DOI: 10.19650/j.cnki.cjsi.J2513669

基于激光雷达与相机融合的 AGV 动态环境目标检测算法*

吴 斌[1],王世杰[1],卢 轶[1],饶 静[2],吴凌昊[3]
(1.南京林业大学机械电子工程学院 南京 210037; 2.北京航空航天大学仪器科学与光电工程学院 北京 100191;
3.中国航发四川燃气涡轮研究院 成都 610500)

摘 要:在人机混合智能仓库等动态环境中,AGV 通常难以精确感知随机出现的人和叉车等障碍物,为仓库的高效、安全运行带来了隐患,故提出一种基于激光雷达与图像融合的轻量化的目标检测方法(L-BEVFusion)。首先,为构建相机图像的鸟瞰视图(BEV)特征,设计了一个轻量化的特征提取网络用于获取图像的 2D 信息,通过引入多尺度语义特征平衡单一尺度语义特征带来的定位偏差;其次,采用基于显式监督方法采用深度真值对其进行监督,实现将图像特征投影到 3D 空间中;然后,分别提取图像和点云特征的预测信息,基于 BEV 特征融合网络,利用通道维度级联图像与点云的 BEV 特征,对其进行目标边界框的回归和分类预测,从而实现对人机混合仓库中动态障碍物的检测;最后,利用 KITTI 数据集和仓库实地采集数据对所提算法进行评估。实验结果表明,在实地采集的人机混合仓库数据集上,L-BEVFusion 方法与常见的点云图像融合方法相比较在工人类和叉车类的检测精度上分别提升了 3.46% 和 2.22%,综合平均检测精度高了 2.97%,在推理速度和检测尺寸精度上也表现更佳,其中法向距离平均误差为 4.02 mm,切向距离的平均绝对误差为 1.75 mm,提高了 AGV 检测的实时性和可靠性,保障了智能仓库物流的高效安全运转,具有较高的实际应用价值。

关键词: 智能仓储;深度预测;BEV 特征融合;目标检测
中图分类号: TH86 TP242 **文献标识码**: A **国家标准学科分类代码**: 510.80

Dynamic environment target detection algorithm for AGV based on lidar and camera fusion

Wu Bin[1], Wang Shijie[1], Lu Yi[1], Rao Jing[2], Wu Linghao[3]
(*1. College of Mechanical and Electronic Engineering, Nanjing Forestry University, Nanjing 210037, China*;
2. School of Instrumentation and Optoelectronic Engineering, Beihang University, Beijing 100191, China;
3. AECC Sichuan Gas Turbine Research Establishment, Chengdu 610500, China)

Abstract: In dynamic settings such as human-machine hybrid intelligent warehouses, Automated Guided Vehicles (AGVs) often face challenges in accurately detecting randomly appearing obstacles like pedestrians and forklifts, which can jeopardize both operational efficiency and safety. This study introduces a lightweight target detection method based on the fusion of LiDAR and image data, termed L-BEVFusion. Firstly, a lightweight feature extraction network is designed to derive 2D image information for constructing bird's eye view (BEV) features. To reduce localization errors caused by relying on single-scale semantic information, multi-scale semantic features are incorporated. Secondly, an explicit supervision strategy utilizing depth ground truth is applied to project image features into 3D space. Predictive features from both image and point cloud data are then extracted. A BEV feature fusion network concatenates these image and point cloud BEV features along the channel dimension, enabling bounding box regression and classification for dynamic obstacle detection in human-machine collaborative warehouses. The proposed algorithm is evaluated on both the KITTI dataset and real warehouse-collected data. Experimental results show that, compared with common point cloud-image fusion methods, L-BEVFusion improves detection accuracy for workers and forklifts by 3.46% and 2.22%, respectively, on the warehouse dataset, with an overall

收稿日期:2024-01-10 Received Date: 2024-01-10
*基金项目:江苏省现代农机装备与技术示范推广项目(NJ2023-16)、国家财政稳定支持项目(GJCZ-0202-2025-0004)资助

average accuracy increase of 2.97%. It also demonstrates superior inference speed and detection size accuracy, achieving an average normal distance error of 4.02 mm and a tangential absolute error of 1.75 mm. These improvements enhance the real-time detection performance and reliability of AGVs, ensuring efficient and safe logistics operations in intelligent warehouses and highlighting strong practical value.

Keywords: intelligent warehousing; depth prediction; BEV feature fusion; object detection

0 引 言

随着智能制造产业链的发展,对仓库进行智能化改造的需求日益增长。在仓库智能化改造中,最常见的技术方案是使用自动导向车辆(automated guided vehicle, AGV)作为货物的运输工具。然而,目前大部分仓库处于人机混合的动态工作环境中,场景中叉车与工人的位置随机,这使得 AGV 的环境感知面临巨大挑战。因此,如何在动态环境中实现 AGV 的精确感知成为亟待攻克的难题,这一技术问题也是保证 AGV 稳定行驶的关键。在环境感知中,基于深度学习的目标检测[1-3]是其中的关键任务之一,尤其在动态环境下,实时、准确地识别目标尤为重要。

目前,尽管基于 2D 图像的目标检测在仓储场景中已取得一定进展[4-6],但由于缺乏深度信息,无法直接获取目标物体的三维数据,导致基于二维数据进行三维空间中目标位置的检测变得较为困难[7]。在动态的仓储环境中,物体的空间位置对于 AGV 的稳定行驶至关重要,特别是在有动态障碍物(如叉车和工人)存在时,精准的空间感知显得尤为重要。因此,如何实时、准确地感知目标的三维位置信息,成为 AGV 稳定运行的关键。目前,基于点云数据的目标检测技术是 3D 目标检测领域的前沿方法[8],但是点云数据具有无序性和稀疏性,导致特征提取困难,对硬件设备要求较高。因此,依靠单一传感器数据难以完成复杂的感知任务,必须依赖多传感器融合技术[9-13],来应对复杂的动态人机混合仓库环境中的目标感知任务。激光雷达通过点云数据提供目标形状和深度等低分辨率信息,而相机则可以提供图像和纹理的高分辨率数据[14],将图像与点云数据融合方法用于 AGV 对于动态环境的感知,不仅能够弥补单一传感器的不足,还能充分利用两者的优势,显著提升 AGV 感知系统在动态场景中的检测精度和鲁棒性[15]。

