文章编号：1003－1251(2024)05－0049－08

基于教师－学生模型的点云目标检测算法

文 峰，石明泽，刘思萌，殷向阳

（沈阳理工大学 信息科学与工程学院，沈阳 110159）

摘 要：针对现有点云目标检测算法对室外自动驾驶场景中遮挡物体检测精度低的问题，提出一种基于知识蒸馏和注意力增强的点云目标检测算法。以CIA-SSD模型为基础，设计了一种基于教师－学生模型的密集特征生成模块，提出基于交并比匹配策略的密集数据生成方法，将稀疏特征转换为密集特征。密集特征生成模块位于学生模型中，学生模型在教师模型生成的软目标监督训练下，推断出完整的密集目标特征，实现目标特征的补全；在教师模型中设计空间注意力和通道注意力机制，增强密集目标点云，提升特征图的质量。在KITTI数据集上的验证实验结果表明：与SE-SSD模型和CIA-SSD模型相比，本文提出的算法保持了单阶段目标检测速度的优势，同时明显提升了检测精度。

关 键 词：点云目标检测；CIA-SSD；教师－学生模型；密集特征模块；注意力机制；交并比匹配策略

中图分类号：TP391.4 **文献标志码**：A **DOI**：10.3969/j.issn.1003－1251.2024.05.008

Point Cloud Object Detection Algorithm Based on Teacher-student Model

WEN Feng，SHI Mingze，LIU Simeng，YIN Xiangyang

（Shenyang Ligong University，Shenyang 110159，China）

Abstract：A point cloud object detection algorithm based on knowledge distillation and attention enhancement is proposed to address the issue of low detection accuracy for occluded objects in outdoor autonomous driving scenarios. Building upon the CIA-SSD algorithm model，a dense feature generation module based on a teacher-student model is designed. A dense data generation method using intersection over union matching strategy is proposed to convert sparse features into dense features. The dense feature generation module is located within the student model，which infers complete dense object features under the soft target supervision generated by the teacher model，achieving the completion of object features. Spatial attention and channel attention mechanisms are designed in the teacher model to enhance the dense point cloud of the objects and improve the quality of the feature maps. Validation results on the KITTI dataset demonstrate that the proposed method significantly improves detection accuracy while maintaining the speed advantage of single-stage object detection compared to the SE-SSD model and CIA-SSD model.

Key words：point cloud object detection；CIA-SSD；teacher-student model；dense feature generation module；attention mechanism；intersection over union matching strategy

收稿日期：2023－10－27

基金项目：国家重点研发计划“社会治理与智慧社会科技支撑”重点专项(2022YFC3302502)

作者简介：文峰(1977—)，男，教授，博士，研究方向为大数据分析技术、机器学习、进化计算、智能交通等。

点云目标检测是利用激光雷达等传感器进行障碍物、行人、车辆等环境感知的一项技术,可实现更安全的自动化巡检[1],在智能机器人和自动驾驶车辆等领域具有重要的应用。点云目标检测算法是计算机视觉领域的重要研究内容,用于从三维点云数据中识别和定位不同的物体和目标,分为单阶段和两阶段算法,单阶段算法具有速度快、精度低的特点,而两阶段算法具有速度慢、精度高的特点[2-3]。

CIA-SSD 算法[4]是单阶段点云目标检测算法的突破性成果,其首先通过 VoxelNet[5]将点云数据体素化处理为鸟瞰视图(bird eyes view,BEV)特征,然后通过检测头部回归目标物体在点云坐标系中的位置。车爱博等[1]在 CIA-SSD 算法的基础上引入注意力机制,提出了 TFAF-SSD 算法,加强了 CIA-SSD 对前景点云的关注度,提升了提取特征图的质量。傅荟璇等[2]在 PV-RCNN 算法的基础上引入了自适应性可变形卷积,改变了特征提取方式,提高了对汽车、骑行者等动态目标的检测精度。陈辉等[6]以 PV-RCNN 作为基础模型,提出了基于点云补全的点云目标检测方法,根据大部分物体形状对称且重复率高的特点,通过自身镜像进行遮挡恢复,并通过预测区域所包含目标形状的占用概率对遮挡点云进行补全,提升了特征图的完整性和丰富性,提高了遮挡目标的检测精度。由于遮挡和某些材料反射的原因会引起激光信号丢失,激光点云实际是不完整的 2.5D 数据,从而导致点云目标检测算法精度较低。

Wang 等[7]提出了一种新的 Sparse2Dense(S2D)方法,通过设计知识蒸馏[8]和 S2D 模块并利用点云重建的思想,为解决点云数据因遮挡导致难以检测的问题提供了新思路。该方法采用时序序列点云帧融合、镜像复制生成密集数据,不仅生成速度慢,而且由于自动驾驶现实场景中动态目标较多,点云的时序融合会导致生成结果中只有静态目标生成了完整密集化点云,动态目标因为坐标改变无法融合生成的数据依然含有点云缺失。当采用缺失点云特征作为蒸馏信号联合训练学生模型时,存在 S2D 模块特征重建不完整的问题。

