特征表示,即被映射到特征空间  $\mathbf{f}_i \in \mathbf{R}^m$  中,特征空间的输入特征维度为 7,输出特征维度为  $m$ 。随后对  $\mathbf{f}_i \in \mathbf{R}^m$  采用最大池化操作,得到体素  $V$  的局部聚合特征  $\tilde{\mathbf{f}} \in \mathbf{R}^m$ ,将逐点特征与局部聚合特征连接,得到输出特征  $\mathbf{f}_i^{\text{out}} = [\mathbf{f}_i^T, \tilde{\mathbf{f}}^T]^T \in \mathbf{R}^{2m}$ ,则输出的特征集合为  $V_{\text{out}} = \{ \mathbf{f}_i^{\text{out}} \}_{i=1, \dots, t \leq T}$ 。

对所有的非空体素进行体素特征编码操作,并共享全连接网络参数,若  $(c_{\text{in}}, c_{\text{out}})$  为特征维度变化,则全连接层的参数矩阵大小为  $c_{\text{in}} \times (c_{\text{out}}/2)$ 。由于输出包含了逐点特征和局部聚合特征,经过 S-VFE 后,网络可以自动学习到每个体素内的特征表示,最终通过 FC 和最大池化将第  $n$  个 VFE 的输出映射到  $\mathbf{R}^C$ ,从而可以得到一系列体素特征,表示为四维张量形式:

$$C \times D' \times H' \times W' \quad (3)$$

水面激光雷达数据稀疏,大部分体素为空,将非空体素特征表示为稀疏张量可有效降低网络反向传播时

的内存和计算消耗。

### 2.2.2 卷积中间层

卷积中间层包含一个三维卷积、BN 和 ReLU,可表示为  $\text{ConMD}(c_{\text{in}}, c_{\text{out}}, \mathbf{k}, \mathbf{s}, \mathbf{p})$ ,其中  $M$  表示卷积操作的维度; $c_{\text{in}}$  和  $c_{\text{out}}$  分别表示输入和输出的通道数; $\mathbf{k}, \mathbf{s}, \mathbf{p}$  均为向量,分别表示卷积操作的卷积核、步长、填充尺寸,对一个三维卷积而言,  $\mathbf{k} = (k, k, k)$ 。

对于水面 USV 目标检测,设计的卷积中间层为

$$\begin{cases} \text{Conv3D}(128, 64, 3, (2, 1, 1), (1, 1, 1)) \\ \text{Conv3D}(64, 64, 3, (1, 1, 1), (0, 1, 1)) \\ \text{Conv3D}(64, 64, 3, (2, 1, 1), (1, 1, 1)) \end{cases} \quad (4)$$

### 2.2.3 RPN 层

区域推荐网络(RPN)<sup>[21]</sup>是用于目标检测的优化算法,该网络与检测网络共享全图像的卷积特征,可以同时每个位置预测目标边界和目标分数,生成高质量的区域提出。在 VoxelNet 中的 RPN 结构如图 9 所示,包含 3 个全卷积块,每个块的第一层通过步长为 2 的卷积操作将特征图下采样为原来的 1/2,随后是一系列步长为 1 的卷积操作。