激光雷达点云数据与相机图像信息的融合技术,包括像素级、决策级与特征级 3 种主要方法[16]。像素级的融合方法的数据信息损失最少,但需要处理大量的数据,且在动态环境的物体检测中难以实时处理快速变化的环境。决策级融合方法结构较为简单,处理速度较快,但由于融合发生在检测结果阶段,原始数据的细节会有较大损失,且在动态环境下,容易受传感器精度和数据对齐问题的影响,导致检测精度下降。上述这两种方法在动态环境检测中均存在一定的不足,尤其是在应对快速变化的动态环境和复杂障碍物时,实时性和精度难以兼顾。因此,针对动态环境中的障碍物检测,还需要更加高效且精确的信息融合策略,以提高在复杂环境中的感知能力和鲁棒性。特征层融合目前应用较为广泛,融合来自不同特征提取方法或者不同表示方式的多个特征,生成更具有代表性和丰富信息的特征向量,提高模型的性能和效果[14]。例如 MVX-Net 方法[17]分别对点云和图像进行特征提取,找到体素与图像的对应关系进行特征融合;PointPainting 方法[18]首先用分割网络对图像执行语义分割任务,预测出分割结果后,将其映射至相应的点云数据中进行三维目标检测。特征级融合方法因数据维度存在差异,需采用复杂的后处理逻辑,使得网络架构变得复杂且易导致信息损耗,影响感知的准确性[19]。且在将二维图像信息投射至三维点云的过程中,二维与三维网络由于相互独立容易造成融合不充分等现象[20]。

随着仓储智能化水平的不断提升,AGV 上安装的传感器类型和数量不断增加,如何融合 AGV 上的多个传感器数据,增强 AGV 对动态环境的感知力与提高其运行效率和决策精度成为仓储领域的研究热点。鸟瞰视图(bird's-eye-view, BEV)[21-22]将多传感器信号统一呈现在俯视视角中,提供了环境的全局视图,可以清楚地呈现 AGV 周围物体的位置和规模,有助于更好的理解周围环境,是当前最为先进的数据融合方法之一[23-24]。然而,现有的基于 BEV 视图的特征融合方法一般先对激光雷达点云进行目标检测,然后将其投影到图像中,未充分挖掘图像在三维目标检测中的潜力[25]。此外,AGV 在实际运行过程中,由于相机镜头污染或震动,可能导致外参矩阵偏差,从而影响数据融合的准确性,进而影响检测结果的可靠性。

因此,亟需一个简单高效的框架,能够充分融合图像和激光点云信息,以提高 AGV 对人机混合仓库等动态环境的感知精度与鲁棒性。提出基于激光雷达与图像融合的轻量化的目标检测方法(lightweight bird's-eye view Fusion, L-BEVFusion),设计了一种轻量化的目标检测方法,在 BEV 特征层面融合激光雷达点云与图像数据。通过引入多尺度语义特征,适应工人与叉车尺寸变化,平衡 AGV 的定位偏差;同时,搭建深度预测网络,通过显式监督利用深度真值训练,完成图像特征的三维投影与 BEV

映射;采用并行网络处理图像与点云数据,确保特征高效融合,提升算法的检测精度与效率。

1　L-BEVFusion 算法实施过程

L-BEVFusion 融合算法采用双分支结构独立处理图像和雷达点云数据。首先,为将图像特征转化到 BEV 特征空间,设计了一个多尺度特征提取网络获取图像的特征,适应人机混合仓库等动态环境下障碍物尺寸变化对图像特征的需求,通过对多尺度特征提取网络进行轻量化改进,提升网络的处理速度;其次,利用深度预测网络实现图像特征向三维空间投影;然后,从图像和点云中分别提取特征信息,将其映射至 BEV 特征空间,并借助 BEV 特征融合网络,在通道维度上级联图像与点云的 BEV 特征;最后,对融合后的特征进行目标边界框的回归和分类预测,得到在人机混合仓库场景中对叉车和工人的检测结果。

L-BEVFusion 通过在 BEV 空间内统一处理激光雷达和图像数据,显著提升对随机障碍物(如行人、叉车等)的检测性能,适用于人机混合仓库这类动态、多变的环境。算法在这些场景中的应用能够有效地识别并追踪目标,在人机协作、自动化搬运等任务中提供实时、高效的支持,提升 AGV 系统的智能化水平与安全性能。算法整体结构如图 1 所示。

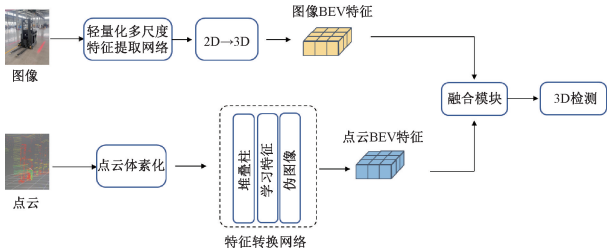

图 1　L-BEVFusion 融合算法整体结构

Fig. 1　Overall structure of L-BEVFusion

2　相机数据分支

构建 BEV 特征的方法大致分为两类,一类是 LSS(lift, splat, shoot)模型[26],该方法依赖图像的隐含深度信息,可能导致相似的图像特征在投影空间的多个位置重复出现,进而影响 BEV 特征的准确性;另一类是 BEVFormer 模型[27],在 BEV 空间下生成三维像素坐标,再将三维坐标投影到图像坐标中,但 BEVFormer 生成的 BEV 特征可能与真实 BEV 特征存在偏差,影响检测精度。L-BEVFusion 采用 BEVFormer 模型策略,将图像的视觉特征投射到相应的三维深度范围中,并引入 Ground Truth BEV 模块[28]来对齐生成的 BEV 特征与真值 BEV 特征,使用深度真值来监督深度信息,相机图像特征转为 BEV 特征过程如图 2 所示。

