计算机科学
COMPUTER SCIENCE

融合动态加权图卷积的三维目标检测

李宗民, 戎光彩, 白云, 徐畅, 鲜世洋

引用本文

李宗民, 戎光彩, 白云, 徐畅, 鲜世洋. 融合动态加权图卷积的三维目标检测[J]. 计算机科学, 2025, 52(3): 104-111.

LI Zongmin, RONG Guangcai, BAI Yun, XU Chang , XIAN Shiyang. 3D Object Detection with Dynamic Weight Graph Convolution [J]. Computer Science, 2025, 52(3): 104-111.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.240700041

融合动态加权图卷积的三维目标检测

李宗民　戎光彩　白　云　徐　畅　鲜世洋
中国石油大学(华东)青岛软件学院、计算机科学与技术学院　山东 青岛 266580

摘　要　三维目标检测是自动驾驶中最关键的技术之一,基于激光雷达的三维目标检测通常在点云构建的场景中进行。目前的三维检测方法不能充分地利用点云的结构信息,这将导致目标物体的误检和漏检。为此,提出了基于动态加权图卷积的DEG R-CNN。首先,在RoI中对节点设置主邻点和次邻点,为目标物体构建点云的图结构,恢复物体的几何信息;然后,在图中利用Gaussian函数和一维卷积,高效地聚合点云的结构特征;最后,使用交叉注意力机制自适应地融合不同粒度的图像特征,为点云补充图像语义信息。在KITTI数据集上进行实验,验证了各个模块的有效性,三维目标检测的3D mAP达到88.80%,相比基线模型提高了1.22%。同时,对三维目标检测的结果进行了可视化,并对可视化结果进行了分析。

关键词:点云;三维目标检测;激光雷达;多模态融合;自动驾驶
中图分类号　TP391

3D Object Detection with Dynamic Weight Graph Convolution

LI Zongmin,RONG Guangcai,BAI Yun,XU Chang and XIAN Shiyang
Qingdao Institute of Software,College of Computer Science and Technology,China University of Petroleum(East China),Qingdao,Shandong 266580,China

Abstract　3D object detection is one of the most critical technologies in autonomous driving,and 3D object detection based on LiDAR is usually carried out in the scene of point cloud construction. The current methods cannot fully use the point cloud's structural information,leading to false and missed target detection. To solve this problem,we propose a DEG R-CNN based on dynamically weighted graph convolution. Firstly,the primary neighbour and subordinate neighbour are set for the node in RoI,and the graph structure of the point cloud is constructed. The geometric information of the object is restored. Then,Gaussian and 1D convolution are used in the graph to efficiently aggregate the point cloud's structural features. Finally,the cross-attention mechanism adaptively fuses image features of different granularities to supplement the image semantic information. Experiments are conducted on KITTI dataset,and the effectiveness of modules is verified. The 3D mAP of the method reaches 88.80%,which is 1.22% higher than that of the baseline model. At the same time,the results of 3D object detection are visualized and analyzed in detail to understand performance and accuracy of the method better.

Keywords　Point clouds,3D object detection,LiDAR,Multimodal fusion,Automatic driving

1　引言

三维目标检测能够感知自然世界的空间信息,并且可以对目标物体进行定位和分类,目前在自动驾驶和机器视觉等领域发挥着至关重要的作用。激光雷达是自动驾驶车辆上最常用的传感器之一,它通过向车辆周围发射和接收激光生成点云,为三维目标检测提供最原始的数据。点云场景中的物体是由若干个无序的点构成的,检测需要从物体的点云结构信息中提取特征。因此,为了能够准确地进行检测,目标物体的点云结构信息应该被充分利用,特别是在检测具有相似几何结构的物体时,如果不能正确区分它们之间的细微差异,就会导致误检和漏检。在真实的驾驶环境中,由于目标物体的点云是通过激光雷达扫描获得,物体原本的形状并不能被完整地呈现,这也会导致原本不同的物体因为关键特征的消失而变得难以辨别。因此,充分利用点云的结构信息对于三维目标检测具有重要意义。

近年来,许多三维目标检测方法[1-6]被提出,它们不同程度地利用了点云的结构信息。基于PointNet和PointNet++的方法[1-2]主要在整个场景中通过点云的结构信息对目标进行检测,其对场景中的点云进行采样并聚合特征来分类和定位待检测的目标物体。由于需要在整个场景中对点云进行操作,这种方法往往会忽略目标物体细节的点云结构信息,而且

到稿日期:2024-07-08　返修日期:2024-09-09
基金项目:国家重点研发计划(2019YFF0301800);国家自然科学基金(61379106);山东省自然科学基金(ZR2013FM036,ZR2015FM011)
This work was supported by the National Key Research and Development Program of China(2019YFF0301800),National Natural Science Foundation of China(61379106) and Natural Science Foundation of Shandong Province,China(ZR2013FM036,ZR2015FM011).
通信作者:李宗民(lizongmin@upc.edu.cn)

检测的过程十分耗时。基于网格的方法[3-4]在特征兴趣区域(Region of Interest，RoI)中均匀设置相同数量的格子点，然后提取每个格子点周围的点云结构信息，并利用 RoI 中所有格子点的特征识别目标物体。此类方法一般在体素中进行，所以检测的速度相对较快，但固定的格子点只能感知 RoI 中特定位置的点云信息，这会造成目标物体点云结构的几何信息丢失。基于图的方法[5-6]用特征提取能力更强的图卷积替代 PointNet[7]和 PointNet++[8]，通过对目标物体的点云建立拓扑结构进行构图，然后使用图卷积提取特征，但是对点云建立简单的拓扑结构依然只能粗略地保留物体的几何信息。