为解决目标检测中因遮挡导致精度低的问题,本文在单阶段 CIA-SSD 网络基础上提出 S2D-SSD 网络结构,并设计基于交并比(IoU)匹配相似目标点云的密集数据生成方法。引入教师-学生模型的知识蒸馏技术,在学生模型的骨干网络之后以残差的方式加入基于点云重建网络的 S2D 模块,在教师模型中引入卷积块注意力模块(convolutional block attention module,CBAM)[9]。通过使用密集点云数据预训练教师模型,使教师模型学习使用密集点云的密集特征信息进行目标检测;学生模型的输入为稀疏点云,通过蒸馏教师模型输出的密集特征训练学生模型,使之可以在稀疏数据的输入下,通过 S2D 模块三维重建稀疏特征为密集特征,利用密集特征完成后续的目标检测。学生模型以教师模型的密集特征作为监督信号,结合 S2D 模块,能够在特征层对因遮挡造成的稀疏点云进行密集化。通过在特征层中重新生成点云特征,有效解决点云目标检测算法在处理稀疏点云数据时精度较低的问题。

1 CIA-SSD 算法

CIA-SSD 是一种基于深度学习的单阶段点云目标检测模型,该模型提出了空间语义特征聚合(spatial semantic feature aggregation,SSFA)模块,可以提取丰富的空间特征和语义特征,提升了检测器的精度。CIA-SSD 骨干网络结构如图 1 所示。

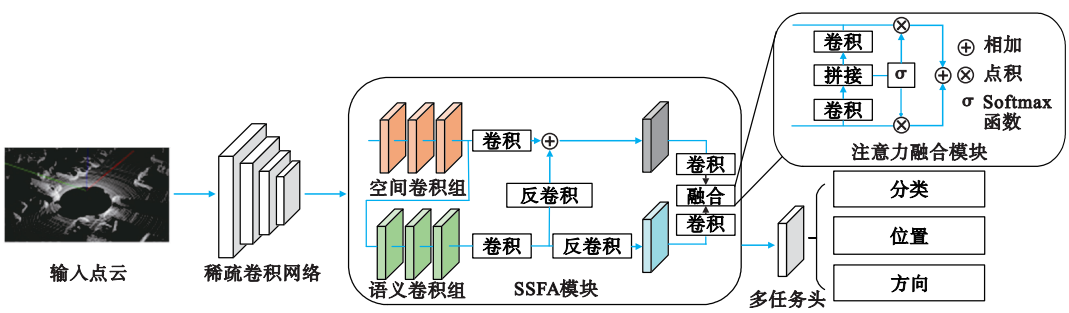


图 1 CIA-SSD 骨干网络结构

Fig. 1 CIA-SSD backbone network architecture

首先,特征提取主干网络使用稀疏卷积网络层提取三维特征,将三维点云数据投影至 BEV,以便进行二维卷积操作。然后,在 SSFA 模块中先通过空间卷积组得到空间特征,再经过语义卷积组得到高级语义特征,高级语义特征经过反卷积恢复特征尺寸并与低级空间特征简单融合。为更好地利用每一级特征,SSFA 模块采用了一种轻量级注意力融合机制,能够自适应地融合高级语义特征和低级空间特征。最后,SSFA 输出的融合特征输入多任务检测头,输出预测结果。

2　改进 CIA-SSD 的 S2D-SSD 算法

2.1　总体结构

由于激光雷达的自遮挡和外部遮挡,CIA-SSD 方法对复杂道路场景中的小目标、存在遮挡的目标检测和定位精度较差[10]。为此,本文将点云目标检测和点云重建算法结合在同一个卷积神经网络模型中,通过引入 S2D 模块,对发生遮挡的稀疏点云进行点云特征重建,提高单阶段检测模型 CIA-SSD 的点云补全能力。改进后的 S2D-SSD 网络结构如图 2 所示,图中变量说明详见下文。

S2D-SSD 模型包括教师模型 D-SSD 和学生模型 S-SSD 两个部分。在 S-SSD 模型中,引入了 S2D 模块并以残差的方式与 CIA-SSD 结合,实现稀疏特征密集化;在 D-SSD 中引入了 CBAM 注意力模块。除 CBAM 和 S2D 模块外,教师和学生模型的其他部分采用了相同的 CIA-SSD 网络结构。区域建议网络(region proposal network,RPN)模块负责提取感兴趣区域,头部则预测目标的位置、分类和方向。