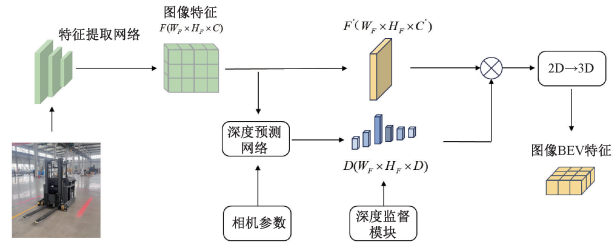

图 2　图像转化为 BEV 特征

Fig. 2　Transformation of image to BEV features

总体而言,2D 图像转换为 BEV 特征的过程包括:首先生成视锥特征,然后从视锥特征转换为体素特征,最后生成 BEV 特征。通过视锥特征提取网络获取图像视锥特征后,使用 BEVFormer 将图像的视觉特征投射到相应的三维深度范围中,在 BEV 坐标系下构建规则的 3D 体素网格,并利用相机内外参数进行投影变换,以获取对应的 2D 图像特征,L-BEVFusion 采用 BEVFormer 结构,首先在 BEV 空间构建 3D 体素网格,并通过投影回 2D 图像坐标获取相应的视觉特征,从而避免 LSS 由于隐含深度信息引起的特征模糊问题。同时,引入 Ground Truth BEV 模块,以真实 BEV 特征进行监督,提高 BEV 特征的匹配精度。

2.1　特征提取网络

将图像信息投影到三维空间前需要将图像特征与估计的深度关联起来。由于人机混合仓库环境复杂,包含多种目标物体,难以对目标进行准确的识别与精准定位。为了解决这个问题,设计了一个轻量化的多尺度特征提取网络获取图像的特征。特征提取网络包含 3 个部分,输入图像数据依次通过输入层 Entry Flow、中间层 Middle Flow 与输出层 Exit Flow 后得到最终输出多尺度图像特征,通过引入多尺度图像特征,适应人机混合仓库等动态环境下障碍物尺寸变化对图像特征的需要。图像特征提取网络如图 3 所示。

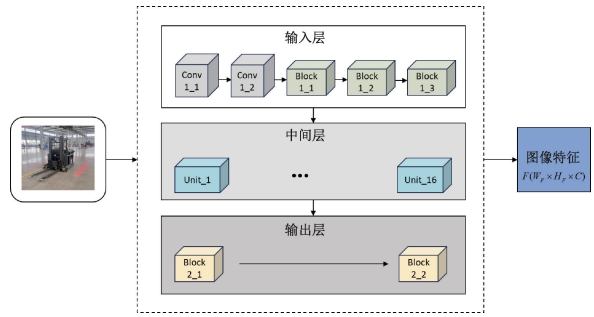

图 3　轻量化的多尺度特征提取网络

Fig. 3　Lightweight multi-scale feature extraction network

输入层包含两层普通卷积与 3 个 block 结构,每一个 block 共有 3 个深度可分离卷积,深度可分离卷积在保持网络深度的同时,显著降低了网络参数量,可以提高参数的使用效率,减少冗余的网络结构参数,加快网络的处理速度。在每个深度可分离卷积后加入了批量归一化处理与 H-Swish 残差结构,批量归一化处理首先对同一批次数据进行处理,然后应用比例系数和比例偏移对数据进行修正,使网络的输出更加稳定,减小参数对后续网络的影响。深度可分离卷积过程如图 4 所示。

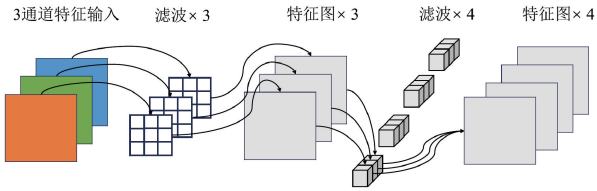

图 4 深度可分离卷积过程

Fig. 4 Depthwise separable convolution

采用深度离散化方法对图像特征进行像素级的估计,得到绝对深度分布,深度离散将连续的深度空间值离散化,使分类任务更加简化。常用深度离散化方法有等距离散化[29]、自适应离散化[30]与线性增加的离散化方法(linear-increasing discretization, LID)[31]。等距离散化对远处目标的深度信息分辨率较低,对远距离目标的深度误差较大,而 LID 采用对数间隔划分,能够在近距离保持较高分辨率,同时逐步增加远距离的间隔,使得整个深度范围内的信息分布更加合理,深度估计更加准确,提高了远距离目标的检测效果。自适应离散化方法通常需要额外的优化步骤,计算复杂度较高,而 LID 方法通过固定的对数间隔划分,既能提升远距离目标的检测精度,又能降低计算开销,适用于实时检测任务。

使用 LID 方法对图像特征进行深度离散,LID 提供了在所有深度上的上的均衡深度估计。L-BEVFusion 融合算法将前景深度限制在 $[d_{\min}, d_{\max}]$ 范围内,LID 离散公式为:

$$d_c = d_{\min} + \frac{d_{\max} - d_{\min}}{D(D+1)} \cdot d_i(d_i + 1) \tag{1}$$

其中,d_c 是连续的深度值,D 是深度区间的数量,d_i 是深度区间的索引。

深度离散化后需要深度分布标签来监督预测的深度分布。将雷达点云投影到图像层生成深度分布标签,通过深度补全生成每个像素处的深度值,通过下采样将尺寸为 $W_1 \times H_1$ 的深度图转换为尺寸为 $W_F \times H_F$ 的图像特征,其中 W 和 H 是特征的宽和高,使用 LID 深度离散方法生成深度分布标签 $\boldsymbol{D} \in \boldsymbol{R}^{W_F \times H_F \times D}$。

原始图像 $\boldsymbol{I} \in \boldsymbol{R}^{W_1 \times H_1 \times 3}$,经过特征提取网络与深度预测,将预测出来的深度块 $\boldsymbol{D}$ 和特征像素 $\boldsymbol{F}$ 做外积得到了带有深度信息的特征图。

$$\boldsymbol{G}(u,v) = \boldsymbol{D}(u,v) \otimes \boldsymbol{F}(u,v) \tag{2}$$