为了解决上述问题，本文提出了一种基于动态加权图卷积的三维目标检测方法(Dynamic Weight Graph R-CNN，DEG R-CNN)，该方法可以在 RoI 中充分提取目标物体的点云结构信息进行检测。动态加权图卷积通过连接节点周围的主邻点和次邻点，为目标物体建立更好的拓扑结构，充分利用了点云的结构信息。相比于 Graph R-CNN[6]使用的动态图卷积(Dynamic Graph CNN，DGCNN)，动态加权图卷积使用的参数更少、速度更快，更符合自动驾驶的应用场景。为了提高三维目标检测的性能，本文还引入了自适应粒度融合，它仅在 RoI 中自适应地融合不同粒度的图像特征，能够灵活地为检测提供丰富的语义信息。

2 相关工作

2.1 基于激光雷达的三维目标检测

点云是一种稀疏的、不规则的三维表示形式，由激光雷达生成，能够描述物体的三维信息。根据点云的表示类型，基于激光雷达的三维目标检测可以分为基于点的方法和基于体素的方法。

基于点的方法直接以原始点作为输入，并采用多层感知机提取点云特征，预测 3D 检测框。PointRCNN[1]使用语义分割网络将点云划分为前景点和背景点，然后在每个前景点上生成提议框。3DSSD[2]在一阶段检测中完全使用点级预测，并且在一个类似于 PointNet[7]特征提取器后面设计了一个无锚框的检测头。Point-GNN[5]提出了一种自动配准机制来减少变换方差，并且设计了边框合并评分操作，来准确地结合多个顶点的预测。这些基于点的方法虽然保留了准确的三维位置信息，但需要在整个场景中进行采样，提取点云特征，这会忽略很多细节的点云结构信息并带来大量的计算和时间开销。

与基于点的方法不同，基于体素的方法将点云场景体素化，生成稀疏的体素场景，然后使用三维稀疏卷积提取特征。VoxelNet[9]是一个端到端的三维目标检测网络，它将点云体素化后，使用体素特征提取层将所有体素中的点转换成统一的表示，再用三维卷积提取体素特征，并将提取的体素特征压缩，送入区域建议网络中进行检测。SECOND[10]提出了一种只在非空体素上进行特征提取的三维稀疏卷积替代三维卷积，由于整个场景中非空体素的占比很小，因此模型训练和推理的速度得到了显著的提升，它也成为目前三维目标检测最常用的骨干网络之一。PointPillar[11]直接将点云转换成伪图像进行目标检测，它把点云场景划分成若干个柱体，在每个柱体中随机采样 N 个点作为柱体的特征，通过这种方法完成点云到伪图像的转换；然后使用二维卷积提取伪图像的特征，将伪图像特征送入检测头中进行检测。虽然 PointPillar 的速度得到了极大的提升，但是由于在点云向伪图像转换的过程中损失了高度信息，物体的空间结构也遭到了破坏，因此检测的精度不如使用体素的检测模型。PV-RCNN[3]将点云场景体素化，使用三维稀疏卷积提取特征，经区域提议网络得到初始的检测框，再提取点的特征对检测框进行细化。它同时利用了体素和点的特征，兼顾了检测的速度和精度。Voxel R-CNN[4]证明了对于高性能的三维目标检测，精确的位置信息并不是必需的，其所有的检测流程都是在分辨率较低的体素空间中进行的，但是性能依然与 PV-RCNN 相近。然而，PV-RCNN 和 Voxel R-CNN 都是通过在 RoI 中设置均匀的格子点聚合物体的点云特征，这在一定程度上造成了物体几何结构信息的损失。

Graph R-CNN[6]使用特征提取能力更强的动态图卷积细化检测框，并使用动态最远体素采样捕获点云，但是它在目标物体的点云上建立的拓扑结构只能粗略地保留物体的几何信息。基于体素的方法使用三维稀疏卷积提取特征，然后在 RoI 中提取点云的结构信息对检测框进行细化。本文通过使用动态加权图卷积为目标物体建立更完备的拓扑结构，以此来充分地利用物体的点云结构信息。

2.2 基于多模态融合的三维目标检测

基于多模态融合的三维目标检测同时利用了图像和点云的优点，使模型的检测性能得到了很大的提升，并成为当前最主流的三维目标检测方法之一。

PointPainting[12]将激光雷达点投影到对应图像的语义分割图中，然后把图像的类别分数附加到对应的激光雷达点上。它利用 RGB 图像的语义信息为点云提供了简洁、明确的特征，使目标与背景之间的辨识度更清晰，显著提高了三维目标检测的性能。PointAugmenting[13]直接使用 2D 检测模型提取的图像特征修饰对应的点云，然后对修饰后的点云进行三维目标检测。与高度抽象的语义分割分数相比，从 2D 检测网络中得到的特征具有更大的感受野，点云能够从图像中获取更丰富的语义信息。然而，简单的拼接只能机械地结合两种模态的信息，因此 DeepFusion[14]利用交叉注意力机制融合点云和图像特征，使得两种异构信息能够更好地相互补充。

常规的融合方法通过标定矩阵将图像和点云进行对齐，但是受设备和外界环境的干扰，图像和点云并不能完全对应，进而产生次优的检测结果。因此，AutoAlign[15]在全局将图像和点云进行对齐，通过自行设计的交叉注意特征对齐模块，自适应地聚合体素对应的图像特征。尽管 AutoAlign 能够将点云和图像自适应对齐，但是它需要在全局使用交叉注意力机制，这无疑会带来很大的计算开销。因此，AutoAlign V2[16]被提出。它通过标定矩阵将体素映射到对应的图像位置，再以映射位置为原点，寻找与体素相关性最高的图像特征。由于只需要在局部使用交叉注意力，因此模型的计算开销也能得到极大的缓解。LoGoNet[17]使用了类似的自适应对齐模块融合图像与点云的特征，并且提出在全局和局部同时进行对齐融合，这样不仅能够使模型对全局有清晰的感知，还能通过融合细粒的局部特征对检测框进行更细致的微调。