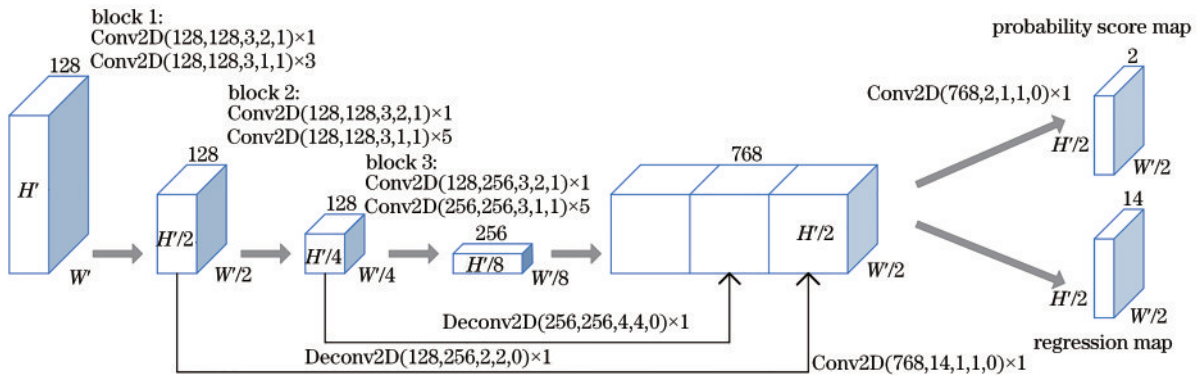


图 9 RPN 结构

Fig. 9 Region proposal network architecture

每个卷积层都包含 BN 和 ReLU 操作,将输出上采样到固定大小并串联,构造高分辨率特征图谱,最后

该特征图通过两种二维卷积映射到概率评分图和回归图。

### 2.2.4 损失函数

定义  $\{a_i^{\text{pos}}\}_{i=1, \dots, N_{\text{pos}}}$  为  $N_{\text{pos}}$  个正样本集合,  $\{a_j^{\text{neg}}\}_{j=1, \dots, N_{\text{neg}}}$  为  $N_{\text{neg}}$  个负样本集合,  $(x_c^g, y_c^g, z_c^g, l^g, w^g, h^g, \theta^g)$  表示一个真实的 3D 标注框, 其中  $(x_c^g, y_c^g, z_c^g)$  是标注框中心坐标,  $l^g, w^g, h^g$  是其长、宽、高,  $\theta^g$  是绕 Z 轴的偏航角, 相应地用  $(x_c^a, y_c^a, z_c^a, l^a, w^a, h^a, \theta^a)$  表示正样本框。

真实标注框与正样本框的残差向量可表示为  $\mathbf{u}^* = (\Delta x, \Delta y, \Delta z, \Delta l, \Delta w, \Delta h, \Delta \theta) \in \mathbf{R}^7$ :

$$\begin{cases} \Delta x = \frac{x_c^g - x_c^a}{d^a} \\ \Delta y = \frac{y_c^g - y_c^a}{d^a} \\ \Delta z = \frac{z_c^g - z_c^a}{d^a} \\ \Delta l = \log \frac{l^g}{l^a} \\ \Delta w = \log \frac{w^g}{w^a} \\ \Delta h = \log \frac{h^g}{h^a} \\ \Delta \theta = \theta^g - \theta^a \end{cases}, \quad (5)$$

式中: 正样本框的对角线长度  $d^a = \sqrt{(l^a)^2 + (w^a)^2}$ , 利用  $d^a$  实现对  $\Delta x, \Delta y, \Delta z$  的归一化, 以便直接估计 3D 标注框, 由此定义的损失函数为

$$L = \alpha \frac{1}{N_{\text{pos}}} \sum_i L_{\text{cls}}(P_i^{\text{pos}}, 1) + \beta \frac{1}{N_{\text{neg}}} \sum_j L_{\text{cls}}(P_j^{\text{neg}}, 1) + \frac{1}{N_{\text{pos}}} \sum_i L_{\text{reg}}(\mathbf{u}_i, \mathbf{u}_i^*), \quad (6)$$

式中:  $P_i^{\text{pos}}$  和  $P_j^{\text{neg}}$  分别表示正样本和负样本的 Softmax 输出;  $\mathbf{u}_i$  和  $\mathbf{u}_i^*$  分别表示正样本回归输出与其残差向量; 损失函数的前两项是正样本和负样本的交叉熵分类损失;  $\alpha$  和  $\beta$  是用于衡量损失函数影响的权重;  $L_{\text{reg}}$  为 Smooth L1 函数, 表示回归损失。

## 3 实验验证

### 3.1 实验平台

#### 3.1.1 SpaitLab-USV120 智能无人艇

实验所用 SpaitLab-USV120 智能无人艇采用玻璃钢制造, 在保证安全稳定的前提下, 尽量减轻质量, 既轻便又灵活。艇体采用双体艇结构设计, 体型较宽, 正常情况下可以搭载 20 kg 除艇体本身基本系统外的其他系统装置, 艇尾采用双喷泵式结构作为驱动装置, 最高时速可达 2 m/s。艇体长为 120 cm, 艇宽为 80 cm, 高度为 60 cm, 自身质量为 60 kg, 如图 10 所示。USV 体积较小, 故可以方便穿梭在近海海域、湖泊、河流、水塘等各种水域采集目标数据。



