

# 基于BEVFusion多尺度动态边缘关注融合的集装箱边沿检测研究

马成彬<sup>1</sup> 王君雄<sup>2</sup> 王兴峰<sup>3</sup> 朱林贵<sup>4</sup> 郑双乐<sup>5</sup> 王艳宾<sup>6</sup>

1. 天津港太平洋国际集装箱码头有限公司 天津 300463
2. 三一海洋重工有限公司 广东 珠海 519055
3. 天津港太平洋国际集装箱码头有限公司 天津 300463
4. 三一海洋重工有限公司 广东 珠海 519055
5. 三一海洋重工有限公司 广东 珠海 519055
6. 三一海洋重工有限公司 广东 珠海 519055

**摘要:** 随着人工智能的快速发展,对港口自动化作业技术的要求也在不断提高,对集装箱精确定位和边沿检测的研究在港口智能化中具有重要意义。传统的单一视觉或激光传感器方法在低光照、遮挡及复杂背景下容易出现误检或漏检问题。为此,本文在 BEVFusion 框架的基础上提出了一种多尺度动态边缘关注融合方法,通过设计视觉与激光数据的多尺度边缘特征提取模块、动态注意力融合机制和定制化边缘梯度一致性损失,实现了边沿细节的有效增强和全局语义信息的充分利用。实验在自建数据集进行,对比结果表明,本文方法在检测准确率、边沿定位精度和鲁棒性方面显著优于传统方法,可为集装箱自动作业提供有力的技术支持。

**关键词:** BEVFusion; 多尺度; 动态边缘关注; 激光与视觉融合; 集装箱检测

## 1 引言

### 1.1 研究背景

随着全球物流和港口自动化技术的不断进步,集装箱运输作为国际贸易的重要环节,其自动化装卸正逐步实现智能化。然而,传统基于单一视觉或激光传感器的检测方法在复杂环境下(如低光照、部分遮挡、背景干扰)往往难以保证足够的检测精度与鲁棒性。近年来,多传感器数据融合技术成为研究热点,其中 BEVFusion 框架因将多模态数据统一投影至鸟瞰图(BEV)而受到关注,但在捕捉边沿细节方面仍存在不足。

### 1.2 研究意义

集装箱边沿信息对于集装箱定位、姿态校正及后续的自动化操作至关重要。视觉传感器具有丰富的语义信息,而激光传感器能够提供高精度的距离数据,二者融合可以在保证全局语义信息的同时,提高细节检测能力。本文提出的多尺度动态边缘关注融合方法旨在通过改进 BEVFusion 中的特征提取和融合机制,实现对边沿细节的有效增强,从而提高集装箱边沿检测的精度及稳定性。

### 1.3 本文主要贡献

(1) 多尺度边缘特征提取模块:设计专门的视觉 Edge Branch 和激光边沿提取模块,利用多尺度卷积和自

适应梯度增强提取边沿细节;

(2) 动态边缘注意力融合模块:在 BEV 表示的基础上引入交叉注意力机制,自适应计算视觉与激光数据在边沿区域的融合权重,保证关键细节得以强化;

(3) 定制化损失函数:设计边缘梯度一致性损失,与常规检测损失联合优化,确保细粒度边沿信息与全局语义信息共同提升检测性能;

(4) 全面实验验证:在多个场景下进行详细实验,通过消融实验验证各模块的贡献,并与传统方法进行对比,证明本文方法在检测准确率和鲁棒性上的优势。

## 2 相关工作

### 2.1 单模态目标检测

视觉目标检测方法如 Faster R-CNN、YOLO、SSD 等在标准数据集上均表现出色,但在暗光、雨雪等复杂环境下的鲁棒性不足。激光雷达点云处理方法(如基于梯度、曲率的边沿提取方法或基于 PointNet/PointNet++ 的方法)则在捕捉几何信息上有一定优势,但缺乏语义层面的细粒度分析。