受 LoGoNet 的启发，考虑到细粒度的图像特征和粗粒度的图像特征关注的语义信息也是不同的，本文在融合多模态特征阶段，使用自适应粒度融合模块在 RoI 中融合不同粒度的图像特征，更好地为点云补充图像的语义特征信息。

3 DEG R-CNN

3.1 网络模型

DEG R-CNN 的网络结构如图 1 所示，它由点云分支和图像分支两部分构成。本文使用 SECOND[10] 作为点云分支的三维骨干网络，使用 CenterNet[18] 作为图像分支的二维骨干网络，并且将点云和图像作为两个分支的输入。其中，SECOND 使用 3D 稀疏卷积提取三维体素空间中的特征，经过多次特征提取和下采样后，沿高度维将三维体素特征压缩，获得二维的鸟瞰图特征图；CenterNet 通过深层聚合的方式迭代、分层合并图像信息，然后使用可变形卷积与上采样提取图像特征。

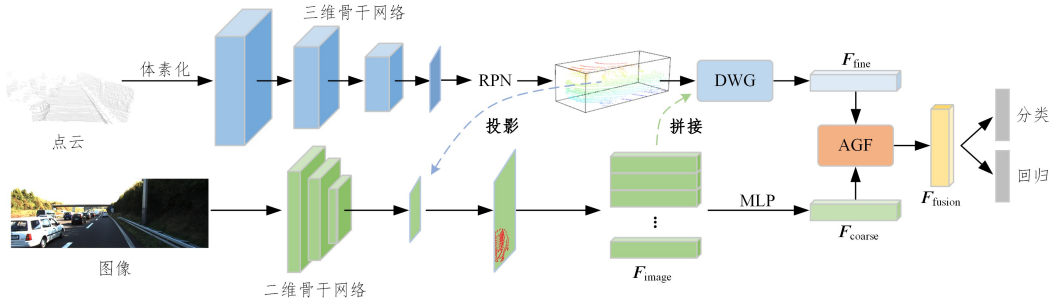

图 1 DEG R-CNN 的网络结构

Fig. 1 DEG R-CNN network structure

在点云分支中，首先将点云进行体素编码，得到体素空间 $V\in\mathbb{R}^{W\times H\times L}$，接着将得到的体素特征送到三维骨干网络，经过后面的区域提议网络（Region Proposal Network，RPN）获得初始提议框，最后将 RoI 中的点云送入动态加权图卷积（Dynamic Weight Graph Convolution，DWG）模块细化边界框。在图像分支中，使用二维骨干网络提取图像特征，然后通过校准矩阵将 RoI 中的点云投影到图像特征图中，找到点云对应的图像特征，最后使用自适应粒度融合（Adaptive Granularity Fusion，AGF）模块融合不同粒度的图像特征，为检测提供丰富的语义信息。

3.2 动态加权图卷积

传统的动态图卷积[19] 根据特征距离为点云构建图结构，然后使用边缘卷积（Edge Convolution）聚合图中每个节点的局部特征。它在构图时仅连接与节点特征距离最近的 K 个邻点，通过这种方式建立的拓扑结构只能粗略地捕获物体的轮廓信息，物体边缘的结构信息很容易被忽略。此外，由于需要对每个节点的所有邻点提取特征，模型的开销也会随着 K 的增大而显著增加。

为了解决上述问题，本文提出了动态加权图卷积，它能为点云建立更完善的拓扑结构，并高效地提取点云特征，充分地利用了目标物体点云的结构信息。该方法在构建图结构时，将节点周围的邻点划分为主邻点和次邻点，每个节点通过它附近的主邻点和次邻点来捕获当前位置的局部特征；在聚合特征时，使用 Gaussian 函数对节点与邻点之间的特征距离进行归一化，然后对每个邻点使用一维卷积提取特征，最后将所有加权特征聚合，更新原节点的特征。图 2 给出了动态加权图卷积的过程。

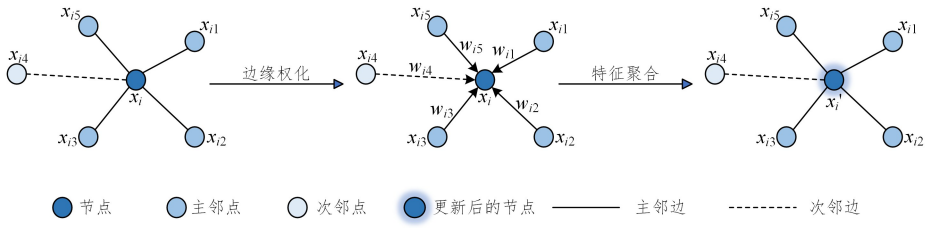

图 2 动态加权图卷积

Fig. 2 Dynamic weight graph convolution

3.2.1 构建图结构

首先，利用 Graph R-CNN[6] 提出的动态最远体素采样得到采样点 $P=\{p_i=[x_i,y_i,z_i,r_i],i=1,2,\cdots,N\}$。然后，在采样点上构建图 $G=\{V,E\}$，并且使用采样点 $p_i\in P$ 表示节点 $v_i\in V$，使用 $e_{im}\in E$ 表示连接节点 v_i 和 v_m 的边。在构造图 G 时，需要连接节点与周围的邻点来捕获当前节点位置的局部特征。