图像特征转化为视锥特征的数据流程如图 5 所示。

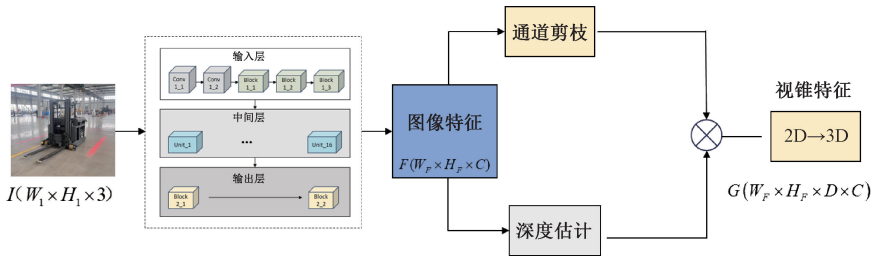

图 5 视锥特征提取网络

Fig. 5 Frustum feature network

2.2 图像 BEV 特征生成

得到视锥特征后需要将特征转换成体素的形式,图像的 BEV 特征与激光雷达的 BEV 特征需要构建到同一个坐标系。借助点云的坐标 $(X_{\min}, Y_{\min}, Z_{\min}, X_{\max}, Y_{\max}, Z_{\max})$ 与体素网格坐标 $V=(x,y,z)$ 来建立体素网格到激光雷达点云的投影矩阵 $\boldsymbol{P}_{VtoL}$,即:

$$\boldsymbol{P}_{VtoL} = \begin{pmatrix} x & 0 & 0 & X_{\min} \\ 0 & y & 0 & Y_{\min} \\ 0 & 0 & z & Z_{\min} \\ 0 & 0 & 0 & 1 \end{pmatrix}_{4\times4} \tag{3}$$

数据采集装置中雷达到相机的外参矩阵为 $\boldsymbol{P}_{LtoC}$。装置中相机内参矩阵为 $\boldsymbol{P}_{CtoI}$,校正矩阵为 $\boldsymbol{R}^0_{rect}$,内参矩阵为 $\boldsymbol{P}_{CtoI}$。

$$\boldsymbol{P}_{LtoC} = \begin{pmatrix} \boldsymbol{R}^{cam}_{LiDAR(3\times3)} & \boldsymbol{T}^{cam}_{LiDAR(3\times1)} \\ 0 & 1 \end{pmatrix}_{4\times4} \tag{4}$$

其中,$\boldsymbol{R}^{cam}_{LiDAR(3\times3)}$ 为旋转矩阵,$\boldsymbol{T}^{cam}_{LiDAR(3\times1)}$ 为平移矩阵。

$$\boldsymbol{R}^0_{rect} = \begin{pmatrix} \boldsymbol{R}^0_{rect(3\times3)} & 0 \\ 0 & 1 \end{pmatrix}_{4\times4} \tag{5}$$

$$\boldsymbol{P}_{CtoI}=\begin{pmatrix} f_u & 0 & c_u & -f_u b_x \\ 0 & f_v & c_v & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}_{3\times 4} \tag{6}$$

其中,f_u 和 f_v 表示输入图像的焦距横焦距和纵焦距,c_u 和 c_v 表示相机光学中心的横向与纵向坐标,b_x 表示相机坐标相对于参考坐标系的位移偏移。

相机在体素空间中的坐标为 $\boldsymbol{Coord}_V=(x_V,y_V,z_V)^{\mathrm{T}}$,雷达坐标系下点坐标为 $\boldsymbol{Coord}_L=(x_L,y_L,z_L)^{\mathrm{T}}$,相机坐标系下点坐标为 $\boldsymbol{Coord}_C=(x_C,y_C,z_C)^{\mathrm{T}}$,图像坐标系下点的像素坐标为$(u,v)^{\mathrm{T}}$,坐标投影计算方式为:

$$\begin{pmatrix}\boldsymbol{Coord}_L\\1\end{pmatrix}=\boldsymbol{P}_{VtoL}\cdot\begin{pmatrix}\boldsymbol{Coord}_V\\1\end{pmatrix} \tag{7}$$

$$\begin{pmatrix}\boldsymbol{Coord}_C\\1\end{pmatrix}=\boldsymbol{P}_{LtoC}\cdot\begin{pmatrix}\boldsymbol{Coord}_L\\1\end{pmatrix} \tag{8}$$

$$z_C\begin{pmatrix}\boldsymbol{Coord}_L\\1\end{pmatrix}=\boldsymbol{P}_{CtoI}\cdot\boldsymbol{R}_{rect}\cdot\begin{pmatrix}\boldsymbol{Coord}_C\\1\end{pmatrix} \tag{9}$$

联立后得到像素坐标到体素坐标的投影关系:

$$z_C\begin{pmatrix}\boldsymbol{Coord}_I\\1\end{pmatrix}=\boldsymbol{P}_{CtoI}\cdot\boldsymbol{R}_{rect}\cdot\boldsymbol{P}_{LtoC}\cdot\boldsymbol{P}_{VtoL}\begin{pmatrix}\boldsymbol{Coord}_V\\1\end{pmatrix} \tag{10}$$

将相机坐标点中的深度信息 z_C 转化为离散的深度区间 d_i,并与像素坐标点进行矩阵拼接。

$$\boldsymbol{Coord}_G=\boldsymbol{Coord}_I\oplus LID(z_C) \tag{11}$$

BEV 特征能够显著降低网络的计算量,同时在检测性能上与 3D 体素特征相当,沿着通道维度使用 1×1 卷积+ BatchNorm + ReLU 层减少通道数量,将通道数从 256 降到 64,在检索原始通道数 C 的同时,学习每个高度切片的相对重要性,最终输出具有调整后通道数的图像 BEV 特征 $\boldsymbol{B}_{Camera}\in\boldsymbol{R}^{X\times Y\times C}$。图像 BEV 特征生成过程如图 6 所示。

