

## 基于置信估计网络和改进D-S的多模态融合方法

程腾<sup>1,2,3</sup>, 郭利港<sup>1,2,3</sup>, 张强<sup>1,2,3,4</sup>, 王文冲<sup>4</sup>, 石琴<sup>1,2,3</sup>, 侯登超<sup>1,2,3</sup>

(1. 合肥工业大学 自动驾驶汽车安全技术安徽省重点实验室, 合肥 230009;  
2. 安徽省智慧交通车路协同工程研究中心, 合肥 230009; 3. 合肥工业大学 汽车与交通工程学院, 合肥 230009;  
4. 奇瑞汽车股份有限公司, 安徽, 芜湖 241007)

**摘要:** 针对神经网络可解释性较低, 且多模态融合中D-S理论易在高冲突场景出现悖论的问题, 提出了一种基于置信估计网络和改进D-S理论的结果级多模态融合方法, 该方法包括一种置信估计网络, 将目标检测中的分类问题表述为置信估计问题, 输出单模态网络检测结果的置信估计分数; 一种改进D-S理论的融合方法, 用检测结果的置信度分数和类别信息合成证据, 最终融合不同模态检测数据。在KITTI数据集上进行的评估试验表明, 该融合方法的mAP值较图像单模态检测最高提升6.64%, 较点云单模态检测最高提升15.43%, 对比结果表明, 该融合方法的mAP提升值高于经典D-S融合0.81%, 能有效减少分类冲突, 克服经典D-S的局限性。

**关键词:** 置信估计; 环境感知; 多模态目标检测; D-S证据理论; 结果级融合

中图分类号: TP391.4 文献标志码: A DOI: 10.3969/j.issn.2095-1469.2025.02.02

## Multimodal Fusion Method Based on Confidence Estimation Network and Improved D-S Theory

CHENG Teng<sup>1,2,3</sup>, GUO Ligang<sup>1,2,3</sup>, ZHANG Qiang<sup>1,2,3,4</sup>, WANG Wenchong<sup>4</sup>,  
SHI Qin<sup>1,2,3</sup>, HOU Dengchao<sup>1,2,3</sup>

(1. Key Laboratory for Automated Vehicle Safety Technology of Anhui Province, Hefei University of Technology, Hefei 230009, China; 2. Engineering Research Center for Intelligent Transportation and Cooperative Vehicle-Infrastructure of Anhui Province, Hefei 230009, China;  
3. School of Automotive and Transportation Engineering, Hefei University of Technology, Hefei 230009, China;  
4. Chery Automobile Co., Ltd., Wuhu 241007, Anhui, China)

**Abstract:** Neural networks lack interpretability and the D-S theory is prone to paradoxes in high-conflict scenarios of multimodal fusion. In response, this paper proposes a result-level multimodal fusion method that integrates a confidence estimation network with an improved D-S theory. The method consist of two key

收稿日期: 2023-12-02 改稿日期: 2024-02-21 网络首发日期: 2024-06-06

基金项目: 国家自然科学基金项目(82171012); 安徽省自然科学基金资助项目(2208085MF171);  
安徽省新能源汽车暨智能网联汽车创新工程项目(JZ2021AFKJ0002);  
汽车标准化公益性开放课题资助(CATARC-Z-2022-01350)

参考文献引用格式:

程腾, 郭利港, 张强, 等. 基于置信估计网络和改进D-S的多模态融合方法[J]. 汽车工程学报, 2025, 15(2): 137-146.

CHENG Teng, GUO Ligang, ZHANG Qiang, et al. Multimodal Fusion Method Based on Confidence Estimation Network and Improved D-S Theory[J]. Chinese Journal of Automotive Engineering, 2025, 15(2): 137-146. (in Chinese)



components. First, a confidence estimation network reframes the classification problem in target detection as a confidence estimation task, providing confidence scores for the detection results of individual unimodal networks. Second, a fusion method with improved D-S theory uses confidence scores and class information to construct evidence, achieving final fusion of detection data from different modalities. Evaluation experiments on the KITTI dataset show that the proposed fusion method improves mAP by up to 6.64% compared to image-based detection and up to 15.43% compared to point cloud-based detection. In the comparison of fusion methods, the proposed fusion method achieves an mAP improvement 0.81% higher than the classical D-S fusion. It effectively reduces classification conflicts and addresses the limitations of the classical D-S theory.