传统的动态图卷积使用 K 近邻算法（K-Nearest Neighbor，KNN）聚合节点附近 K 个最近邻点的特征，但是仅通过近邻点建立的拓扑结构并不能充分地利用点云的结构信息。因此，根据邻点对节点提供的结构信息的不同，将节点的邻点划分为主邻点和次邻点。其中，距离当前节点最近的 K 个邻点称为当前节点的主邻点；除 K 个最近邻点外，以当前节点为主邻点的节点称为当前节点的次邻点。主邻点和次邻点的生成如图 3 所示。

在图 3 中，取最近邻点数量 $K=4$，并分别以点 x_0，x_4，x_5 作为节点。当以 x_0 作为节点时，距离 x_0 最近的 K 个点为 x_1，x_2，x_3，x_4，因此 x_0 的主邻点为 x_1，x_2，x_3，x_4，没有次邻点（假设此时没有点以 x_0 作为主邻点）。当以 x_4 作为节点时，距离 x_4 最近的 K 个点为 x_5，x_6，x_7，x_8，因此 x_4 的主邻点为 x_5，x_6，

x_7，x_8；由于 x_0 以 x_4 作为主邻点且不在 x_4 的 K 个最近邻点中，因此 x_0 是 x_4 的次邻点。当以 x_5 作为节点时，距离 x_5 最近的 K 个点为 x_2，x_4，x_6，x_7，因此 x_5 的主邻点为 x_2，x_4，x_6，x_7；由于 x_4 属于 x_5 最近的 K 个邻点中，因此 x_4 即使以 x_5 作为主邻点，也不是它的次邻点。

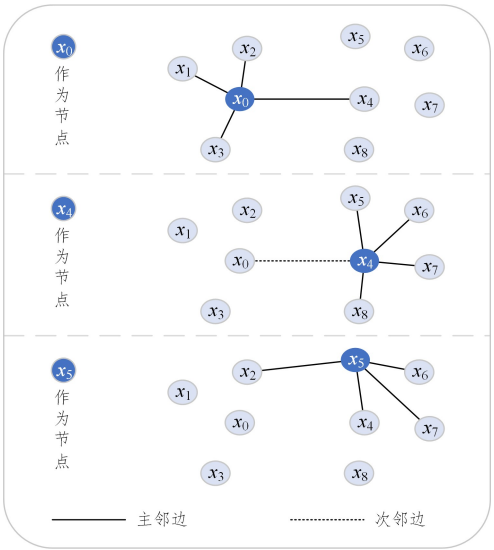

图 3　主、次邻点生成示意图($K=4$)

Fig. 3　Generation of primary and subordinate neighbors

DGCNN[19]在构造图结构时只关注到节点的主邻点，忽略了节点的次邻点。在点云中，由于主邻点距离节点较近，主邻点与节点之间的相关性比次邻点与节点之间的相关性更强。但从全局视角看，次邻点通常位于目标物体的边缘，具有丰富的边缘信息，它们包含的这些细节的边缘信息具有物体最重要的结构属性，是区分相似物体的关键。因此，DWG 能够通过聚合次邻点的特征，更好地获取相似物体之间的差异，减少模型的误检。图 4 给出了主邻点和次邻点的分布。

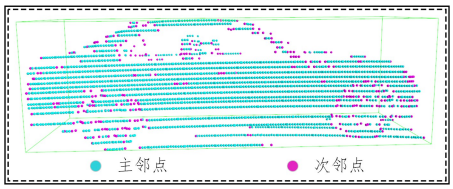

图 4　主邻点和次邻点的分布

Fig. 4　Distribution of primary and subordinate points

本文方法考虑了主邻点与次邻点对节点的影响，为 RoI 中物体的点云建立了一个更好的拓扑结构。它不仅能够完整地聚合点云的空间结构特征，而且可以更好地捕捉目标的边缘信息，对目标物体的分类和回归具有重要意义。

3.2.2　聚合特征

在构造初始的图结构后，需要聚合图中节点与其邻点的特征。DGCNN[19]将每个节点和邻点的关联信息存储在连接邻点的边上，对连接节点和邻点的边使用边缘卷积。但是它需要对所有的邻边使用边缘卷积，这大大增加了模型的参数量和推理时间。因此，本文提出了一种新的操作来聚合图的特征。首先，根据节点与其邻点之间的距离，使用 Gaussian 函数计算它们的相关性。计算方法如下：

$$w_m=\frac{e^{-d_m}}{\sum_{m=1}^{M}e^{-d_m}} \tag{1}$$

其中，M 为邻点的数量，d_m 为节点到第 m 个邻点的距离，w_m 为节点到第 m 个邻点的权重。在特征空间中，两点之间的距离越近，它们的相关性越强[19]。这里使用 Gaussian 函数对节点与邻点的距离进行归一化，将结果存储在连接各自邻点的边上，并将其记作对应邻点的权重，再利用得到的边缘权重衡量每个邻点与对应节点的相关性。然后，通过边缘权重对节点与邻点的特征进行聚合。聚合公式如下：

$$x_i'=Convld\left(\sum_{m=1}^{M}w_m\cdot x_{im},x_i\right) \tag{2}$$

其中，x_i 为第 i 个节点的特征，x_{im} 为第 i 个节点的第 m 个邻点的特征，x_i' 为节点 x_i 更新后的特征。为了能够同时利用节点和邻点的局部信息，将节点特征与邻点特征的加权和进行拼接，然后对拼接后的特征进行一维卷积。图结构中所有节点执行此操作后，每个更新的节点将捕获对应位置的邻域信息。由于此时每个节点都包含各自的局部信息，下一次迭代会继续将当前节点的局部信息传播到邻域。若干次迭代后，图中每个节点将会捕获目标物体的全局信息。