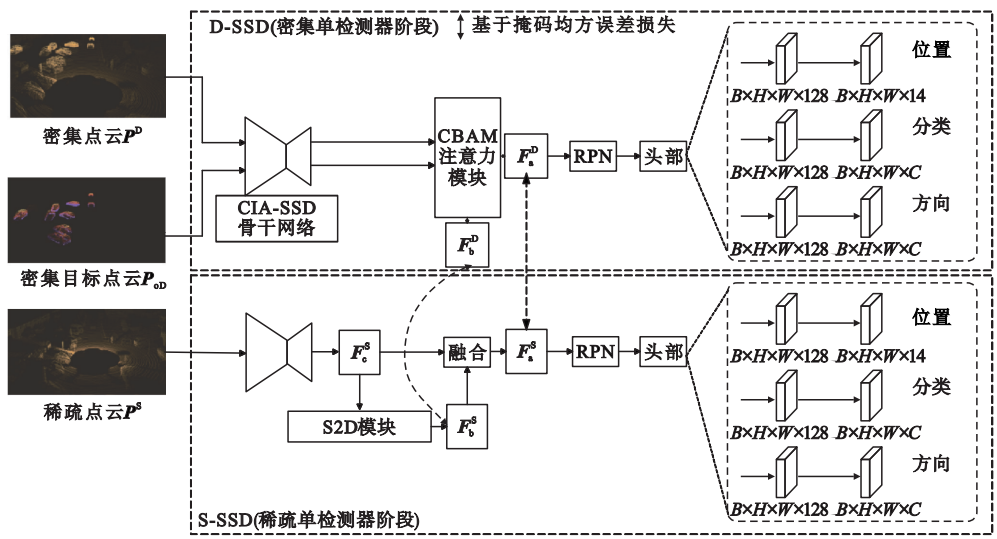


图 2　S2D-SSD 网络结构图

Fig. 2　S2D-SSD network architecture

以 $\boldsymbol{P}^{\mathrm{D}}$ 表示包含密集目标和背景的密集点云,由数据预处理得到;$\boldsymbol{P}_{\mathrm{oD}}$ 表示只有目标物体的密集点云,由 $\boldsymbol{P}^{\mathrm{D}}$ 和点云标签提取目标得到;$\boldsymbol{P}^{\mathrm{S}}$ 表示未经处理的稀疏点云,为数据集原始单帧点云数据;B 为训练参数批大小;H、W 表示特征图的长、宽;C 为特征通道数。S2D-SSD 网络模型的训练分为两个阶段,在第一训练阶段,由密集点云及标签训练 D-SSD 网络,训练完成后进入第二阶段,进行教师 – 学生模型的训练。相关公式如下。

$$\boldsymbol{F}_{\mathrm{a}}^{\mathrm{D}}=f_{\mathrm{Attention}}(f_{\mathrm{D\text{-}SSD}}(\boldsymbol{P}^{\mathrm{D}})) \tag{1}$$

$$\boldsymbol{F}_{\mathrm{b}}^{\mathrm{D}}=f_{\mathrm{Attention}}(f_{\mathrm{D\text{-}SSD}}(\boldsymbol{P}_{\mathrm{oD}})) \tag{2}$$

$$\boldsymbol{F}_{\mathrm{c}}^{\mathrm{S}}=f_{\mathrm{S\text{-}SSD}}(\boldsymbol{P}^{\mathrm{S}}) \tag{3}$$

$$\boldsymbol{F}_{\mathrm{b}}^{\mathrm{S}}=f_{\mathrm{S2D}}(\boldsymbol{F}_{\mathrm{c}}^{\mathrm{S}}) \tag{4}$$

$$\boldsymbol{F}_{\mathrm{a}}^{\mathrm{S}}=\boldsymbol{F}_{\mathrm{c}}^{\mathrm{S}}\oplus\boldsymbol{F}_{\mathrm{b}}^{\mathrm{S}} \tag{5}$$

式中:$f_{\mathrm{Attention}}$ 表示注意力模块;$f_{\mathrm{S\text{-}SSD}}$ 表示 S-SSD 骨干网络部分计算;$f_{\mathrm{D\text{-}SSD}}$ 表示 D-SSD 骨干网络部分计算;f_{S2D} 表示 S2D 模块部分计算;$\boldsymbol{F}_{\mathrm{a}}^{\mathrm{D}}$ 和 $\boldsymbol{F}_{\mathrm{b}}^{\mathrm{D}}$ 是来自 D-SSD 的监督信号,其中 $\boldsymbol{F}_{\mathrm{a}}^{\mathrm{D}}$ 是带有背景信息的密集特征,由密集点云 $\boldsymbol{P}^{\mathrm{D}}$ 输入 D-SSD,并在 D-SSD 的骨干网络后输出,$\boldsymbol{F}_{\mathrm{b}}^{\mathrm{D}}$ 是在 D-SSD 只有前景目标下,由密集目标点云 $\boldsymbol{P}_{\mathrm{oD}}$ 经过骨干网络生成的密集特征;$\boldsymbol{F}_{\mathrm{c}}^{\mathrm{S}}$ 是 S-SSD 骨干网络输出的特征值;$\boldsymbol{F}_{\mathrm{b}}^{\mathrm{S}}$ 为 S2D 模块三维重建后的特征值;$\boldsymbol{F}_{\mathrm{a}}^{\mathrm{S}}$ 是 $\boldsymbol{F}_{\mathrm{c}}^{\mathrm{S}}$ 和 $\boldsymbol{F}_{\mathrm{b}}^{\mathrm{S}}$ 特征值的融合结果。

在教师 – 学生模型联合训练阶段,D-SSD 网络并行输入两个点云数据 $\boldsymbol{P}^{\mathrm{D}}$ 和 $\boldsymbol{P}_{\mathrm{oD}}$,S-SSD 网络

使用原始点云 $\boldsymbol{P}^{\mathrm{S}}$ 为输入数据;在测试阶段,只需要 S-SSD 网络进行推理。为训练 S2D 模块,本文将 D-SSD 和 S-SSD 中对应的特征进行匹配,并在掩码均方差(masked mean squared error,MMSE)损失函数的训练下实现 $\boldsymbol{F}_{\mathrm{b}}^{\mathrm{S}}$ 和 $\boldsymbol{F}_{\mathrm{b}}^{\mathrm{D}}$、$\boldsymbol{F}_{\mathrm{a}}^{\mathrm{S}}$ 和 $\boldsymbol{F}_{\mathrm{a}}^{\mathrm{D}}$ 两对特征图中对应像素特征值之间的距离最小化。S-SSD 在 D-SSD 和 MMSE 损失函数的训练下,使 S2D 模块的输出完成目标特征的密集化重建,实现两个特征对 $\boldsymbol{F}_{\mathrm{b}}^{\mathrm{S}}$ 和 $\boldsymbol{F}_{\mathrm{b}}^{\mathrm{D}}$、$\boldsymbol{F}_{\mathrm{a}}^{\mathrm{S}}$ 和 $\boldsymbol{F}_{\mathrm{a}}^{\mathrm{D}}$ 对齐。