**Keywords:** confidence estimate; environment perception; multimodal target detection; dempster-shafer evidence theory; result-level multimode fusion

自动驾驶技术<sup>[1]</sup>、多模态协同感知已成为当前的研究热点之一。多模态融合方法根据融合层次可分为数据级融合、特征级融合和结果级融合。数据级和特征级融合在充分利用不同模态的信息方面有较大优势。LI Yingwei等<sup>[2]</sup>提出一种深度特征融合方法,用于提高激光雷达和摄像头数据的对齐质量。该方法应用反向数据增强技术,将点云中的关键点与摄像头像素数据对齐,一定程度上解决了旋转变换导致数据发生扭曲时的对齐问题。应用基于学习的特征配准技术,通过机器学习优化摄像头和激光雷达数据的深度特征对齐,综合了不同传感器模态数据的特点。WU Xiaopei等<sup>[3]</sup>提出焦点稀疏卷积及其多模态版本,考虑了空间中数据特征分布不均匀的特点。该模块通过学习空间特征稀疏程度,找到数据分布的关键位置。在卷积过程中对不同重要程度的数据使用不同的卷积方法,在学习特征空间分布的同时提取多模态信息。

但数据级与特征级融合通常对不同模态数据对齐有较高要求,并且由于传感器的原始数据量较大、网络模型架构较复杂,所以融合难度较高。结果级融合对单模态检测器的识别结果进行融合,是更高层次的数据融合,融合时能有效避免信息的损失,并且由于融合时计算量较少,所以在实时性及模型复杂度方面有更卓越的优势。PANG Su等<sup>[4]</sup>考虑常规特征级融合方法中粗暴地将不同稠密度模态对应的不同维度特征向量变换到相同大小,不能很好地融合多模态特征,因此,提出CLOCs融合方法,在多模态对应的神经网络完成预测进行非极

大值抑制处理前,将预测的候选框作为结果级融合模块的输入,基于几何一致性与语义一致性的约束,使用卷积神经网络提取输入的特征,并预测出融合结果。此外,为降低模型复杂性,PANG Su等<sup>[5]</sup>在CLOCs融合方法基础上提出Fast CLOCs方法,将3D检测候选框投影至图像坐标系下,并基于投影的候选框作为图像识别分支的建议区域,用于修正2D检测结果,最后利用CLOCs方法融合2D检测结果与3D检测结果,该方法在检测精度与实时性方面均有较好的表现。

现有基于计算机视觉和深度学习的目标检测方法在特定数据集上取得了优异的成绩,但在现实场景中的部署仍然面临限制<sup>[6]</sup>,主要原因在于神经网络自身的不确定性,无法提供对其预测结果的可靠置信估计。ZHAO Xujiang等<sup>[7]</sup>针对传统证据神经网络没有考虑不同数据本身不确定性的问题,提出正则化证据神经网络。使用迪利克雷分布函数替换传统神经网络中的softmax函数,使模型输出具有置信度以及不确定性的结果。通过在合成数据集与真实世界数据集上的训练,使模型能分辨不同训练样本来源,得到分类任务中的不确定度。HAN Zongbo等<sup>[8]</sup>认为在多分类任务中,传统神经网络使用softmax函数输出类别预测的置信度会导致神经网络对检测结果过度自信。使用ReLU激活函数对神经网络输出进行激活处理,获得非负神经网络,然后使用迪利克雷分布获得目标分类置信度以及结果不确定度。最终在多源数据任务中使用D-S证据理论融合多源检测结果。CORBIÈRE等<sup>[9]</sup>为

可靠评估神经网络预测结果的置信度，以及有效预测模型的错误分类，提出了真实类别概率 (True Class Probability, TCP) 作为衡量模型置信度的指标。通过构建置信度神经网络，结合新的学习方式让模型在训练时学习真实类别概率。DEVRIES等<sup>[10]</sup>为增强神经网络对域外数据的检测能力，提出一种使神经网络学习自身预测结果置信度的方法。在训练分类模型的同时训练一个置信估计模型，对于一个给定输入，该网络输出分类结果以及对应于该分类结果的置信估计，通过分别构造二者的损失函数对网络进行训练。通过对网络的置信估计输出大小判断检测样本是否为域外数据，达到在普通数据集下训练域外检测任务的目的。