3.3　自适应粒度融合

不同粒度的图像特征携带的语义信息不同，为了充分利用粗粒度和细粒度的图像特征，本文提出了一个 AGF 模块，它利用注意力机制自适应地融合这两种粒度的特征。由于户外场景中的激光雷达点数量较多，将所有的点投影到图像上来会产生很大的开销，因此仅将 RoI 中的点通过校准矩阵投影到图像上获得细粒度的图像特征，以最大程度地降低不必要的开销。获得细粒度的图像特征后，将其与对应的点拼接，然后送入到 DWG 模块，获得具有图像细粒度信息的细粒度 RoI 特征。同时，将上面的细粒度图像特征再送入 MaxPooling 层，获得粗粒度 RoI 特征。值得注意的是，这里获得的 RoI 图像特征不同于以前的方法[20-22]获得的图像特征。具体如图 5(a)所示，之前的方法将三维空间中的 RoI 投影到图像中，得到投影在图像上的二维 RoI，然后从二维 RoI 中提取图像特征。

现实场景在二维空间中的表示与三维空间不同，二维图像 RoI 中的物体会受其他物体或背景的干扰，尤其是对被遮挡的物体。本文方法将三维 RoI 中的点云投影到图像中，如图 5(b)所示，此时点云投影位置的图像特征恰好是当前物体的图像特征，受到其他物体或背景特征的干扰很小。在获得两种不同粒度的 RoI 特征后，利用 AGF 模块自适应地融合两种不同粒度的图像信息，得到最终融合的 RoI 特征。AGF 模块的流程如图 6 所示。

(a)2D RoI 点云投影　　(b)3D RoI 点云投影

图 5　2D RoI 和 3D RoI 中投影点的对比

Fig. 5　Comparison of points in 2D RoI and 3D RoI

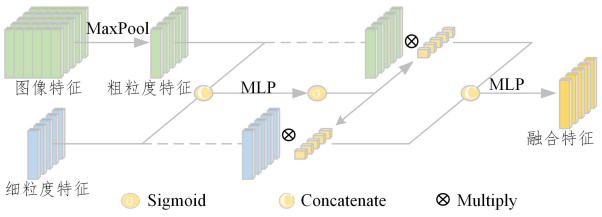

图 6 AGF 模块

Fig. 6 AGF module

在获得了 RoI 中的细粒度特征和粗粒度特征后，需要先将两种不同粒度的特征进行拼接，再把拼接后的特征送入到多层感知机(Multi-layer Perceptron，MLP)中，使用 Sigmoid 函数将得到的结果归一化，并将归一化结果作为 RoI 中两种粒度特征的权重，最后再将两种加权后的 RoI 特征送入到下一个 MLP 中，从而得到最终 RoI 中的融合特征。RoI 中的融合特征的计算如下：

$$(w_i^{\text{coarse}}, w_i^{\text{fine}}) = \sigma(MLP(Concat(f_i^{\text{coarse}}, f_i^{\text{fine}}))) \tag{3}$$

$$f_i = MLP(Concat(w_i^{\text{coarse}} f_i^{\text{coarse}}, w_i^{\text{fine}} f_i^{\text{fine}})) \tag{4}$$

其中，$f_i^{\text{coarse}} \in \mathbb{R}^{n \times C}$ 和 $f_i^{\text{fine}} \in \mathbb{R}^{n \times C}$ 分别表示第 i 个 RoI 中的粗粒度特征和细粒度特征，n 为 RoI 的个数，C 是 RoI 特征通道数，f_i 表示第 i 个 RoI 中的融合特征。

3.4 损失函数

本文检测网络的总损失与之前的工作[3-4]相同，由 RPN 损失 L_{rpn} 和 RoI 损失 L_{roi} 两部分组成。计算方法如下：

$$L_{\text{total}} = L_{\text{rpn}} + L_{\text{roi}} \tag{5}$$

其中，RoI 损失 L_{roi} 与 RPN 损失 L_{rpn} 类似，由分类损失 L_{cls} 和回归损失 L_{reg} 两部分组成。以 L_{rpn} 为例，L_{rpn} 的计算如下：

$$L_{\text{rpn}} = L_{\text{cls}} + \lambda L_{\text{reg}} \tag{6}$$

其中，λ 是平衡这两个损失的超参数，默认设置为 1。

4 实验结果与分析

4.1 数据集与评估指标

本文使用 KITTI 数据集[23]对 DEG R-CNN 进行评估。该数据集已在 KITTI 官方网站上公开，包含自动驾驶场景的 7481 个样本，这些样本通常被分为 3712 个训练样本和3769 个测试样本。为了保证结果的公平性，本文将遵循上述的划分原则来比较三维目标检测的性能，并且使用 KITTI 官方提供的评估指标即平均准确度(Mean Average Precision，mAP)来衡量模型性能。其中，交并比(Intersection over Union，IoU)阈值设置为 0.7，并通过 40 个召回位置计算三维目标检测的准确率。

4.2 实验细节

根据之前的工作，将点云划分为体素后送入到检测网络中，使用 SECOND[10]作为 RPN，在 RPN 后生成若干提案框。在 RoI 中进行构图时，设置主邻点 $K=4$，迭代次数 $T=3$。为了验证本文的方法对多模态数据同样有效，同时使用 CenterNet[18]提取图像特征作为输入。

首先，使用 4 个 GTX 1080Ti GPU 训练骨干网络和 RPN。训练周期设置为 80，每个 GPU 的 batchsize 设置为 4。然后，冻结预训练的 RPN 和骨干网络参数，使用 2 张 Tesla P100 GPU 训练检测头，并将训练周期设置为 80，学习率设置为 0.0015，每个 GPU 的 batchsize 设置为 16。