图 10 实验平台

Fig. 10 Experimental platform

#### 3.1.2 RS-LiDAR-16 激光雷达

SpaitLab-USV120 智能 USV 搭载的激光雷达如图 10 所示, 为速腾聚创科技有限公司生产的 16 线激光雷达 RS-LiDAR-16。该激光雷达采用混合固态激光雷达方式, 集成了 16 个激光收发组件, 测量距离高达 150 m, 测量精度在  $\pm 2$  cm 以内, 出点数高达 320000 point/s, 水平测角为  $360^\circ$ , 垂直测角为  $-15^\circ \sim 15^\circ$ 。产品传感器规格如表 1 所示。

表 1 RS-LiDAR-16 传感器参数

Table 1 RS-LiDAR-16 sensor parameters

Performance	Parameter
Number of channels	TOF ranging 16 channels
Ranging	20 cm to 150 m (target reflectivity is 20%)
Accuracy	Within $\pm 2$ cm (typical value)
Vertical view	$\pm 15^\circ$ ( $30^\circ$ in total)
Vertical angular resolution	$2^\circ$
Horizontal perspective	$360^\circ$
Azimuth resolution	$0.09^\circ$ (5 Hz) to $0.36^\circ$ (20 Hz)
Rotation speed	$300/600/1200 \text{ rad} \cdot \text{min}^{-1}$ (5/10/20 Hz)

激光雷达的数据解析在 USV 中的嵌入式控制板进行, 数据采用 TCP/IP 协议传输, 在将工作模式等参数导入后, 开始启动数据采集流程, 逐帧读取点云信息。

### 3.2 实验结果分析

#### 3.2.1 实测数据

实验采用 SpaitLab-USV120 搭载 RS-LiDAR-16 激光雷达, 在湖面针对三种不同形状的立方体进行数据采集, 其中球体直径为 0.5 m, 三棱锥的高为 1 m, 圆柱的底面直径为 0.5 m, 高为 1 m。针对这三种尺寸目标物, USV 制动反应距离为 10 m, 因此在实验中, 分别选取 5, 10, 15 m 三种不同距离采集多份数据, 较为合理。图 11 给出了视觉传感器和 3D 激光雷达的采集数据。



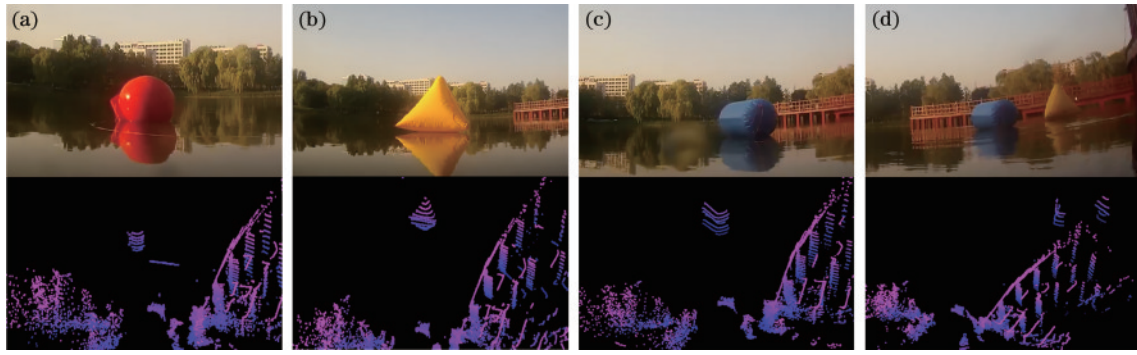


图 11 实测数据。(a)球体;(b)三棱锥;(c)圆柱;(d)多目标

Fig. 11 Measured data. (a) Spheres; (b) tri-pyramid; (c) cylindrical; (d) multi-objective

### 3.2.2 水面目标检测结果对比

实验环境配置如下:CPU为Intel Xeon E5-2678 v3;内存为10 G;GPU为NVIDIA GTX 2018Ti;操作系统为64位Ubuntu16.04 LTS;实验框架为PyTorch开源框架;编程语言为Python 3.6;第三方库OpenCV和CUDA10.0。在实验过程中,测试集数据共513张,验证集数据共169张。

模型训练过程中的训练损失如图12所示。可以发现,DBSCAN-VoxelNet可以很好地对水面目标检测任务进行拟合,训练损失在0~10 epoch下降速度很快,40 epoch之后下降速度降低,80 epoch后趋于稳定,可以很好地收敛于水面数据集。