D-S 证据理论<sup>[11]</sup>是一种处理不确定性、不完整性以及不一致性的数学框架。黄建根等<sup>[12]</sup>通过对车辆的不同传感器数据构造概率分配函数来形成证据，然后使用D-S理论预测车辆的运动轨迹，提高预测的准确性。传统D-S证据理论合成规则在处理分配冲突时存在“一票否决”缺陷，针对该缺陷，目前已有的改进方法可以分为两类<sup>[13]</sup>：一部分学者认为D-S证据理论的缺陷源自其组合规则，因此，通过研究组合规则的改进解决D-S证据理论的缺陷，比如Yager方法、Dubois方法和Semts方法等；另一部分学者认为D-S证据理论本身并不存在缺陷，高度冲突的本质在于证据分配了极端的概率密度值，因此，通过加权等形式修正证据源，比如Murphy方法、邓勇方法等。

基于以上问题，提出一种基于置信估计网络及改进D-S证据理论的结果级多模态融合方法，主要贡献如下。

1) 量化单模态神经网络检测结果置信度：针对神经网络难以衡量检测结果不确定性的问题，提出一种置信估计网络，将目标检测中的分类问题表述为置信估计问题，输出单模态感知网络检测结果的置信估计分数。

2) 改进经典证据理论：针对经典D-S理论在高冲突场景中引发悖论的问题，提出结合置信估计网络改进经典D-S证据理论的方法，克服了经典证

据理论的局限性。

3) 解决分类冲突：针对多模态识别结果不一致的问题，在结果级融合中使用改进的D-S证据理论融合多模态信息，并在KITTI数据集上进行评估试验，结果表明，改进D-S的融合方法能有效解决分类冲突，并在一定程度上提高多模态神经网络检测精度。

## 1 模型架构

提出的多模态融合方法总体结构如图1所示，主要包括：一种置信估计网络，该网络用于评估单模态神经网络识别结果的不确定性，量化识别结果的可靠程度；一种多模态融合方法，以置信估计网络的输出作为权重对目标检测网络的输出结果进行加权、修正以合成证据源，最后利用改进的D-S证据理论进行类别融合得到最终检测结果，弥补了现有改进后的D-S证据理论在多模态融合领域的不足。

模型整体架构如图1所示，首先，接收点云和图像两种模态数据，将其并行地输入到目标检测网络（蓝色虚线框）和置信估计网络（左红色虚线框）。目标检测网络的分类回归分支和边框回归分支分别生成网络预测的类别置信度以及目标边框参数；置信估计网络使用一系列全连接层对网络预测结果进行置信度特征提取，并使用sigmoid函数将结果映射为0~1的标量以表征神经网络预测结果的可靠度。然后将点云的结果坐标进行转换，使其映射至图像坐标系。

在该坐标系下，结合目标检测框的交并比 (IoU) 对神经网络的检测结果进行目标匹配。最终结合置信估计分数、3种分类置信度以及边框参数合成D-S理论的证据信息，并利用改进的D-S证据理论完成对检测结果的类别融合。

### 1.1 置信估计网络

为衡量神经网络检测结果的可靠程度，借鉴域外识别领域 (Out-of-Distribution, OOD)<sup>[14]</sup>以及证据神经网络的工作，设计了一种置信估计网络，以