在模型推理结果时不再需要 D-SSD 的监督,只保留 S-SSD 网络模型,且此时在第二阶段训练完成后、输入稀疏点云时,S2D 模块会将稀疏点云在特征层密集化,从而达到被遮挡目标检测精度的提升。

2.2 S2D 密集特征生成模块

在自动驾驶场景中,车辆、行人等动态目标往往具有对称性,故引入 S2D 模块解决点云目标检测中因遮挡造成的严重误检、漏检问题。在 S2D-SSD 网络模型中通过挖掘特征图之间的关联性,提取 D-SSD 网络生成的密集特征 $\boldsymbol{F}_{\mathrm{b}}^{\mathrm{D}}$ 和 $\boldsymbol{F}_{\mathrm{a}}^{\mathrm{D}}$ 作为监督信号,作用于 S-SSD 中 S2D 模块的输出。S2D 模块的网络结构如图 3 所示,该模块以稀疏特征 $\boldsymbol{F}_{\mathrm{c}}^{\mathrm{S}}$ 作为输入,并输出生成的密集特征 $\boldsymbol{F}_{\mathrm{b}}^{\mathrm{S}}$。

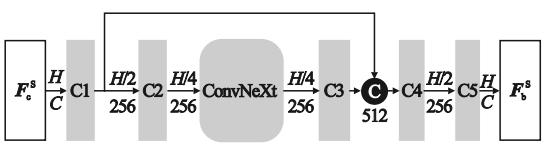


图 3 S2D 模块结构图

Fig. 3 S2D module architecture

首先稀疏点云 $\boldsymbol{P}^{\mathrm{S}}$ 经过骨干网络生成稀疏特征 $\boldsymbol{F}_{\mathrm{c}}^{\mathrm{S}}$,网络可以在特征图空间中采取有效的二维卷积操作。然后通过使用步长为 2 的 C1、C2 卷积层对特征图进行下采样,将输入稀疏特征的尺寸调整到与之相匹配的大小,设置通道数为 256。

ConvNeXt 聚合特征信息的能力比 ResNet、transformer 更强[11-13],因此本文在 S2D 模块中嵌入了三个 ConvNeXt 残差块,形成一个 ConvNeXt 卷积组,ConvNeXt 块的输入输出特征图尺寸相同。ConvNeXt 卷积块如图 4 所示,每个卷积块含三层卷积结构,第一层是 7×7 的深度卷积和一个归一化层,第二层是一个 1×1 的卷积和一个 GELU 激活函数[11],第三层是一个 1×1 的卷积,使特征通道降维至 256,并与原特征图以残差的方式等尺寸融合相加。

通过转置二维卷积组 C3 对特征进行上采样

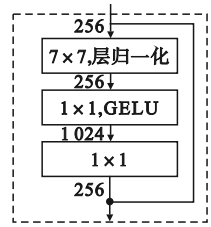


图 4 ConvNeXt 卷积块

Fig. 4 ConvNeXt block

恢复特征图尺寸,并将结果与 C1 卷积后的特征进行拼接;将拼接后的特征送到 3×3 的 C4 卷积层,并使用 C4、C5 卷积对特征进行上采样操作,获得最终的加密特征 $\boldsymbol{F}_{\mathrm{b}}^{\mathrm{S}}$。除 ConvNeXt 块中的卷积层之外,每个卷积操作后面都有一个批量归一化和一个 GELU 非线性操作。

本文采用基于教师模型 D-SSD 的密集特征对学生模型 S-SSD 中的 S2D 模块进行指导型训练,在训练 S-SSD 时,D-SSD 模型参数在反向传播时不做更新,D-SSD 输出的密集特征用于监督 S2D 模块的输出。

2.3 CBAM 注意力机制

当点云被骨干网络处理为 BEV 特征时,直接提取的 BEV 特征表达能力较弱。因此,本文在教师模型骨干网络之后引入 CBAM 注意力机制,该模块结构如图 5 所示。

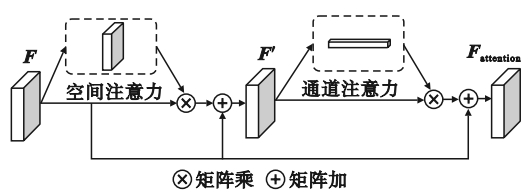


图 5 CBAM 注意力机制

Fig. 5 CBAM attention mechanism

CBAM 模块由两个注意力子模块组成:通道注意力模块和空间注意力模块。通道注意力模块捕捉特征图中不同通道之间的相互关系,使网络能够专注于更具信息量的通道;空间注意力模块捕捉特征图中每个通道内的空间关系,使网络能够关注相关的空间区域。将骨干网络输出的特征图 $\boldsymbol{F}$ 作为该模块的输入特征图。首先通过计算尺寸为 $C\times H\times W$ 的空间注意力权重得到每个像素点的权重,将权重与原来的特征图相乘,并进行短路连接;然后将上一步的计算结果 $\boldsymbol{F}'$ 经过通道注意力增强,通过计算通道注意力权重,将尺寸为 $C\times 1\times 1$ 的通道注意力权重与 $\boldsymbol{F}'$ 相乘,再与原特征图进行短路连接,该步骤的目的是计算特征图

每个通道的权重并进行分配;最后输出增强后的特征图 $\boldsymbol{F}_{\text{attention}}$。相关公式表达如下。

$$\boldsymbol{F}' = f_{\text{spatial}}(\boldsymbol{F}) \cdot \boldsymbol{F} + \boldsymbol{F} \tag{6}$$

$$\boldsymbol{F}_{\text{attention}} = f_{\text{channel}}(\boldsymbol{F}') \cdot \boldsymbol{F}' + \boldsymbol{F} \tag{7}$$