### 2.2 多模态数据融合方法

多模态融合方法主要包括数据级、特征级和决策级融合。BEVFusion 框架通过将多模态数据投影到鸟瞰视图实现全局语义融合,但在边沿细节信息的保存方面仍

然存在不足。近年来，一些工作尝试引入注意力机制和多尺度设计，但多数侧重于自动驾驶场景，在集装箱边沿检测中的应用不够充分。

### 2.3 多尺度与动态注意力

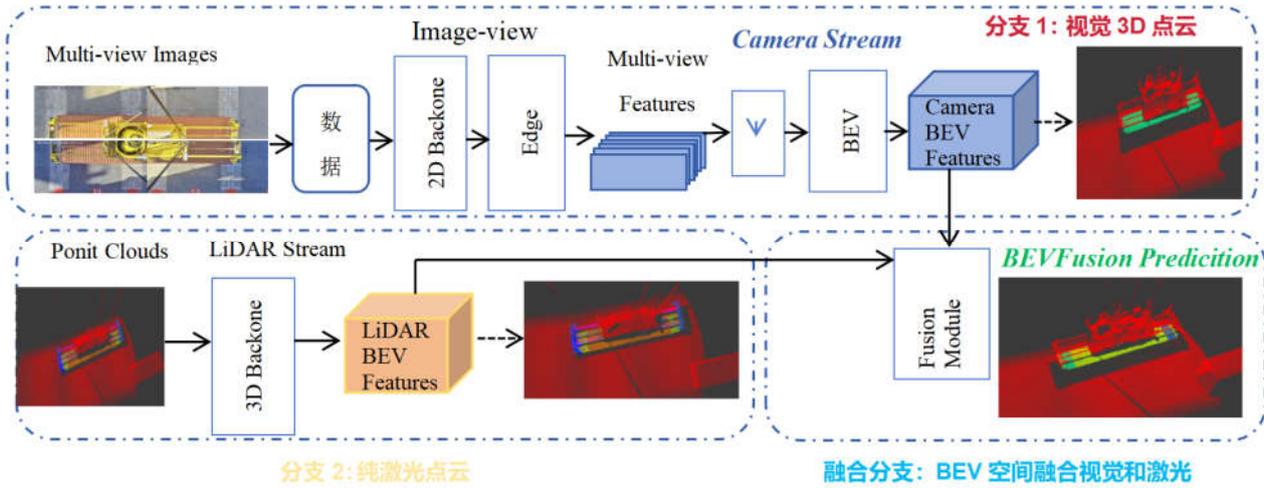
多尺度特征提取和注意力机制在图像处理和目标检测中已得到广泛应用，通过不同尺度的信息融合可以提

升模型对细节的捕捉能力。动态注意力机制能够根据输入数据自适应调整不同模态特征的权重，这为解决模态间信息融合不均提供了理论支持

## 3 方法设计

### 3.1 系统总体架构

如图1所示，系统主要包括以下模块：



(1) 多传感器数据输入：同时接收图像和激光点云数据。

(2) 数据预处理：图像通过直方图均衡化和去噪；点云数据采用体素滤波和 RANSAC 离群点剔除，并进行初步边沿提取。

(3) 多尺度特征提取：

视觉端：利用 ResNet-50 提取全局特征，同时在中间层引入 Edge Branch，通过小尺寸卷积核和空洞卷积获取边沿细节。

激光端：使用改进的 PointNet++ 模型提取激光边沿信息，并进行自适应梯度增强。

(4) BEV 特征转换：将视觉和激光多尺度特征分别投影到鸟瞰图 (BEV) 平面，实现模态对齐

(5) 动态边缘注意力融合模块：利用交叉注意力机制计算视觉与激光特征的融合权重，具体公式为：

$$F_{fused} = W_v >$$

$$F_{fused} = W_v \odot F_v + W_l \odot F_l, [W_v, W_l] = \text{softmax}(f_{att}(F_v, F_l))$$