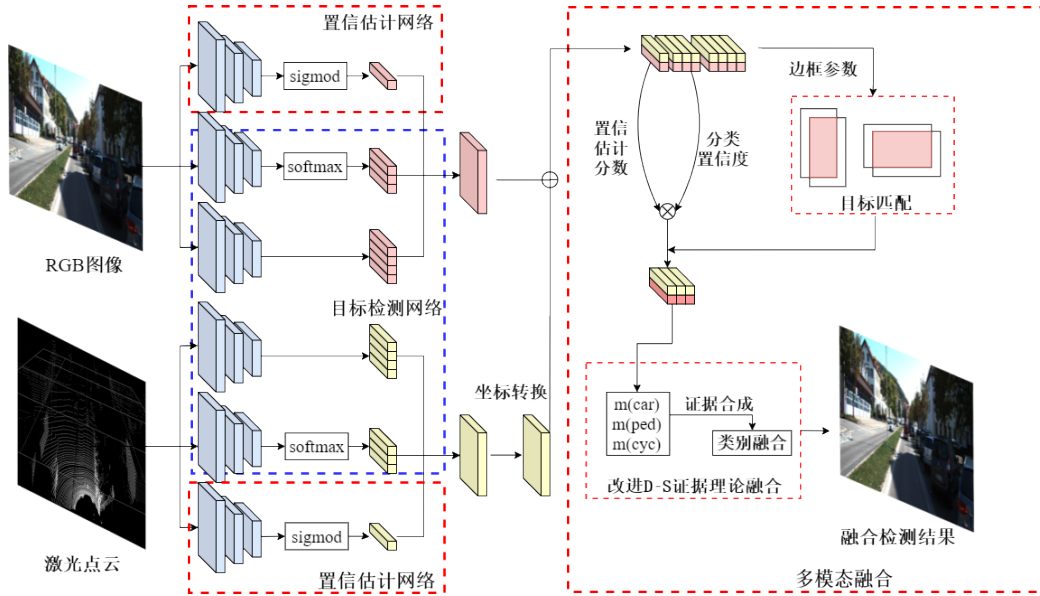


图 1 基于置信估计网络和改进 D-S 证据理论的结果级多模态融合方法

置信估计分数作为网络输出，分数值越大，表示神经网络检测结果越可靠。如图 2 所示，置信估计网络并行于分类预测网络，并与其共享数据特征，在经过一系列线性层提取特征后，将输出值使用 sigmoid 函数进行处理，最终以一个值为 0~1 的一维预测值作为神经网络预测结果的置信度分数。

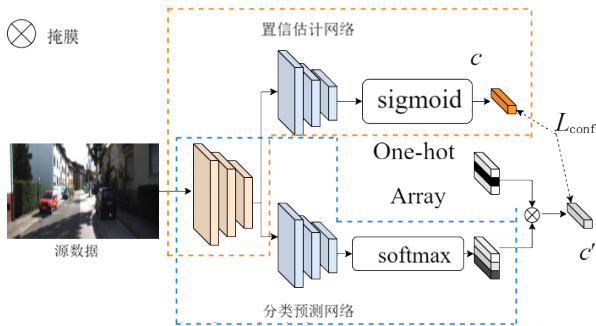


图 2 置信估计网络

网络训练过程中，先对分类预测网络进行训练，使其具备目标分类与识别的能力。分类预测部分训练完成后进行置信估计网络训练，将分类预测部分权重数据作为置信估计网络的预训练权重。在置信估计网络训练过程中冻结分类预测层，使分类层与置信估计层解耦，避免置信估计层干扰原始神经网络的预测。

针对置信估计分数标签难以获取的问题，采用训练时生成标签的方式，将分类预测网络输出的正

确分类类别置信度作为置信估计网络的真值标签。具体做法为根据数据集标签，构建对应正确分类的独热编码 (One-Hot Array)，利用独热编码制作掩膜对分类预测网络的结果进行过滤，得到神经网络检测该目标时，真实类别对应的置信度分数。以本文的三分类任务 (图 2) 为例，针对源数据中的目标，分类预测网络输出该目标的车、人、非机动车三类置信度  $[c_1, c_2, c_3]$ 。独热编码  $[1, 0, 0]$  由标签中该目标的真实类别构建，真实类别置 1，错误类别置 0。使用独热编码向量对三类置信度进行过滤，得到置信度分数  $c'$ 。最后将此置信度分数作为置信估计网络训练的真值标签，并将该分数与置信估计网络输出的置信估计分数联合以计算损失。

训练置信估计网络时，因引入了预训练权重，分类层网络已经具备了一定程度的特征提取、目标分类的能力，因此，错误分类的概率较小，置信估计网络难以学习到预测错误时的置信估计分数。针对该问题，设计了式 (1) 作为置信估计网络的损失函数，以增大网络在学习过程中发生错误时的惩罚度。

$$L_{\text{conf}} = \frac{1}{N} \sum_{i=1}^N -\log \left[ \left( 1 - |c_i - c'_i| \right)^2 \right] \quad (1)$$

式中： $L_{\text{conf}}$  为损失值； $N$  为神经网络检测的目标数量； $c_i$  和  $c'_i$  分别为置信估计网络的真值和预测值。

由于 $c_i$ 和 $c_j$ 的值均在0~1之间，为提高神经网络的训练效果，在损失函数中引入了log函数，当预测值与标签的曼哈顿距离接近1时，损失值较小，接近0时，损失值趋向于正无穷。

### 1.2 基于KM算法的目标边框匹配

为便于进行目标匹配，利用相机的内外参矩阵将激光雷达的检测结果坐标投影至图像坐标系下。在同一场景下的同一时刻，激光雷达与图像的检测边框是天然的二分图数据结构，因此，使用二分图匹配算法将两种模态的目标边框进行融合。CARION等<sup>[15]</sup>在将transformer网络用于图像识别时，使用匈牙利算法代替NMS过滤，匈牙利算法的O(MN)时间复杂度使检测效率得到极大提升。在匈牙利算法的基础上，KM算法面向带有权重信息的二分图进行匹配，可保证结果是当前二分图中的最优匹配。以点云、图像目标边框信息作为二分图的左右顶标，以二者的交并比(IoU)作为二分图的权值，利用KM算法对多模态识别结果的目标边框进行匹配，可以在一定程度上避免出现某一检测器一个检测结果对应另一检测器多个检测结果的情况。