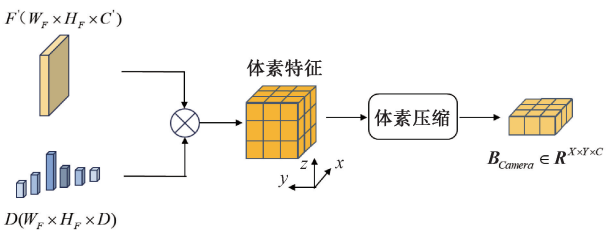

图 6　图像 BEV 特征生成结构

Fig. 6　Structure for generating image BEV features

3　构建点云 BEV 特征

点云的 BEV 特征由点云沿着 Z 轴方向合并折叠得到,人机混合仓库环境复杂,点云数据比较复杂且数据量庞大,直接进行合并折叠会增大计算量。PointPillars 简化了主网络结构,将输入激光雷达点云按 X 与 Y 轴划分为网格,落入到一个网格中的点云构成一个 Pillars,对点云的 D 个维度进行升维,得到 $\boldsymbol{V}_L'\in\boldsymbol{R}^{64\times P\times N}$,然后使用最大池化操作输出 $\boldsymbol{L}_M\in\boldsymbol{R}^{64\times P}$,使用深度索引生成点云 BEV 特征 $\boldsymbol{V}_L'\in\boldsymbol{R}^{64\times H\times W}$。以伪图像特征进行编码,避免了复杂的 3D 卷积,使得计算速度更快,更适用于本文提出的人机混合仓库中的点云数据,最终得到点云的 BEV 特征 $\boldsymbol{B}_{LiDAR}\in\boldsymbol{R}^{X\times Y\times C_{LiDAR}}$。障碍物点云映射到 BEV 特征空间的过程如图 7 所示。

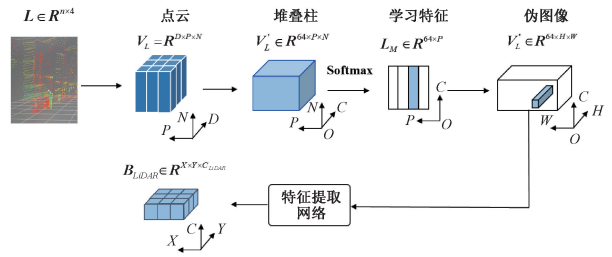

图 7　点云转化为 BEV 特征

Fig. 7　Conversion of point clouds into BEV features

4　图像与点云在 BEV 特征空间融合

为了有效地融合来自相机的 BEV 特征 $\boldsymbol{B}_{Camera}$ 和激光雷达传感器的 BEV 特征 $\boldsymbol{B}_{LiDAR}$,采用动态融合模块进行特征融合,特征融合模块如图 8 所示。

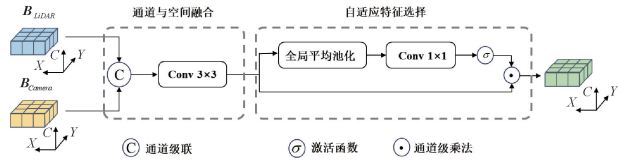

图 8　BEV 特征融合模块

Fig. 8　BEV feature fusion module

给定相同空间维度下的 2 个特征,可以采用静态权值将其融合,融合后的 BEV 特征可以公式化表述为:

$$\boldsymbol{B}_{fused}=f_{adaptive}(f_{static}([\boldsymbol{B}_{Camera},\boldsymbol{B}_{LiDAR}])) \tag{12}$$

其中 $[\boldsymbol{B}_{Camera},\boldsymbol{B}_{LiDAR}]$ 表示将相机和激光雷达传感器的 BEV 特征沿着通道维度进行特征拼接,f_{static} 表示静态的通道与空间融合函数,通过 3 × 3 的卷积层将拼接特征在通道维度进行降维操作,$f_{adaptive}$ 表示融合模块中的自适应特征,计算方法为:

$$f_{adaptive}(\boldsymbol{B})=\sigma(\boldsymbol{W}f_{avg}(\boldsymbol{B}))\cdot\boldsymbol{B} \tag{13}$$

其中,$\boldsymbol{W}$ 表示线性变换矩阵,f_{avg} 表示全局平均池化,σ 表示 sigmoid 函数。

本研究提出的 L-BEVFusion 目标检测网络将图像数据与点云数据结合,既保留了图像的颜色和细节信息,又

引入了点云的全局几何信息,弥补了单一传感器数据带来的局限性。

5 实验结果分析

5.1 数据集与实验环境

为了评价 L-BEVFusion 融合算法的性能,模型对比实验在公开的 KITTI 数据集上进行[32]。由于数据集中自行车类别的样本较少,因此仅使用 KITTI 数据集中的行人和汽车类别进行训练,网络模型在训练集上进行训练,并在验证集上进行算法验证。设置不同类别的交并比(IoU)阈值:汽车类的 IoU 阈值为 0.7,行人类的 IoU 阈值为 0.5。随后,分别计算各类别的检测精度和召回率,并使用所有类别的平均精度均值(mean average precision,mAP)和帧率作为算法的评估指标。

5.2 对比实验及分析

采用 Ubuntu16.04 和 Windows 双操作系统,在 Ubuntu16.04 系统下,利用 ROS 采集仓库实时点云数据,深度学习框架配置为 python3.10、pytorch1.10、cuda11.3。在 KITTI 数据集上,对 L-BEVFusion 模型与主流的三维检测模型进行了对比评估。评估指标采用帧率来衡量目标检测模型的推理速度[33],结果如表 1 所示。

表 1 KITTI 数据集下的检测结果

Table 1 Detection results on KITTI dataset

方法	模态	汽车 mAP /%	行人 mAP /%	帧率 /fps
Pointpillars	点云	74.02	53.76	31.5
F-PointNet	点云+图像	72.89	65.01	6.8
CLOCs	点云+图像	72.15	64.18	7.9
本研究	点云+图像	74.79	72.89	13.9

由表 1 检测结果可知,L-BEVFusion 在叉车和工人检测任务中各项评价指标均取得了显著提升。在检测精度方面,L-BEVFusion 相比于单模态方法 Pointpillars 有所提升,且优于其他点云图像融合方法,这表明改进的多模态特征融合策略增强了目标的几何与语义特征表达,在检测任务中具有更强的特征提取能力和目标识别能力。在检测速度方面,L-BEVFusion 在帧率上达到了 13.9 fps,表明其在提升精度的同时可以保持不错的推理速度。