式中 f_{spatial} 和 f_{channel} 分别代表空间注意力和通道注意力操作。空间注意力操作关注特征图的空间布局,而通道注意力操作则关注特征图的通道重要性,通过这两种操作捕捉关键特征,提高检测质量,抑制点云噪声和背景特征,增强目标物体 BEV 特征表达能力,进而提高点云目标检测的准确度。

2.4　密集数据生成方法

本文在 KITTI 公开数据集[14]上通过 3D 标注框从多帧数据中提取稀疏目标车辆点云,然后采用基于 IoU 匹配的相似目标进行融合。密集目标点云生成过程如图 6 所示。选定数据序列 01,根据 01 数据序列的三维标注框,在时序帧点云数据前后帧中基于 IoU 对标注框进行匹配,设定 IoU 匹配阈值为 0.9,若标注框和数据序列 01 的标注框 IoU 大于 0.9,将其选为待融合点云数据。按照数据序列 01 的匹配方法计算,直到待融合点云数据可融合为完整目标点云数据。点云融合后做镜像复制得出密集目标。

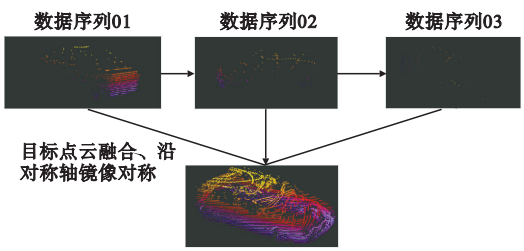
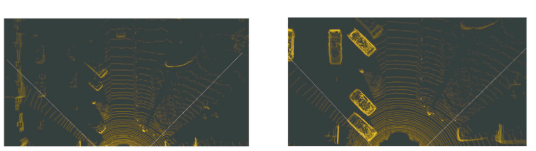


图 6　密集目标点云生成示意图

Fig. 6　Dense object point cloud generation illustration

采用 Sparse2Dense 方法进行目标点云检测时存在时序融合点云帧时动态目标无法完整融合的问题,故本文提出基于 IoU 匹配策略的密集数据生成方法,仅选取大小相匹配的目标点云进行融合,得到密集目标点云。密集点云生成前后可视化图如图 7 所示。



(a) 原KITTI数据可视化图　　(b) 点云目标密集化后可视化图

图 7　密集点云生成前后可视化图

Fig. 7　Visualization of dense point cloud generation before and after

2.5　检测头部及损失函数

教师模型和学生模型的检测分支均使用 CIA-SSD 的检测头结构,如图 8 所示。

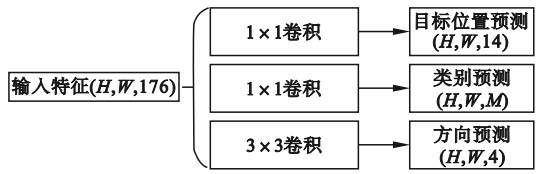


图 8　CIA-SSD 检测头部结构

Fig. 8　CIA-SSD head architecture

检测分支输入特征图尺寸为 $H \times W \times 176$,其中 176 为特征图通道数。得到四组平行输出结果:目标位置结果即预测框的坐标信息,类别结果即类别信息,方向结果即方向信息,交并比预测结果即预测框的 IoU。图中类别预测中的 M 为检测类别数。

教师模型 D-SSD 检测分支的总损失函数 L 定义为

$$L = L_{\text{cls}} + \omega L_{\text{box}} + \mu L_{\text{dir}} + \lambda L_{\text{iou}} \tag{8}$$

式中:L_{cls} 为分类损失;L_{box} 为边界框回归损失;L_{dir} 为方向分类损失;L_{iou} 为 IoU 预测损失,本文只计算正样本的 IoU 预测损失;ω、μ、λ 为平衡损失函数的权重参数,$\omega = 2.0$,$\mu = 0.2$,$\lambda = 1.0$。

学生模型 S-SSD 检测分支的总损失函数定义为

$$L_{\text{S}} = L_{\text{cls}} + \omega L_{\text{box}} + \mu L_{\text{dir}} + \lambda L_{\text{iou}} + L_{\text{S2D}} \tag{9}$$

式中 L_{S2D} 是通过知识蒸馏优化密集特征生成模块的掩码均方差损失函数。L_{S2D} 计算式为

$$\begin{aligned} L_{\text{S2D}} = {} & \beta \frac{1}{|N|} \sum_{i=1}^{N} (\boldsymbol{F}_{\text{a},i}^{\text{S}} - \boldsymbol{F}_{\text{a},i}^{\text{D}})^2 + \\ & \gamma \frac{1}{|\widetilde{N}|} \sum_{i=1}^{\widetilde{N}} (\boldsymbol{F}_{\text{a},i}^{\text{S}} - \boldsymbol{F}_{\text{a},i}^{\text{D}})^2 + \\ & \beta \frac{1}{|N'|} \sum_{i=1}^{N'} (\boldsymbol{F}_{\text{b},i}^{\text{S}} - \boldsymbol{F}_{\text{b},i}^{\text{D}})^2 + \\ & \gamma \frac{1}{|\widetilde{N}'|} \sum_{i=1}^{\widetilde{N}'} (\boldsymbol{F}_{\text{b},i}^{\text{S}} - \boldsymbol{F}_{\text{b},i}^{\text{D}})^2 \end{aligned} \tag{10}$$