采用平均精度均值(mAP)对模型的性能进行评估,目标在不同距离下的AP如表2所示,mAP如表3

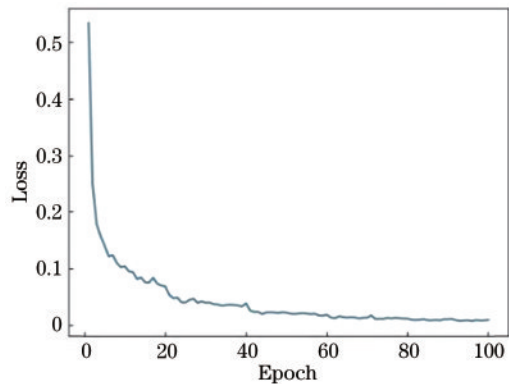


图 12 DBSCAN-VoxelNet 训练过程损失值

Fig. 12 DBSCAN-VoxelNet loss value in the training process 所示,从表中可以看出,DBSCAN-VoxelNet在保证检测速度的同时提高了对水面目标的检测精度。

表 2 DBSCAN-VoxelNet在水面目标数据集上的检测结果

Table 2 Detection results of DBSCAN-VoxelNet on water surface target dataset

AP <sub>spheres</sub>			AP <sub>tri-pyramid</sub>			AP <sub>cylindrical</sub>			AP <sub>multi-objective</sub>		
5 m	10 m	15 m	5 m	10 m	15 m	5 m	10 m	15 m	5 m	10 m	15 m
0.876	0.865	0.793	0.881	0.873	0.866	0.858	0.852	0.844	0.812	0.798	0.776

表 3 水面目标检测的 mAP 结果

Table 3 mAP detection results for water surface target

Parameter	VoxelNet	DBSCAN-VoxelNet
mAP	0.812	0.824
Average detection speed / (frame·s <sup>-1</sup> )	0.07	0.08

实验过程中,利用了消融实验的思想,首先从距离层面分析,三种形状目标分别展示了不同的效果。其中球形目标受距离影响较大,当目标距离激光雷达越远时,有用信息模糊的同时引入了很多噪声,尽管PRN层提高了对小目标识别的鲁棒性,但15 m的效果与10 m与5 m相比差距较大;算法对三棱锥目标和圆柱目标在不同距离上的鲁棒性较球体要强一些,目标反射面形状较为稳定,在远距离时仍能保持端点信息,但仍会随着距离的增加测试精度降低。其次从形状来分析,为了避免其他信息的干扰,选择5 m的距离进行

分析,从实验结果上看,对网络检测精度从高到低的物体排序,依次为三棱锥、球体和圆柱。三棱锥的特征最显著,在检测过程中DBSCAN-VoxelNet网络模型可以有效提取目标特征,与周边环境中的水草和桥墩等干扰信息进行区分;球体和圆柱体存在类似的曲面特征,且易被桥墩干扰,从Badcase的例子中分析发现,准确率低的部分原因在于错误识别了桥墩等类圆形物体。因此通过DBSCAN进行聚类时,DBSCAN-VoxelNet网络模型有效利用距离信息,在后续处理过程中对神经网络错误判断的桥墩进行抑制。进一步对多目标检测进行分析,三棱锥和圆柱、球体之间的区分效果较好,但是当距离较远时(15 m)圆柱和球体之间的误检率依然较高,仅仅通过形状层面的信息无法进行区分,因此加入DBSCAN聚类信息滤除波浪的影响,能够有效提高检测精度。为了更好地突出模型性能,在非极大值抑制时,当球体和圆柱的分类置信度之



间差值大于 40% 时,保留神经网络预测的结果,当球体和圆柱体分类置信度差值小于 40% 时,对预测到的物体进行抛弃,最终将 DBSCAN-VoxelNet 的 mAP 提高了 1.2 个百分点。

为更好地验证算法对水面波浪干扰条件下的目标检测任务的有效性,建立虚拟仿真场景下的水面波浪环境,测试算法的 mAP 值。在仿真测试的过程中,需要模拟激光雷达对水面波浪和目标的反射点云数据,使其尽量接近现实数据,采用 Unity3D 仿真平台完成虚拟环境搭建,利用 Terrain 提供的纹理绘制工具添加了树木、草地等元素的表面细节信息,达到逼真的水域模拟效果。最终建立的水面仿真环境如图 13 所示。