4.3 实验结果对比

在 KITTI[23]数据集上进行实验，并将所提方法与主流的三维目标检测方法进行对比。实验结果如表 1 所列。其中 DEG-S 仅使用激光雷达作为传感器，DEG-M 同时使用激光雷达和相机作为传感器。可以看到，本文方法获得了具有竞争力的检测结果。

表 1 KITTI 数据集三维目标检测结果的对比

Table 1 Comparison with state-of-the-art methods on KITTI for 3D detection

(%)

方法	模态	3D 检测准确率($IoU=0.7$)			
		简单	中等	困难	mAP
VoxelNet[9]	LiDAR	81.97	65.46	62.85	70.10
Point RCNN[1]	LiDAR	88.88	78.63	77.38	81.63
SECOND[10]	LiDAR	88.15	78.33	75.25	81.54
Point-GNN[5]	LiDAR	87.89	78.34	77.38	81.20
3DSSD[2]	LiDAR	89.71	79.45	78.67	82.61
SA-SSD[24]	LiDAR	90.15	79.91	78.78	82.95
PV-RCNN[3]	LiDAR	92.57	84.83	82.69	86.70
Voxel R-CNN[4]	LiDAR	92.38	85.29	82.86	86.84
OcTr[25]	LiDAR	88.47	78.57	77.16	81.40
Graph RCNN[6]	LiDAR	93.33	86.12	83.29	87.58
MV3D[22]	LiDAR+RGB	71.29	62.68	56.56	63.51
3D-CVF[26]	LiDAR+RGB	92.28	79.88	78.47	82.67
CLOCs[27]	LiDAR+RGB	89.67	85.94	83.25	87.25
EPNet[28]	LiDAR+RGB	92.28	82.59	80.14	85.00
FocalsConv[29]	LiDAR+RGB	92.26	85.32	82.95	86.84
VFF[30]	LiDAR+RGB	92.31	85.51	82.92	86.91
LoGoNet[17]	LiDAR+RGB	92.04	85.04	**84.31**	87.13
DVF-V[31]	LiDAR+RGB	93.07	85.84	83.13	87.34
PA3DNet[32]	LiDAR+RGB	93.18	86.02	83.74	87.65
VoPiFNet[33]	LiDAR+RGB	—	—	—	84.24
DEG-S(本文)	LiDAR	95.08	85.75	82.74	87.86
DEG-M(本文)	LiDAR+RGB	**95.71**	**86.77**	83.93	**88.80**

根据实验结果，本文提出的方法在简单、中等、困难 3 个难度级别较之前的方法均有不同程度的提升。在仅使用激光雷达作为传感器的方法中，DEG-S 的 3D mAP 超过了其他所有的单模态方法，尤其是在检测的难度级别为简单时。DEG-M 额外融合了来自相机的图像信息，在多模态融合方法中，它的 3D mAP 也是最高的，甚至优于最新的多模态方法 PA3DNet[32]。

此外，将所提方法与其他基于图的三维目标检测方法进行了比较，如表 2 所列。相比于 Point-GNN[5]和 Graph RCNN[6]，DEG-S 的 3D mAP 具有更显著的优势。其中，DEG-S 的 3D mAP 比 Point-GNN 高 6.66%。

表 2 基于图的方法的对比

Table 2 Comparison with graph-based methods

(%)

方法	3D 检测准确率($IoU=0.7$)			
	简单	中等	困难	mAP
Point-GNN[5]	87.89	78.34	77.38	81.20
Graph RCNN[6]	93.33	86.12	83.29	87.58
DEG-S	95.08	85.75	82.74	87.86

因为 Point-GNN 需要在整个点云场景中构建图，所以它

的参数量和推理时间远远高于 DEG-S 和 Graph RCNN。因此，这里只将 Graph RCNN 与 DEG-S 进行对比，结果如表 3 所列。可以看到，DEG-S 中 DWG 的参数量和推理时间都小于 Graph RCNN 中 DGCNN 的参数量和推理时间。这是由于提取特征时，DEG R-CNN 使用 Gaussin 函数和 1D 卷积代替了 Graph RCNN 的 2D 卷积，提高了特征提取的效率。

表 3 参数与推理时长的对比

Table 3 Comparison on parameters and runtime

方法	参数	时长/ms
Graph RCNN[6]	5.04×10^7	39.2
DEG-S	1.28×10^7	36.3

Point-GNN，Graph RCNN 与 DEG R-CNN 的可视化结果如图 7 所示。其中，第一排是相机视图，后面 3 排分别为 Point-GNN，Graph RCNN 和 DEG R-CNN 在点云场景中的检测结果。可以看到，在第一、第二个场景中的货车和第三个场景中的围墙与汽车具有相似的点云结构，这容易造成模型的误检。而本文使用的 DWG 能够为物体点云建立更完备的拓扑结构，捕捉目标物体的细节信息，提高检测相似目标物体的性能。在第四、第五个场景中，Point-GNN 和 Graph RCNN 漏检的目标物体受视角的遮挡，只能显示部分点云；但是 DEG R-CNN 能够充分地利用点云的结构信息进行推理，并得到目标物体完整的检测框。