式中:N 和 $\widetilde{N}$ 分别是 $\boldsymbol{F}_{\text{a}}^{\text{D}}$ 的非零值和零值特征元素数量;N' 和 $\widetilde{N}'$ 是 $\boldsymbol{F}_{\text{b}}^{\text{D}}$ 的非零值和零值特征元素数量;下角标 i 表示第 i 个特征元素;β、γ 为平衡非空特征和空特征的损失权重,$\beta = 10$,$\gamma = 20$。

3　实验及分析

3.1　实验数据和评价指标

为验证本文密集特征生成模块和注意力机制

的有效性,在 KITTI 数据集上评估本文提出的 S2D-SSD 算法模型。KITTI 数据集根据目标被遮挡情况、与当前视点距离考虑了三个检测难度级别:简单、中等和复杂。在简单级别中,目标具有明显特征,且点云目标遮挡比例小于等于 15%;在中等级别中,目标的遮挡状态处于中间水平,点云目标遮挡比例在 15% ~50% 之间;在复杂级别中,目标通常被严重遮挡,导致仅含有少量点云,遮挡比例大于等于 50%。因为车辆类别数量较多且为样本中最大的类别,故本文只在车辆类别上做训练和测试。

按照 KITTI 官方评价指标,以平均精度(AP)评价三维空间和 BEV 场景下的检测结果,作为检测性能的评估指标,分别采用 AP_{3D} 和 AP_{BEV} 表示。AP_{3D} 考虑到了三维空间中的所有信息,可以更准确地评估模型的性能,IoU 阈值为 0.7,当 IoU 大于 0.7 时,则视为检测准确;使用 AP_{BEV} 进行评估时,计算 IoU 时只考虑预测的边界框和真实边界框在地面平面上的投影。

3.2 实验结果分析

3.2.1 消融实验

以 CIA-SSD 模型为基础,引入 S2D 模块和 CBAM 模块进行消融实验,结果如表 1 所示。由表 1 可见:当在 CIA-SSD 基础上增加 CBAM 模块时,简单、中等、复杂三种检测难度下的 AP_{3D} 和 AP_{BEV} 均有提升,说明 CBAM 模块可明显提升特征的表达能力;当采用 CIA-SSD 构成教师 - 学生模型,在学生模型中增加 S2D 模块后(表 1 中第三行数据),AP_{3D} 和 AP_{BEV} 明显提升,且在三种遮挡难度上,复杂遮挡的检测精度提升最大,其中 AP_{3D} 比基础模型 CIA-SSD 提高了 7.04%,说明 S2D 模块可有效解决点云遮挡导致检测精度低的问题。当 S2D 模块和 CBAM 模块同时与 CIA-SSD 结合,所有遮挡程度的检测精度 AP_{3D} 均获得大幅度提升,其中复杂遮挡的检测精度较基础模型提高了 7.99%。

表 1 消融实验结果

Table 1 Results of ablation experiment

CIA-SSD 基础模型	S2D 模块	CBAM 模块	AP_{3D}/%			AP_{BEV}/%		
			简单	中等	复杂	简单	中等	复杂
√			89.59	80.28	72.87	93.74	89.84	82.39
√		√	90.99(+1.40)	81.23(+0.95)	73.44(+0.57)	95.18(+1.44)	90.92(+1.08)	84.12(+1.73)
√	√		91.77(+2.18)	83.02(+2.74)	78.00(+5.13)	96.44(+2.70)	92.60(+2.76)	86.81(+4.42)
√	√	√	**92.04(+2.45)**	**83.25(+2.97)**	**78.69(+5.82)**	**98.34(+4.60)**	**93.28(+3.44)**	**87.89(+5.20)**

在教师模型中选择不同的蒸馏位置进行实验,结果如表 2 所示。由表 2 可见:蒸馏位置选择为 CBAM 模块之前时,AP_{3D} 和 AP_{BEV} 与表 1 中第三行实验结果相当,即与未加入 CBAM 模块时精度持平,此时 CBAM 的加入没有效果;蒸馏位置选择为 CBAM 模块之后时,三种检测难度下的 AP_{3D} 和 AP_{BEV} 均得到提升。在 CBAM 之后蒸馏,可以发挥 CBAM 的特征增强能力,密集点云生成的密集目标特征被进一步增强。

表 2 知识蒸馏实验

Table 2 Knowledge distillation experiment

教师模型蒸馏位置	AP_{3D}/%			AP_{BEV}/%		
	简单	中等	复杂	简单	中等	复杂
CBAM 之前	91.71	83.00	78.00	96.41	92.60	86.81
CBAM 之后	92.04	83.25	78.69	98.34	93.28	87.89

3.2.2 实验结果及对比分析

为进一步验证本文算法在点云目标检测中的有效性,将本文算法与 KITTI 目标检测数据集中排名第一的 SOTA 方法中 GLENet-C 网络[15]、基于教师 - 学生模型且以 CIA-SSD 作为基础模型改进的 SE-SSD 网络[16]以及 CIA-SSD 网络进行比较,在 KITTI 数据集中类别为车的目标检测效果对比如表 3 所示。