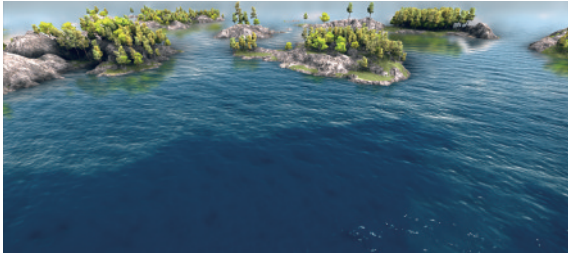


图 13 水面目标检测仿真环境构建

Fig. 13 Construction of surface target detection simulation environment

为突出波浪对水面目标检测的影响,调整虚拟场景中波浪的起伏大小,在 Ocean 组件中,可以通过 Wave 参数调整波浪分支的姿态,包括波浪角度 (Angle)、速度 (Speed)、等级 (Scale)、长度 (Length) 和锐度 (Sharpness),波浪起伏状态如图 14 所示。

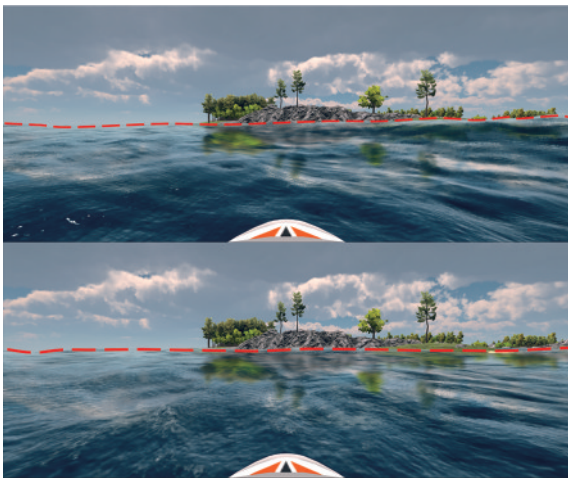


图 14 船舶第一视角虚拟波浪起伏状态

Fig. 14 Virtual wave fluctuation state in the first view of ship

要想在搭建完成的虚拟环境中完成水面目标检测任务,还需要向环境中添加目标,生成障碍物反射回波,本文的虚拟实验环境中,共添加了 5 种不同的虚拟目标,第 1 种是直径为 1 m 的球形,第 2 种是边长为 1 m 的正方体,第 3 种是长轴为 3 m、短轴为 1 m 的椭球形,

第 4 种是长为 2 m、宽和高均为 1 m 的长方体,第 5 种是长为 5 m、宽为 1 m 的船舶模型,如图 15 所示。

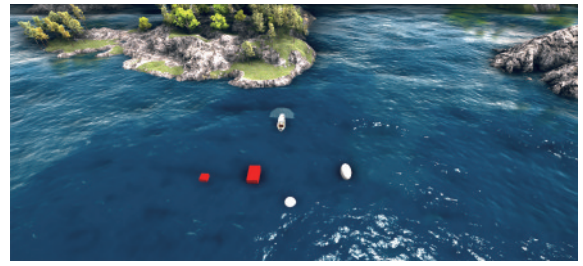


图 15 水面虚拟环境目标设置

Fig. 15 Target setting of water virtual environment

利用 Unity3D 中的虚拟激光雷达,依次按照 5, 10, 15 m 的距离采集目标回波并用于算法模型训练,得到的检测结果如表 4 所示。

表 4 DBSCAN-VoxelNet 算法在不同环境中的 mAP  
Table 4 mAP of DBSCAN-VoxelNet at different environments

Environment	mAP	Average detection speed / (frame · s <sup>-1</sup> )
Without wave	0.841	0.08
With wave	0.897	0.08

对仿真水面波浪环境中的目标点云数据进行测试,发现在有波浪的情况下,算法的 mAP 相比于实际平静水面提高了 5.6 个百分点,说明针对水面波浪杂波, DBSCAN-VoxelNet 能够有效干扰噪声,提高对目标的检测精度,满足船舶对环境感知的适应性要求。

## 4 结 论

环境感知作为 USV 自主航行的先决条件,为航行任务调度的安全执行发挥了举足轻重的作用。作为环境感知的重要组件,3D 激光雷达能够以高测量精度、较低漏检率完成目标检测任务。虽然在作用场景上,3D 雷达在水面检测比在地面检测能更出色完成需求,但依然受到杂波的影响。因此利用环境信息对水面噪声进行抑制,提出一种基于激光雷达的 DBSCAN-VoxelNet 检测算法,降低波浪等干扰信息对检测精度的影响,同时将水面稀疏点云划分为体素形式,再利用神经网络结构提取目标特征信息,实现多模态融合,在保证检测速度的条件下提高了检测精度,最终 mAP 为 82.4%。所提 DBSCAN-VoxelNet 检测算法可以满足 USV 航行要求,为路径规划和避障等任务提供有效环境信息。

## 参 考 文 献

- [1] Liu Z X, Zhang Y M, Yu X, et al. Unmanned surface vehicles: an overview of developments and challenges[J]. Annual Reviews in Control, 2016, 41: 71-93.
- [2] 蔡喜平, 赵远, 黄建明, 等. 成像激光雷达系统性能的研究[J]. 光学技术, 2001, 27(1): 60-62.