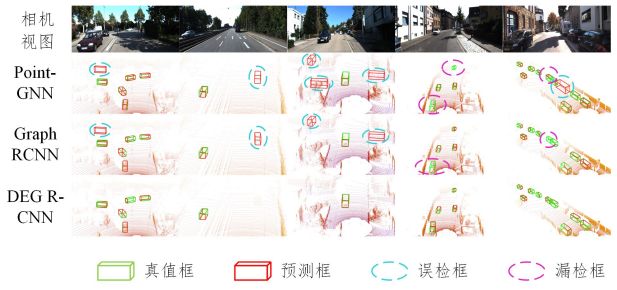

图 7 Point-GNN，Graph RCNN 与 DEG R-CNN 的可视化结果

Fig. 7 Visualization of Point-GNN，Graph RCNN and DEG R-CNN

4.4 消融实验

本文在 KITTI[23] 数据集上进行消融实验，验证了 DEG R-CNN 模块的有效性；同时对图卷积的迭代次数进行探索，得到了最优的三维目标检测结果。

4.4.1 模块有效性

将 DEG R-CNN 的 DWG 与 Graph RCNN 的 DGCNN[19] 进行比较，如表 4 所列。结果表明，使用 DWG 替代 DGCNN 可以将检测的 3D mAP 提高 0.28%。值得注意的是，在简单的难度级别，DWG 的检测准确率比 DGCNN 高 1.75%。这是因为简单级别的目标物体通常具有充足的原始点云表示，DWG 更容易在这些原始点云上为目标物体构造更好的拓扑结构。

然后，在 DWG 模块后加入 AGF 模块自适应地融合不同粒度的图像特征。与仅使用 DWG 模块的 DEG-S 相比，AGF 模块使检测的 3D mAP 提高了 0.94%，并且在所有难度的级别中，检测结果都有不同程度的提升。这证明了对于各个难度级别的目标物体，AGF 模块都能有效地融合点云与图像的特征。

表 4 各个模块的影响

Table 4 Impact of each module

(%)

DGCNN[19]	DWG	AGF	3D 检测准确率($IoU=0.7$)			
			简单	中等	困难	mAP
√			93.33	86.12	83.29	87.58
	√		95.08	85.75	82.74	87.86
	√	√	95.71	86.77	83.93	88.80

4.4.2 融合方法

为了证明 AGF 模块能够有效地融合不同粒度的图像特征，将 AGF 与 Add 和 Concat 方法进行比较，结果如表 5 所列。Concat 比 Add 的 3D mAP 高 0.55%，这是由于不同粒度的图像特征表示的信息具有差异，Add 对特征进行加和会造成图像信息的损失，Concat 则会最大程度地保留两种粒度的图像信息。AGF 比 Concat 更有效，并且在简单、中等和困难 3 个难度级别的检测准确率分别比 Concat 高出 0.27%，0.21%和 0.18%。这是由于不同粒度的图像特征侧重的信息不同，简单的拼接只能粗略地保留它们各自的特征。然而，AGF 可以通过注意力机制有选择地融合不同粒度的图像特征，重点关注表示物体信息的关键特征。

表 5 融合结果的对比

Table 5 Comparison of fusion results

(%)

融合方法	3D 检测准确率($IoU=0.7$)			
	简单	中等	困难	mAP
Add	94.89	86.31	82.88	88.03
Concat	95.44	86.56	83.75	88.58
AGF	95.71	86.77	83.93	88.80

4.4.3 图卷积迭代

为了探究图卷积迭代次数对检测性能的影响，将迭代次数 T 分别设置为 4，6，8，10，比较不同迭代次数 T 的检测结果，如表 6 所列。可以看到，简单级别的检测性能随着迭代次数的增加逐渐提高，但检测的 3D mAP 在 $T=6$ 时最高。这是因为简单级别中目标物体的点云数量较多，迭代次数越多，则可以更充分地提取目标物体的点云结构特征。但是，中等和困难级别中物体的点云相对稀疏，经过较少的迭代，就可以提取出物体的结构信息，更多的迭代只会导致模型性能的降级。

表 6 图迭代次数 T 的影响

Table 6 Influence of the number of convolution iterations of graph T

(%)

T	简单	中等	困难	mAP
4	95.61	86.39	83.57	88.52
6	95.71	86.77	83.93	88.80
8	95.73	86.57	83.76	88.69
10	95.82	86.31	83.44	88.52

结束语 针对现有方法不能充分地利用点云结构信息的问题，提出了 DEG R-CNN。该方法利用主邻点和次邻点，在原始点云中建立更完备的拓扑结构来提取物体的结构信息。此外，本文还引入了自适应粒度融合模块来融合不同粒度的图像特征。在 KITTI 数据集上进行的广泛实验中，DEG R-

CNN获得了具有竞争力的结果。未来将会考虑使用交叉注意力机制，根据目标物体点云的稀疏来自适应融合点云与图像的特征，提高三维目标检测的性能。

参 考 文 献

[1] SHI S S, WANG X G, LI H S. PointRCNN: 3D object proposal generation and detection from point cloud[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 770-779.

[2] YANG Z T, SUN Y N, LIU S, et al. 3DSSD: Point-based 3D single stage object detector[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 11040-11048.

[3] SHI S S, GUO C X, JIANG L, et al. PV-RCNN: Point-voxel feature set abstraction for 3D object detection[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 10529-10538.

[4] DENG J J, SHI S S, LI P W, et al. Voxel R-CNN: Towards high performance voxel-based 3D object detection[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2021: 1201-1209.

[5] SHI W, RAJKUMAR R. Point-GNN: Graph neural network for 3D object detection in a point cloud[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 1711-1719.

[6] YANG H L, LIU Z L, WU X P, et al. Graph R-CNN: Towards accurate 3D object detection with semantic-decorated local graph [C]// European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022: 662-679.

[7] QI C R, SU H, MO K, et al. PointNet: Deep learning on point sets for 3D classification and segmentation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 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[J]. arXiv: 1706.02413, 2017.