由表 3 可见:CIA-SSD 受稀疏点云数据的限制,其检测精度较低,均值平均精度(mAP_{3D})值为

80.91%,复杂遮挡的目标检测 AP_{3D} 值为 72.87%;与 CIA-SSD 相比,本文算法在密集特征生成模块和教师学生模型的作用下,提高了补全稀疏特征为密集特征的能力,在特征图中重建出目标的缺失部分,解决了激光雷达的自遮挡问题;相比 SOTA 方法中 GLENet-C 在简单遮挡时的数据表现,本文算法的 AP_{3D} 值提高了 2.23%,与 SOTA 方法在中等、复杂遮挡难度下的数据相比,AP_{3D} 分别相差 1.29%、0.13%,GLENet-C 方法对中等、复杂遮挡目标的检测精度提升更高。

为解决由于人为标签手动注释错误和遮挡引起数据集中 3D 包围框的不确定性,导致模型学习错误标签使得检测精度下降的问题,GLENet-C 引入了概率 3D 检测网络和标签不确定性估计网络,标签不确定性估计网络将问题建模为对非确定性潜在空间的包围框的预测,使用 KL 散度优化 3D 检测器和标签不确定性估计网络分布之间的差异,避免了模型对人为错误标注数据的学习。当目标发生严重遮挡时,一个点云目标可能对应多个可能的 3D 包围框;当目标发生简单遮挡时,人工标签大概率是明确的,且不具有模糊性。因此,GLENet-C 方法解决简单遮挡问题的效果不好。由于本文从点云重建的思路入手解决遮挡问题,未考虑严重遮挡的点云数据存在手动标签错误,导致本文算法针对中等、复杂遮挡目标的检测精度低于 GLENet-C 算法。

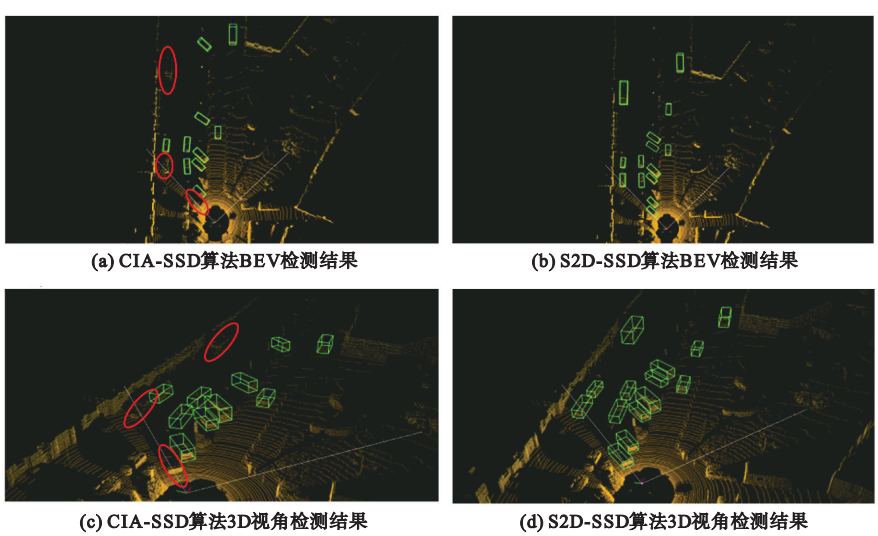
本文 S2D-SSD 模型的 mAP_{3D} 值提升到 84.66%,算法推理时间比 SE-SSD 增加了 0.23 ms,本文算法引入的模块比较简单,保持了单阶段点云目标检测算法的推理时间,可见改进后的 S2D-SSD 模型综合性能较好。

表 3　实验结果对比

Table 3　Comparison of experimental results

方法	AP_{3D}/%				AP_{BEV}/%				时间/ms
	简单	中等	复杂	mAP_{3D}	简单	中等	复杂	mAP_{BEV}	
CIA-SSD	89.59	80.28	72.87	80.91	93.74	89.84	82.39	88.66	30.76
SE-SSD	91.49	82.54	77.15	83.72	95.68	91.84	86.72	91.41	30.56
GLENet-C	89.82	**84.59**	**78.78**	84.39	–	–	–	–	31.01
本文网络	**92.04**	83.25	78.69	**84.66**	98.34	93.28	87.89	93.17	30.79

对改进前后网络推理结果的 BEV 和三维视图进行可视化,结果如图 9 所示。可以发现,CIA-SSD 算法产生了漏检,漏检原因为目标发生严重遮挡,本文 S2D-SSD 算法的推理结果中严重遮挡目标均被检测成功,且获得了精准的定位(图中红色椭圆标出)。对比实验结果表明,本文提出算法提升了检测精度,验证了其有效性。



(a) CIA-SSD算法BEV检测结果　(b) S2D-SSD算法BEV检测结果

(c) CIA-SSD算法3D视角检测结果　(d) S2D-SSD算法3D视角检测结果

图 9　密集数据生成前后检测结果

Fig. 9　Detection results before and after dense data generation

4 结论

针对目前点云目标检测方法存在的遮挡严重目标检测精度低的问题,提出一种基于教师 - 学生模型的点云目标检测算法,通过引入知识蒸馏、注意力增强和生成密集数据,提升三维目标检测性能。知识蒸馏使密集点云三维检测器(D-SSD)中获得的密集点知识转移到稀疏点云三维检测器(S-SSD),使得 S-SSD 能够在潜在空间中学习到密集化的三维特征;注意力模块通过提取更加丰富的密集特征,增强密集点云在特征图中的表达能力;密集点云生成方法可以生成完整密集点云,为训练提供完整的车辆数据。实验结果表明,本文算法在 CIA-SSD 的基础上提升了点云目标检测的精度。

本文未考虑人为标注标签的不准确性所导致的学习误差,故本文算法在解决遮挡问题上与 GLENet-C 算法相比,存在一定差距。未来研究可以着重于解决标签不准确性问题、KL 散度的学习方式、相机图片和点云信息的多模态融合等方向。