结合KM算法逻辑，边框匹配逻辑如图3所示。以目标数量较少的模态边框数据初始化为左顶标，另一模态初始化为右顶标，以交并比为权重建立有权二分图并初始化邻接矩阵；使用匈牙利算法匹配二分图，若未找到完备匹配则修改可行顶标值并重新匹配。迭代过程中对比两次迭代结果的子图是否相等，不相等则重新迭代。迭代结束后，对结果中匹配项的边框坐标、中心点坐标取均值后存入结果集，未匹配项根据置信估计分数进行筛选，若该分数大于0.5则直接存入结果集，否则直接丢弃。

### 1.3 基于改进D-S证据理论的目标类别融合

实际交通场景中由于外界因素的干扰，不同检测器的识别结果会产生差异，在进行数据融合时，分类冲突时有发生。假设有A1、A2、A3三个独立事件，分别有m1、m2、m3三个证据源获得证据，

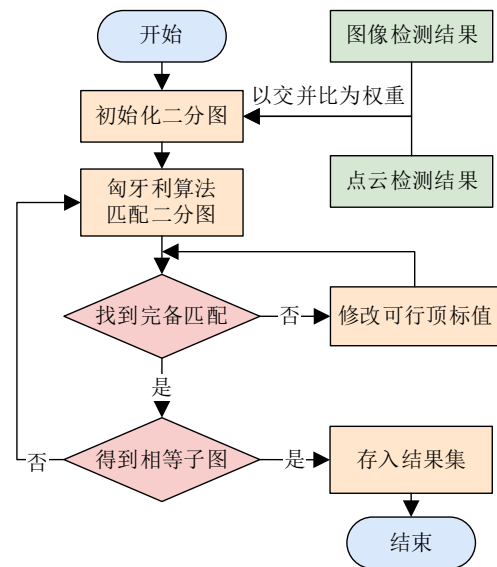


图3 基于KM算法的目标边框匹配逻辑

D-S理论在多分类应用中，A1事件的合成计算式为：

$$m(A) = \frac{1}{1-k} \sum_{A_1 \cap A_2 \cap A_3 = A} m_1(A_1)m_2(A_2)m_3(A_3)。 \quad (2)$$

式中： $m$ 函数反映事件置信度的大小； $k$ 为冲突因子，表征了证据之间的冲突程度，其计算式为：

$$k = \sum_{A_1 \cap A_2 \cap A_3 = \emptyset} m_1(A_1)m_2(A_2)m_3(A_3)。 \quad (3)$$

在出现分类冲突时，D-S理论基于置信度分布对证据进行重新排布，最终取置信度分数最大值对应类别作为冲突修正结果。经典D-S的局限性体现在证据间的高冲突场景，即出现某证据源一票否决给出为0或接近0的证据得分，此时经典D-S理论会合成与真实情况不符的结果。计算过程中表现为当冲突因子 $k$ 较大即 $k$ 接近于1时，会产生Zadeh悖论，影响结论合成。针对此局限性，对D-S证据理论进行了改进，以不同模态的检测结果作为证据源，将证据源对应的置信估计分数视为证据源概率密度的正确程度，对证据源进行加权修正，可以得到维度为 $n$ 的权重系数向量。

$$W = [w_1, \dots, w_i, \dots, w_n]。 \quad (4)$$

式中： $W$ 为权重系数向量； $w_i$ 为向量中置信估计网络输出的权重系数。权重系数对 $n$ 条证据加权平均后得到修正后的期望证据 $M$ 。

$$M = \sum_{i=1}^n w_i m_i \circ \quad (5)$$

在高冲突场景中,使用修正后的公式重新计算冲突因子,修正后的冲突因子计算式为:

$$k = \sum_{A_1 \cap A_2 \cap A_3 = \emptyset} w_1 m_1(A_1) w_2 m_2(A_2) w_3 m_3(A_3) \circ \quad (6)$$