- Cai X P, Zhao Y, Huang J M, et al. Research on the performance of imaging laser radar[J]. *Optical Technology*, 2001, 27(1): 60-62.
- [3] 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, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 77-85.
- [4] Charles R Q, Yi L, Su H, et al. Pointnet++: deep hierarchical feature learning on point sets in a metric space [EB/OL]. (2017-06-07)[2021-05-04]. <https://arxiv.org/abs/1706.02413>.
- [5] González A, Villalonga G, Xu J L, et al. Multiview random forest of local experts combining RGB and LIDAR data for pedestrian detection[C]//2015 IEEE Intelligent Vehicles Symposium, June 28 - July 1, 2015, Seoul, Korea (South). New York: IEEE Press, 2015: 356-361.
- [6] Li B, Zhang T L, Xia T. Vehicle detection from 3D lidar using fully convolutional network[C]//Robotics: Science and Systems XII, Robotics: Science and Systems Foundation, June 18-22, 2016, Ann Arbor, Michigan, USA. [S.l.: s.n.], 2016.
- [7] 李小毛, 张鑫, 王文涛, 等. 基于 3D 激光雷达的无人水面艇海上目标检测[J]. *上海大学学报(自然科学版)*, 2017, 23(1): 27-36.  
Li X M, Zhang X, Wang W T, et al. 3D Lidar-based marine object detection for USV[J]. *Journal of Shanghai University (Natural Science Edition)*, 2017, 23(1): 27-36.
- [8] Chen X Z, Ma H M, Wan J, et al. Multi-view 3D object detection network for autonomous driving[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 6526-6534.
- [9] 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), October 1-5, 2018, Madrid, Spain. New York: IEEE Press, 2018.
- [10] Chen X Z, Ma H M, Wan J, et al. Multi-view 3D object detection network for autonomous driving[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 6526-6534.
- [11] Maturana D, Scherer S. VoxNet: a 3D Convolutional Neural Network for real-time object recognition[C]//2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), September 28-October 2, 2015, Hamburg, Germany. New York: IEEE Press, 2015: 922-928.
- [12] Zhou Y, Tuzel O. VoxelNet: end-to-end learning for point cloud based 3D object detection[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 4490-4499.
- [13] Yan Y, Mao Y X, Li B. SECOND: sparsely embedded convolutional detection[J]. *Sensors*, 2018, 18(10): 3337.
- [14] Graham B, Engelcke M, van der Maaten L. 3D semantic segmentation with submanifold sparse convolutional networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018.
- [15] Luo W J, Yang B, Urtasun R. Fast and furious: real time end-to-end 3D detection, tracking and motion forecasting with a single convolutional net[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 3569-3577.
- [16] Shi S S, Wang X G, Li H S. PointRCNN: 3D object proposal generation and detection from point cloud[EB/OL]. (2018-12-11)[2021-02-03]. <https://arxiv.org/abs/1812.04244v2>.
- [17] Lang A H, Vora S, Caesar H, et al. PointPillars: fast encoders for object detection from point clouds[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: 12689-12697.
- [18] Simon M, Milz S, Amende K, et al. Complex-YOLO: an Euler-region-proposal for real-time 3D object detection on point clouds[M]//Leal-Taixé L, Roth S. Computer vision-ECCV 2018 workshops. Lecture notes in computer science. Cham: Springer, 2019, 11129: 197-209.
- [19] Kwon S S, Park T H. Channel-based network for fast object detection of 3D LiDAR[J]. *Electronics*, 2020, 9(7): 1122.
- [20] Yi H W, Shi S S, Ding M Y, et al. SegVoxelNet: exploring semantic context and depth-aware features for 3D vehicle detection from point cloud[EB/OL]. (2020-02-13)[2021-05-06]. <https://arxiv.org/abs/2002.05316>.
- [21] 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 and Machine Intelligence*, 2017, 39(6): 1137-1149.