[9] ZHOU Y, TUZEL O. Voxelnet: End-to-end learning for point cloud based 3d object detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 4490-4499.

[10] YAN Y, MAO Y X, LI B. SECOND: Sparsely embedded convolutional detection[J]. Sensors, 2018, 18(10): 3337.

[11] LANG A H, VORA S, CAESAR H, et al. PointPillars: Fast encoders for object detection from point clouds[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 12697-12705.

[12] VORA S, LANG A H, HELOU B, et al. PointPainting: Sequential fusion for 3D object detection [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 4604-4612.

[13] WANG C W, MA C, ZHU M, et al. PointAugmenting: Cross-modal augmentation for 3D object detection[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 11794-11803.

[14] LI Y, YU A W, MENG T, et al. DeepFusion: Lidar-camera deep fusion for multi-modal 3D object detection[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 17182-17191.

[15] CHEN Z, LI Z, ZHANG S, et al. AutoAlign: Pixel-instance feature aggregation for multi-modal 3D object detection[J]. arXiv: 2201.06493, 2022.

[16] CHEN Z, LI Z, ZHANG S, et al. Deformable feature aggregation for dynamic multi-modal 3D object detection [C] // European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022: 628-644.

[17] LI X, MA T, HOU Y N, et al. LoGoNet: Towards accurate 3D object detection with local-to-global cross-modal fusion [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 17524-17534.

[18] ZHOU X Y, WANG D Q, KRÄHENBÜHL P. Objects as points [J]. arXiv: 1904.07850, 2019.

[19] WANG Y, SUN Y B, LIU Z W, et al. Dynamic graph cnn for learning on point clouds [J]. ACM Transactions on Graphics (tog), 2019, 38(5): 1-12.

[20] 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). IEEE, 2018: 1-8.

[21] LIANG M, YANG B, CHEN Y, et al. Multi-task multi-sensor fusion for 3D object detection[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 7345-7353.

[22] CHEN X Z, MA H M, WAN J, et al. Multi-view 3D object detection network for autonomous driving[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 1907-1915.

[23] 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.

[24] HE C H, ZENG H, HUANG J Q, et al. Structure aware single-stage 3D object detection from point cloud[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 11873-11882.

[25] ZHOU C, ZHANG Y N, CHEN J X, et al. OcTr: Octree-based transformer for 3D object detection [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 5166-5175.

[26] YOO J H, KIM Y, KIM J, et al. 3D-CVF: Generating joint camera and lidar features using cross-view spatial feature fusion for 3D object detection [C] // Computer Vision-ECCV 2020: 16th European Conference. Springer International Publishing, 2020: 720-736.

[27] PANG S, MORRIS D, RADHA H. CLOCs: Camera-LiDAR object candidates fusion for 3D object detection[C]// 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems

(IROS). IEEE，2020：10386-10393.

[28] HUANG T T，LIU Z，CHEN X W，et al. EPNet：Enhancing point features with image semantics for 3D object detection[C]// Computer Vision-ECCV 2020：16th European Conference. Springer International Publishing，2020：35-52.

[29] CHEN Y K，LI Y W，ZHANG X Y，et al. Focal sparse convolutional networks for 3D object detection[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022：5428-5437.

[30] LI Y W，QI X J，CHEN Y K，et al. Voxel field fusion for 3D object detection[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022：1120-1129.

[31] MAHMOUD A，HU J S K，WASLANDER S L. Dense voxel fusion for 3D object detection[C]// Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2023：663-672.

[32] WANG M L，ZHAO L，YUE Y F. PA3DNet：3-D vehicle detection with pseudo shape segmentation and adaptive camera-LiDAR fusion[J]. IEEE Transactions on Industrial Informatics，2023，19(11)：10693-10703.

[33] WANG C H，CHEN H W，CHEN Y，et al. VoPiFNet：Voxel-Pixel Fusion Network for Multi-Class 3D Object Detection[J]. IEEE Transactions on Intelligent Transportation Systems，2024，25(8)：8527-8537.

LI Zongmin，born in 1965，Ph.D，professor，Ph. D supervisor，is a member of CCF(No. 11175S). His main research interests include computer graphics，digital image processing and pattern recognition.

（责任编辑：柯颖）

CCF 科技成果评价委员会召开 2025 年首次工作会议

2025 年 2 月 19 日，CCF 科技成果评价委员会召开 2025 年首次工作会议。CCF 副理事长、CCF 科技成果评价委员会主任胡事民院士，委员中科院计算所陈云霁研究员、中国人民大学杜小勇教授、上海交通大学过敏意教授、北京大学胡振江教授、哈尔滨工业大学刘挺教授、东南大学罗军舟教授、南京大学马晓星教授、北京航空航天大学王莉莉教授、中科院软件所张健研究员、浙江大学庄越挺教授，以及主任助理清华大学张松海副教授参加会议。会议就科技成果评价委员会 2024 年工作进行了总结，并对 2025 年的工作开展了研讨。各位委员就成果评价工作规范性、服务水平提升、成果评价工作宣传等方面充分发表了意见和建议。

委员们提出要充分发挥中国计算机学会在计算机领域的独特优势和专家资源，为成果团队做好专家推介，以及 CCF 报奖联动服务；在专业、合规的基础上提高成果鉴定效率，进一步缩短鉴定周期；并通过 CNCC、学会网站和公众号、CCCF 等途径为鉴定成果提供宣传渠道。

2025 年 CCF 科技成果奖的申报将于 5 月份启动，科技成果评价工作即将全面展开，有意开展成果评价的项目团队可联系：CCF 成果评价办公室　贺亚楠　0512－65900856-21　staec@ccf. org. cn

据 CCF 微信公众号