最后,将利用D-S证据理论的组合规则对期望证据 $M$ 迭代组合 $n-1$ 次后的结果作为 $n$ 条证据的合成结果,合成结果最大值对应的类别即类别融合结果。

## 2 试验

设计了置信估计网络试验评估置信估计分数的有效性。基于KITTI数据集选择了3个经典图像感知网络(YOLOV8<sup>[16]</sup>、FasterRCNN<sup>[17]</sup>、CenterNet<sup>[18]</sup>),以及3个点云感知网络(PVRCNN<sup>[19]</sup>、SECOND<sup>[20]</sup>、PointPillar<sup>[21]</sup>)进行融合试验,用于验证本文融合方法的通用性。对比了本文融合方法、单一感知网络和经典D-S融合结果,用于验证本文融合方法的识别准确性,解决分类冲突的能力与克服经典D-S证据理论局限性的能力。

### 2.1 置信估计网络试验

为验证置信估计网络在分辨神经网络错误检测结果的有效性,统计了图像检测网络CenterNet与点云检测网络PVRCNN的错误检测结果数量,并记录对应的置信估计网络输出。对于1 821个错误检测结果,其置信度分数分布如图4所示,图4横轴描述单模态检测器以及检测类别,纵轴描述目标数量,柱状图以不同颜色区分各个分数区间的目标数量。由统计结果可知,置信估计网络能以较低的置信度分数( $<0.5$ )有效分辨错误检测结果。图5分别展示了常见错检结果即检测重复、分类错误以及检测错误场景下的检测类别、边框及其对应的置信估计分数。

由于非机动车与行人大小、形状相近,神经网络在分类时常会将两种类别混淆,而在受阴影、光照、小目标的影响时,神经网络往往会错误地识别一些目标。

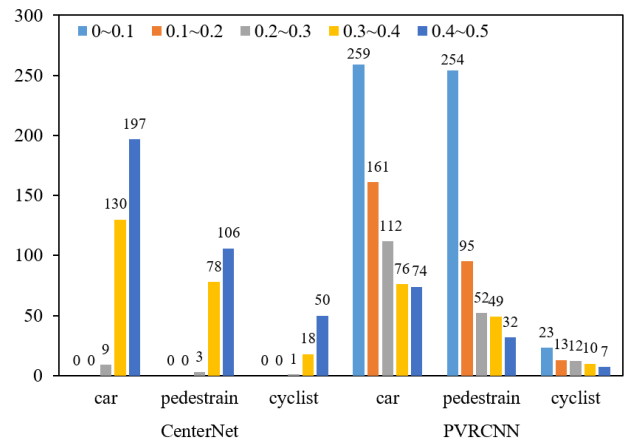


图4 错检目标的置信估计分数分布

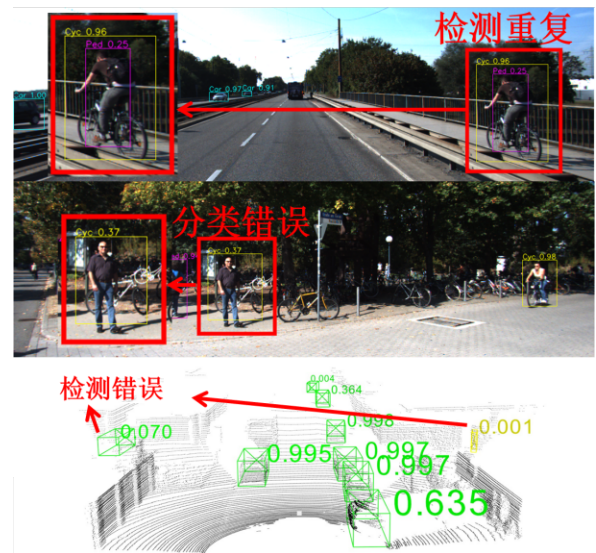


图5 错检情况下的置信估计分数

图5中,在检测重复所对应的场景中,对于同一目标的重复检测结果,置信估计网络对于行人类的置信分数为0.96,对于非机动车类的置信分数为0.25,表示置信估计网络认为非机动车的检测结果更可信。在分类错误所对应的场景中,置信估计网络对于分类错误结果的置信分数为0.37,表示该识别结果可信度较低。在检测错误所对应的场景中,神经网络将左侧建筑物点云识别为车辆,右侧植物识别为行人,两处检测错误的置信分数分别为0.07和0.001,表示该识别结果可信度较低。

通过置信估计网络试验分析可以得出:一方面,置信估计网络可以对错检目标生成较低分数值,表示该检测结果不可信,因此,在目标匹配过

程中，可以根据置信估计分数判断未完成关联的单模态检测结果是否应该加入融合识别结果集中，以减少单模态网络错检对融合结果精确度造成的影响；另一方面，置信估计网络可以对错误分类生成较低分数值，表示该分类不可信，因此，置信估计网络是改进 D-S 理论解决分类冲突的有效工具。