5.3 实地实验结果及分析

为了进一步验证 L-BEVFusion 融合算法的检测性能,使用 AGV 在人机混合仓库中进行实地检测实验,采用大疆 Mid-360 形激光雷达和 320 万像素的工业相机分别作为点云数据和图像数据采集传感器。Mid-360 形激光雷达每秒钟输出点云点数可以达到 20 万个,点云帧率为 10 Hz,远探测距离可以达到 70 m,采用混合固态技术,可以帮助 AGV 感知人机混合仓库等复杂环境,满足人机混合仓库中 AGV 对叉车和工人的检测需求。随机截取实地采集数据,采集结果如图 9 所示。

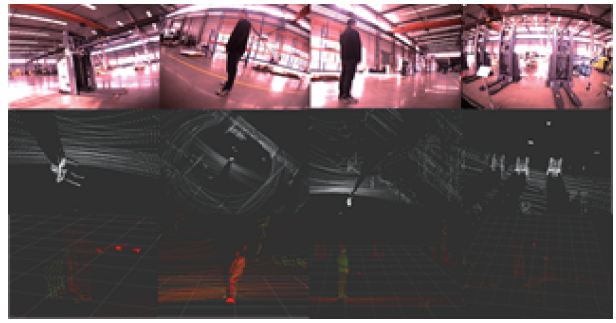

图 9 人机混合仓库实地数据

Fig. 9 Field data of human-machine hybrid warehouse

本研究提出的 L-BEVFusion 与 Pointpillars 在 KITTI 数据集中在处理速度和检测精度综合能力上有着优异的表现,因此将两个网络在仓库实地采集的数据集上进行比较,比较结果如表 2 所示。

表 2 仓库实地数据集下的检测结果

Table 2 Detection results on the warehouse field dataset (%)

方法	叉车	工人	mAP
Pointpillars	63.48	61.09	61.91
本研究	81.78	76.45	79.42

由检测结果可知,相比于检测速度较快的单模态检测模型 Pointpillars,L-BEVFusion 模型在叉车检测任务中检测精度提升了 18.3%;在工人任务中检测精度提升了 15.36%;总体检测精度提升了 17.51%。实验结果表明,L-BEVFusion 提高了 AGV 检测的实时性和可靠性。

为了验证 L-BEVFusion 融合算法的可靠性,将提出的融合算法与 Pointpillars 进行误差比较。分别计算预测包围框与针织包围框法向距离(沿着目标中心到传感器的方向)和切向距离的平均绝对误差(mean absolute error, MAE)以及包围框的高度误差,对比结果如表 3 所示。

表 3 误差结果对比

Table 3 Comparison of error results (mm)

方法	法向 MAE	切向 MAE	高度误差
Pointpillars	4.87	3.17	0.58
本研究	4.02	1.75	0.43

由误差对比结果可知,在法向 MAE、切向 MAE 和高度误差 3 个指标上,L-BEVFusion 表现均显著优于 Pointpillars。法向方向上 L-BEVFusion 的测量平均绝对误差为 4.02 mm,相比 Pointpillars 降低了约 17.5%。切向方向上平均绝对误差为 1.75 mm,相比 Pointpillars 降低了约 44.8%。在整体高度误差方面 L-BEVFusion 的平均绝对误差为 0.43 mm,相比 Pointpillars 降低了约 25.9%。实验结果表明,算法融合机制的改进使得各个方向的数据处理更加均衡,减少了单一方向上的误差积累。

为验证本研究方法的有效性,选择了实地人机混合仓库采集的数据进行可视化展示。结果如图 10 所示。

图 10　仓库实地数据集上 L-BEVFusion 可视化结果

Fig. 10　Visualization results of L-BEVFusion on warehouse field dataset

从可视化结果可以看出,在光线较弱且环境复杂的人机混合仓库中,L-BEVFusion 能够准确感知行驶过程中环境中出现的工人与叉车。因此,本研究提出的 L-BEVFusion 算法通过有效融合激光雷达和相机 2 种传感器数据,能够有效应对人机混合的动态仓库环境中的障碍物检测问题,具有较高的实际应用价值。

5.4　消融实验

为评估 BEV 特征空间融合与深度检测模块性能,在人机混合仓库实地环境获取的数据集上开展消融实验,实验结果对比如表 4 所示。其中,Baseline 表示纯点云检测网络基础框架;Fusion 表示图像与点云数据在 BEV 特征空间下进行融合;本研究为在构建图像 BEV 特征的过程中融入真值监督模块。

表 4　不同模块对 L-BEVFusion 检测精度的影响

Table 4　Impact of different modules on the detection accuracy of L-BEVFusion

模块类型	BEV	深度监督	叉车/%	工人/%	mAP/%
Baseline	×	×	65.42	63.29	64.71
Fusion	√	×	78.32	74.23	76.45
本研究	√	√	81.78	76.45	79.42

由表 4 可以看出,将图像信息融入 Baseline 中进行 BEV 特征融合后,叉车与工人的检测准确率均有所提高,其中叉车类和工人类检测精度分别得到提升了,叉车类检测精度提升较明显,这是因为叉车类别的点云数据较为稀缺,纯点云检测框架在处理叉车这一具有挑战性的类别时表现欠佳,因此,叉车目标的检测更加依赖于图像信息的辅助。在加入了真值监督模块后,基于 BEVFormer 的 BEV 特征构建方法更适应于人机混合动态仓库等动态场景,显著增强了 AGV 对动态环境的感知和检测精度。具体来说,叉车类别和工人类别的检测精度分别提升了 3.46% 和 2.22%,综合平均准确率提高了 2.97%。这些改进显著提升了网络的检测效能,证实了图像特征和深度预测模块在动态复杂环境下对于提高网络目标检测能力的有效性。

6　结　　论

针对在人机混合仓库等动态环境中,由于障碍物状态不确定导致的 AGV 难以精准感知问题,本研究提出了一种轻量化目标检测算法 L-BEVFusion,在 BEV 特征空间融合数据。为适应目标物体尺寸的多样性,设计了一个多尺度特征提取网络获取图像特征,为满足 AGV 感知实时性的要求,对多尺度特征提取网络进行轻量化改进,提升网络处理速度;同时,使用深度信息真值来直接监督其预测的深度信息,预估图像的深度信息;设计 BEV 特征融合网络,按照通道维度级联图像与点云的 BEV 特征,充分利用激光雷达与相机数据的互补性,从而提升检测精度与效率。