参考文献(References):

[1] 车爱博,张辉,李晨,等. 基于点云数据的交通环境下单阶段三维目标检测方法[J]. 计算机科学,2022,49(S2):567 - 572.
CHE A B,ZHANG H,LI C,et al. Single-stage 3D object detector in traffic environment based on point cloud data[J]. Computer Science,2022,49(S2):567 - 572. (in Chinese)

[2] 傅荟璇,刘凌风,王宇超. 基于改进 PV-RCNN 的 3D 目标检测算法实验研究[J]. 实验技术与管理,2022,39(2):23 - 28.
FU H X,LIU L F,WANG Y C. Experimental research on 3D target detection algorithm based on improved PV-RCNN[J]. Experimental Technology and Management,2022,39(2):23 - 28. (in Chinese)

[3] SHI S S,JIANG L,DENG J J,et al. PV-RCNN + +:point-voxel feature set abstraction with local vector representation for 3D object detection[J]. International Journal of Computer Vision,2023,131(2):531 - 551.

[4] ZHENG W,TANG W L,CHEN S J,et al. CIA-SSD:confident IoU-aware single-stage object detector from point cloud[J]. Proceedings of the AAAI Conference on Artificial Intelligence,2021,35(4):3555 - 3562.

[5] 周治国,李怡瑶,曹江微,等. 基于 3D 激光雷达的水面目标检测算法研究[J]. 激光与光电子学进展,2022,59(18):278 - 287.
ZHOU Z G,LI Y Y,CAO J W,et al. Surface target detection algorithm based on 3D lidar[J]. Laser & Optoelectronics Progress,2022,59(18):278 - 287. (in Chinese)

[6] 陈辉,王帅杰,蔡晗. 基于点云补全的三维目标检测[J]. 电子技术应用,2023,49(8):1 - 6.
CHEN H,WANG S J,CAI H. 3D object detection based on point cloud completion[J]. Application of Electronic Technique,2023,49(8):1 - 6. (in Chinese)

[7] WANG T Y,HU X W,LIU Z Z,et al. Sparse2Dense:learning to densify 3D features for 3D object detection[EB/OL]. arXiv:2211.13067(2022 - 11 - 23)[2023 - 10 - 17]. https://arxiv.org/abs/2211.13067.

[8] 单铭琦,文峰,高文印. 基于模型压缩算法改进 YOLOv4-tiny 模型的车辆检测技术研究[J]. 沈阳理工大学学报,2023,42(2):36 - 42.
SHAN M Q,WEN F,GAO W Y. Research on vehicle detection technology based on model compression algorithm to improve YOLOv4-tiny model[J]. Journal of Shenyang Ligong University,2023,42(2):36 - 42. (in Chinese)

[9] 殷向阳,李树辉,陈禹铭,等. 面向注意力增强和特征选择的多目标跟踪算法[J]. 沈阳理工大学学报,2022,41(4):26 - 31.
YIN X Y,LI S H,CHEN Y M,et al. Multi-object tracking algorithm based on attention enhancementand feature selection[J]. Journal of Shenyang Ligong University,2022,41(4):26 - 31. (in Chinese)

[10] 李宇轩,陈壹华,温兴,等. 改进 Point-Voxel 特征提取的 3D 小目标检测[J]. 微电子学与计算机,2023,40(2):50 - 58.
LI Y X,CHEN Y H,WEN X,et al. 3D small object detection based on improved Point-Voxel feature extraction[J]. Microelectronics & Computer,2023,40(2):50 - 58. (in Chinese)

[11] 杨文哲,郝如江,郭梓良,等. 基于一维 ConvNeXt 网络的齿轮箱故障诊断[J]. 国防交通工程与技术,2023,21(4):28 - 31,61.
YANG W Z,HAO R J,GUO Z L,et al. Fault diagnosis of gearbox based on 1D ConvNeXt[J]. Traffic Engineering and Technology for National Defence,2023,21(4):28 - 31,61. (in Chinese)

[12] 朱超平,杨艺. 基于 YOLO2 和 ResNet 算法的监控视频中的人脸检测与识别[J]. 重庆理工大学学报(自然科学),2018,32(8):170 - 175.
ZHU C P,YANG Y. Face detection and recognition in monitoring video based on YOLO2 and ResNet algorithm[J]. Journal of Chongqing University of Technology(Natural Science),2018,32(8):170 - 175. (in Chinese)

[13] 阚亚亚,张孙杰,熊娟,等. 结合 transformer 多尺度实例交互的稀疏集目标检测[J]. 应用科学学报,2023,41(5):777 - 788.
KAN Y Y,ZHANG S J,XIONG J,et al. Sparse set object detection combined with transformer multi-scale instance interaction[J]. Journal of Applied Sciences,2023,41(5):777 - 788. (in Chinese)

[14] GEIGER A,LENZ P,STILLER C,et al. Vision meets robotics:the KITTI dataset[J]. The International Journal of Robotics Research,2013,32(11):1231 - 1237.

[15] ZHANG Y F,ZHANG Q J,ZHU Z Y,et al. GLENet:boosting 3D object detectors with generative label uncertainty estimation[J]. International Journal of Computer Vision,2023,131(12):3332 - 3352.

[16] ZHENG W,TANG W L,JIANG L,et al. SE-SSD:self-ensembling single-stage object detector from point cloud[C]// 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville,TN,USA:IEEE,2021:14489 - 14498.

(责任编辑:宋颖韬)