### 2.2 融合方法试验

结合单模态神经网络与置信估计网络的输出，使融合前的结果携带置信估计分数，根据 1.3 节提出的方法进行证据合成，若出现分类冲突则通过 D-S 理论融合多模态预测的分类置信度，得到融合后的重新分配的类别置信度得分，选取得分最高的分类作为融合结果的类别。

为验证本文融合方法是否能克服经典 D-S 证据理论的局限性，设计了消融试验。以图像检测算法 CenterNet 与点云检测算法 PVRCNN 的多模态融合为例，在测试集中目标类别融合共发生 1 037 次分类冲突。经对比试验，将冲突系数阈值设置在 0.9 时，本文融合方法能在检测结果 mAP 上较经典 D-S 理论有稳定提升。所以将冲突系数  $k$  高于 0.9 时的证据合成场景称为高冲突场景，当冲突系数高于此阈值时，使用改进 D-S 合成结果，否则使用经典 D-S 合成结果。以其中一次高冲突分类冲突场景为例，如图 6 所示，当证据间发生冲突时，经典 D-S 证据理论合成了错误的识别结果，而本文融合方法基于置信估计分数加权修正证据后，能克服经典 D-S 理论的上述问题，合成正确识别结果。

表 1 列举了图 6 场景下不同证据源的证据合成分数以及经典 D-S 与本文融合方法置信度分配的结果。为防止网络模型对检测结果过度自信<sup>[8]</sup>，使用未经过 softmax 处理的目标置信度进行证据合成。该场景下的证据源有两类（相机、激光雷达），对于同一个目标仅对应一个真实类别，所以车、人、非机动车之间的交集为空。冲突因子可由式 (7) 计算得出。

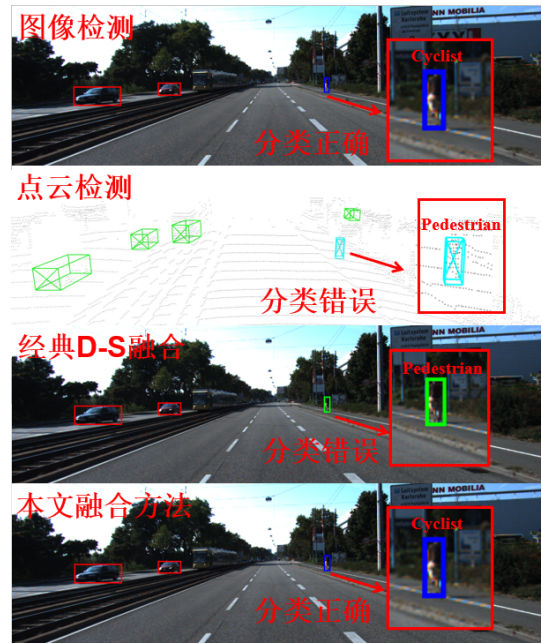


图 6 证据源高冲突场景下的分类表现

$$k = \sum_{A_1 \cap A_2 \cap A_3 = \emptyset} m_1(A_i)m_2(A_i) = 1 - m_1(A_1)m_2(A_1) - m_1(A_2)m_2(A_2) - m_1(A_3)m_2(A_3) = 0.9875 \quad (7)$$

此时冲突因子  $k$  大于阈值 0.9，经典 D-S 失效。由改进 D-S 合成的冲突因子可计算为：

$$k = \sum_{A_1 \cap A_2 \cap A_3 = \emptyset} m_1(A_i)m_2(A_i)c_i = 0.5900 \quad (8)$$

表 1 证据源高冲突场景下的置信度重分配 %

检测目标	检测器	车	人	非机动车	置信估计分数
非机动车	相机	3.83	2.94	59.85	59.84
	激光雷达	0.23	37.64	0.22	2.99
	经典 D-S	0.78	98.74	1.18	
	本文方法	1.72	2.81	26.89	

改进后的冲突因子未超过阈值，使用该冲突因子进行后续合成，合成结果见表 1。在不同检测器间出现分类冲突时，置信估计网络能以分数的高低区分识别结果的正误。而结合置信估计分数修正证据后的融合方法能克服经典 D-S 证据理论局限性，输出正确检测结果。