实验结果表明,L-BEVFusion 在公开的 KITTI 数据集上取得了较大的性能提升,在自建人机混合仓库环境数据集上对叉车类、工人类和整体的的检测精度达到了 81.78%、76.45% 与 79.42%,此外,本研究提出的融合算法相较于经典的融合算法在推理速度和尺寸精度上表现更佳,符合人机混合仓库中对于检测实时性和可靠性的需求。在和工人和叉车的检测精度上分别提升了 3.46% 和 2.22%,在推理速度和检测尺寸精度上也表现更佳,其中法向距离平均误差达到 4.02 mm,切向距离的平均绝对误差为 1.75 mm,符合人机混合仓库中 AGV 检测实时性、可靠性的需求,具有较好的实际应用价值。

目前 L-BEVFusion 在现有人机混合仓库中取得了良好表现,但是仍有进一步提升的空间。在未来的研究中,可以通过改进特征提取网络,提升 AGV 对小目标与远距离目标的检测能力,进一步增强网络对动态环境的适应性。还可以将研究拓展应用到更广泛的动态环境,并通过引入多模态融合技术进一步提升 AGV 的环境感知能力。

参考文献

[1]　FELZENSZWALB P F, GIRSHICK R B, MCALLESTER D, et al. Object detection with discriminatively trained

part-based models[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(9): 1627-1645.

[2] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature, 2015, 521(7553): 436-444.

[3] QI C R, SU H, MO K, et al. Pointnet: Deep learning on point sets for 3d classification and segmentation[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition, 2017:77-85.

[4] 武文汉, 杨明, 王冰, 等. 一种基于轮廓匹配的仓储机器人托盘检测方法[J]. 上海交通大学学报, 2019, 53(2): 197-202.
WU W H, YANG M, WANG B, et al. A contour matching-based method for pallet detection of warehouse robots[J]. Journal of Shanghai Jiao Tong University, 2019, 53(2): 197-202.

[5] 王晨, 袁庆霓, 白欢, 等. 面向仓储货物的轻量化目标检测算法[J]. 激光与光电子学进展, 2022, 59(24): 74-80.
WANG CH, YUAN Q N, BAI H, et al. A lightweight target detection algorithm for warehouse goods[J]. Progress in Optics and Laser, 2022, 59(24): 74-80.

[6] 周诗捷, 王玉槐, 沈思橙, 等. 基于改进型 Faster R-CNN 的仓储环境物体识别技术研究[J]. 计算技术与自动化, 2024, 43(2): 187-191.
ZHOU SH J, WANG Y H, SHEN S CH, et al. Research on object recognition technology in warehouse environments based on improved Faster R-CNN[J]. Computer Technology and Automation, 2024, 43(2): 187-191.

[7] FERNANDES D, SILVA A, NÉVOA R, et al. Point-cloud based 3D object detection and classification methods for self-driving applications: A survey and taxonomy[J]. Information Fusion, 2021, 68: 161-191.

[8] 李思纯, 王建军, 宋伟润, 等. 无人驾驶扫地机道路可行驶区域的融合提取研究[J]. 仪器仪表学报, 2024,45(12):190-200.
LI S CH, WANG J J, SONG W R, et al. Fusion extraction of drivable areas for unmanned sweeping vehicles[J]. Chinese Journal of Scientific Instrument, 2024, 45(12): 190-200.

[9] 任泽裕, 王振超, 柯尊旺, 等. 多模态数据融合综述[J]. 计算机工程与应用, 2021, 57(18): 49-64.
REN Z Y, WANG ZH CH, KE Z W, et al. A review of multimodal data fusion[J]. Computer Engineering and Applications, 2021, 57(18): 49-64.

[10] 刘永刚, 于丰宁, 章新杰, 等. 基于激光点云与图像融合的 3D 目标检测研究[J]. 机械工程学报, 2022, 58(24): 289-299.
LIU Y G, YU F N, ZHANG X J, et al. Research on 3D object detection based on fusion of LiDAR point clouds and images[J]. Journal of Mechanical Engineering, 2022, 58(24): 289-299.

[11] 洪诚康, 杨力, 江文松, 等. 基于多深度相机融合的机械臂抓取系统[J]. 计算机集成制造系统, 2024, 30(2): 435-444.
HONG CH K, YANG L, JIANG W S, et al. A robotic arm grasping system based on multi-depth camera fusion[J]. Computer Integrated Manufacturing Systems, 2024, 30(2): 435-444.

[12] 苏增辉, 马向宇, 白静, 等. 基于层次掩码及多尺度特征融合的 CAD 模型表征[J/OL]. 计算机集成制造系统,1-23[2025-04-08].
SU Z H, MA X Y, BAI J, et al. CAD model representation based on hierarchical masks and multi-scale feature fusion[J/OL]. Computer Integrated Manufacturing Systems, 1-23[2025-04-08].

[13] 汪鑫, 廖小平, 刘树胜, 等. 多传感器融合下多工况刀具磨损状态预测的深度森林方法研究[J]. 仪器仪表学报, 2023, 44(9): 265-274.
WANG X, LIAO X P, LIU SH SH, et al. Deep forest-based method for multi-condition tool wear state prediction under multi-sensor fusion[J]. Chinese Journal of Scientific Instrument, 2023, 44(9): 265-274.

[14] CUI Y D, CHEN R, CHU W B, et al. Deep learning for image and point cloud fusion in autonomous driving: A review[J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 23(2): 722-739.

[15] FANG D J, HAN J L, SHAO J J, et al. Pallet detection and localization based on RGB image and point cloud data for automated forklift[C]. 2024 7th International Conference on Advanced Algorithms and Control Engineering, 2024:642-647.

[16] 司垒, 谭超, 朱嘉皓, 等. 基于 X 射线图像和激光点云的煤矸识别方法[J]. 仪器仪表学报, 2022, 43(9): 193-205.
SI L, TAN CH, ZHU J H, et al. Coal gangue recognition method based on X-ray images and laser point clouds[J].

Chinese Journal of Scientific Instrument, 2022, 43(9): 193-205.

[17] SANDLER M, HOWARD A, ZHU M, et al. Mobilenetv2: Inverted residuals and linear bottlenecks[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018:4510-4520.