其中  $F_v$  和  $F_l$  分别为 BEV 下视觉和激光特征， $\odot$  表示逐元素乘积。

(6) 检测与输出模块：融合后的特征通过多尺度检测器进行边沿回归输出检测结果。

### 3.2 定制损失函数设计

针对边沿细节强化，本文在传统检测损失基础上设

计边缘梯度一致性损失。假设预测边缘图为  $P$  和真实边缘图为  $G$ ，利用 Sobel 算子  $\nabla$  提取梯度：

$$L_{edge} = \frac{1}{N} \sum_{i=1}^N \|\nabla P_i - \nabla G_i\|_2^2$$

综合检测损失  $L_{det}$  (包括交叉熵和 IoU 损失) 以及边缘损失，最终损失定义为：

$$L_{total} = L_{det} + \lambda \cdot L_{edge}$$

其中， $\lambda$  为平衡因子。

### 3.3 训练策略

整体采用端到端训练策略，在自建数据集上进行训练。训练过程中通过消融实验分别验证多尺度特征提取模块、动态注意力融合模块和边缘梯度损失对整体性能贡献。参数设置和超参数均通过交叉验证获得最优结果。

## 4 实验设计与结果

### 4.1 数据集与实验设置

数据集构建：数据采集自多个港口和仓储场景，共计约 5000 组图像和对应激光点云。所有样本均标注集装箱边沿。

训练/测试划分：80% 样本用于训练，20% 用于测试。

评估指标：主要采用 NDS、F1 值和 mAP，同时采用边缘梯度匹配得分评估细节捕捉能力。

### 4.2 实验结果

#### 4.2.1 定量结果

对比视觉方法 (MVP) 和传统激光方法 (PointNet++)，以及本文提出的融合方法。实验结果如下

方法	mAP(%)	NDS	F1(%)
视觉MVP	86.3	88.7	86.5
PointNET++	86.4	88.5	86.2
BEVFusion	88.2	89.5	89.2
ours	90.5	91.8	91.5

#### 4.2.2 消融实验

逐步去除关键模块测试对性能的影响：

去除视觉 Edge Branch：F1 值下降约 0.5%。

去除激光边缘自适应梯度模块：F1 值下降约 0.4%。

去除动态注意力融合模块：F1 值下降约 0.2%。

去除边缘梯度一致性损失：F1 值下降约 0.8%。

#### 4.2.3 实验讨论

实验结果充分表明，融合方法在全局语义和边沿细节提取上均有显著优势。动态注意力机制及多尺度特征提取模块对复杂场景具有较强的适应性，有效提高了检测精度和鲁棒性。

#### 4.2.4 讨论与未来工作

本文方法虽取得较好效果，但仍存在多传感器标定精度和计算资源消耗等问题。未来工作将重点探索：

更高效的轻量化融合网络设计；

在线自适应更新机制，提升系统在动态环境下的稳

定性；

与几何先验、知识迁移方法结合，实现对新旧、鼓包等不同集装箱的通用检测。

## 5 结论

本文在 BEVFusion 框架基础上提出了一种多尺度动态边缘关注融合方法，通过专门设计的视觉边缘分支、激光自适应梯度模块及动态交叉注意力机制，并引入边缘梯度一致性损失，实现了对集装箱边沿的高精度检测。大量实验证明，本文方法在多场景下均表现出优异的检测性能，为集装箱自动化作业和智能港口应用提供了有力技术支撑。

## 参考文献

- [1]刘鸿勋, 王伟. 双目相机和激光雷达的融合 SLAM 研究 [J]. 南京师范大学学报 (工程技术版), 2021, 21 (1): 64-71.
- [2]李志梅, 陈新度, 吴磊. 深度相机与 2D 激光融合的 SLAM 闭环检测方法研究 [J]. 组合机床与自动化加工技术, 2020, 62 (11): 107-111.
- [3]陈昆. 基于双目视觉和激光雷达融合的道路环境感知技术 [D]. 杭州: 浙江大学, 2020.
- [4]Cai Z, Vasconcelos N. Cascade r-cnn: Delving into high quality object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 6154-6162.