[18] VORA S, LANG A H, HELOU B, et al. Pointpainting: Sequential fusion for 3D object detection[C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020:4603-4611.

[19] 康国华, 张琪, 张晗, 等. 基于点云中心的激光雷达与相机联合标定方法研究[J]. 仪器仪表学报, 2019, 40(12): 118-126.
KANG G H, ZHANG Q, ZHANG H, et al. Joint calibration method of LiDAR and camera based on point cloud centroid[J]. Chinese Journal of Scientific Instrument, 2019, 40(12): 118-126.

[20] 李文礼, 喻飞, 石晓辉, 等. BEV 特征下激光雷达和单目相机融合的目标检测算法研究[J]. 计算机工程与应用, 2024, 60(11): 182-193.
LI W L, YU F, SHI X H, et al. Research on target detection algorithm based on LiDAR and monocular camera fusion under BEV features[J]. Computer Engineering and Applications, 2024, 60(11): 182-193.

[21] LI H Y, SIMA CH H, DAI J F, et al. Delving into the devils of bird′s-eye-view perception: A review, evaluation and recipe[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 46(4): 2151-2170.

[22] MA Y X, WANG T, BAI X Y, et al. Vision-centric bev perception: A survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024, 46(12): 10978-10997.

[23] 刘明杰, 何峥言, 陈俊生, 等. 基于循环跨视图转换和多状态特征融合的鸟瞰图生成方法[J]. 仪器仪表学报, 2024, 45(10): 133-142.
LIU M J, HE ZH Y, CHEN J SH, et al. Bird′s-eye view generation method based on cyclic cross-view transformation and multi-state feature fusion[J]. Chinese Journal of Scientific Instrument, 2024, 45(10): 133-142.

[24] 黄德启, 黄海峰, 黄德意, 等. BEV 感知学习在自动驾驶中的应用综述[J]. 计算机工程与应用, 2025, 61(6):1-21.
HUANG D Q, HUANG H F, HUANG D Y, et al. A review of BEV perception learning in autonomous driving[J]. Computer Engineering and Applications, 2025,61(6):1-21.

[25] SRIVASTAVA S, JURIE F, SHARMA G. Learning 2D to 3D lifting for object detection in 3D for autonomous vehicles[C]. 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2019: 3657-3664.

[26] PHILION J, FIDLER S. Lift, splat, shoot: Encoding images from arbitrary camera rigs by implicitly unprojecting to 3D[C]. Computer Vision-ECCV 2020, 2020: 194-210.

[27] LI ZH Q, WANG W H, LI H Y, et al. Bevformer: Learning bird′s-eye-view representation from multi-camera images via spatiotemporal transformers[C]. Computer Vision-ECCV 2022, 2022:1-18.

[28] PAN CH B, YAMAN B, VELIPASALAR S, et al. Clip-bevformer: Enhancing multi-view image-based bev detector with ground truth flow[C]. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024:15216-15225.

[29] 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, 2018:4490-4499.

[30] LIU ZH J, TANG H T, AMINI A, et al. Bevfusion: Multi-task multi-sensor fusion with unified bird′s-eye view representation[C]. 2023 IEEE International Conference on Robotics and Automation, 2023: 2774-2781.

[31] READING C, HARAKEH A, CHAE J, et al. Categorical depth distribution network for monocular 3D object detection[C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 8555-8564.

[32] LIAO Y Y, XIE J, GEIGER A. Kitti-360: A novel dataset and benchmarks for urban scene understanding in 2D and 3D[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(3): 3292-3310.

[33] 董钰婷, 官磊. 基于自适应加权融合激光雷达和相机的三维目标检测方法[J]. 计算机应用, 2024, 44(S1): 250-255.
DONG Y T, GUAN L. A 3D object detection method based on adaptive weighted fusion of LiDAR and

camera[J]. Computer Applications, 2024, 44(S1): 250-255.

作者简介

吴斌,2003 年于景德镇陶瓷学院获得学士学位,2006 年于东南大学获得硕士学位,2011 年于东南大学获得博士学位,现为南京林业大学副教授,主要研究方向为数字化设计与制造。

E-mail:wubin@njfu.edu.cn

Wu Bin received his B. Sc. degree from Jingdezhen Ceramic Institute in 2003, received his M. Sc. degree from Southeast University in 2006, and received his Ph. D. degree from Southeast University in 2011. He is currently an associate professor at Nanjing Forestry University. His research interests include digital design and manufacturing.

卢轶(通信作者),2014 年于江苏大学获得硕士学位,2020 年于东南大学获得博士学位,现为南京林业大学讲师。主要研究方向为机器视觉、深度学习、制造过程的智能检测和数字孪生等方向。

E-mail:juliusx@163.com

Lu Yi (Corresponding author) received his M. Sc. degree from Jiangsu University in 2014 and his Ph. D. degree from Southeast University in 2020. He is currently a lecturer at Nanjing Forestry University. His research interests include machine vision, deep learning, intelligent inspection in manufacturing processes, and digital twin.

饶静,2009 年于山东大学获得学士学位,2012 年于浙江大学获得硕士学位,2018 年于南洋理工大学获得博士学位,现为北京航空航天大学教授,主要研究方向为无损检测,柔性传感器和基于人工智能的缺陷检测。

E-mail:jingrao@buaa.edu.cn

Rao Jing received her B. Sc. degree from Shandong University in 2009, received her M. Sc. degree from Zhejiang University in 2012, and received her Ph. D. degree from Nanyang Technological University in 2018. She is currently a professor at Beihang University. Her research interests include nondestructive testing, flexible sensors, and defect detection based on artificial intelligence.

吴凌昊,2014 年于西北工业大学获得学士学位,2017 年于西北工业大学获得硕士学位,现为中国航发四川燃气涡轮研究院工程师,主要研究方向为航空发动机光学测试。

E-mail:wlhgte@163.com

Wu Linghao received his B. Sc. degree from Northwestern Polytechnical University in 2014 and his M. Sc. degree from the same university in 2017. He is currently an engineer at the Sichuan Gas Turbine Research Institute of China Aviation Engine Corporation (AECC). His research interest includes optical testing for aircraft engines.