### 2.3 融合结果对比

为对比融合方法之间的效果，设计了融合结果对比试验，表 2 中，在单模态检测列，列举了使用

图像检测、点云检测的单模态神经网络的 mAP 值。其余三列数据展示了使用经典 D-S 证据理论融合、本文融合方法的 mAP 值,以及本文融合的 mAP 值较经典 D-S 理论的 mAP 值提升。由表中数据可知,进行的 6 组试验中:在多模态融合结果与图像检测器检测结果的对比中,其中 5 组都能对检测结果提升 0.59%~2.00%;在本文融合方法与经典 D-S 融合的

对比中,以任意方式组合单模态检测器,本文融合方法较经典 D-S 理论检测结果提升 0.01%~0.81%。

图 7 以柱状图的形式展示了多模态融合结果与点云检测器的检测结果对比。在与点云检测器的对比中,经典 D-S 融合和本文融合方法在任意组合方式中均较点云检测器提升 10% 左右,且本文融合方法的 mAP 值均高于经典 D-S 融合。

表 2 融合试验结果数据对比

检测器	mAP @0.5%			
	单模态检测	经典 D-S 融合 (较图像提升)	本文方法 (较图像提升)	本文方法较 经典 D-S 提升
FasterRCNN	74.58			
FasterRCNN+PVR CNN	69.14	75.97 (+1.39)	76.58 (+2.00)	+0.61
FasterRCNN+PointPillar	62.25	74.23 (-0.35)	74.51 (-0.07)	+0.28
FasterRCNN+SECOND	67.01	75.16 (+0.58)	75.17 (+0.59)	+0.01
CenterNet	75.29			
CenterNet+PVR CNN	69.14	75.55 (+0.26)	75.91 (+0.62)	+0.36
CenterNet+PointPillar	62.25	75.88 (+0.59)	76.69 (+1.40)	+0.81
CenterNet+SECOND	67.01	76.14 (+0.85)	76.86 (+1.57)	+0.72
YOLO V8	72.71			
YOLO V8 + PVR CNN	69.14	78.89 (+6.18)	79.35 (+6.64)	+0.46
YOLO V8 + PointPillar	62.25	77.59 (+4.88)	77.68 (+4.97)	+0.09
YOLO V8 + SECOND	67.01	77.60 (+4.89)	77.79 (+5.08)	+0.19

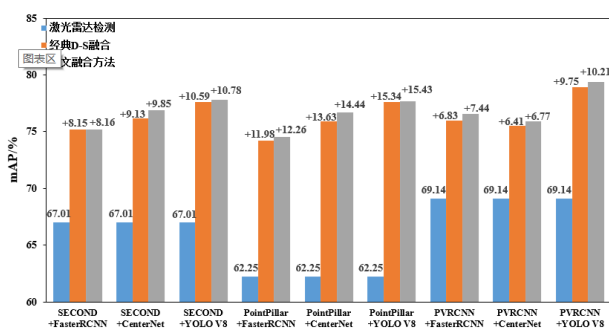


图 7 融合试验结果对比

由试验数据可知,提出的改进 D-S 融合方法能在一定程度上提高基础感知网络的识别准确性,还能有效减少分类冲突,并克服经典 D-S 证据理论的局限性。

### 3 结论

1) 基于置信估计网络和改进 D-S 证据理论的

多模态融合方法能有效地利用多模态间的互补性与一致性,使融合结果精确度相较单一感知网络有较大提升。本文方法较单模态图像检测器最高提升 6.64%,较单模态点云检测器最高提升 15.43%,且融合部分计算量主要集中在检测结果生成后的阶段,即检测框级的融合,其时间复杂度为  $O(n^3)$ 。融合算法不依赖 GPU 硬件,有一定的实际应用价值。

2) 经试验验证,本文融合方法的 mAP 提升值均高于经典 D-S 融合,其中 PointPillar 与 CenterNet 的融合试验中,本文融合方法较经典 D-S 融合高 0.81%。此部分提升量主要发生在冲突因子  $k$  高于 0.9 的高冲突场景中,体现了本文融合方法克服经典 D-S 理论局限性的能力。通过以置信估计分数加

权形成证据,以此改进D-S证据理论并完成目标类别融合,有效减少了分类冲突情况,并克服了经典D-S证据理论的局限性,具有较好的应用前景。

3) 由于试验所使用的数据集中恶劣天气场景较少,源数据因天气原因造成的分类冲突场景较少,且相比车辆,非机动车与行人的目标数量较少。因此,能应用改进D-S证据理论克服经典D-S

理论的场景相对较少,所以试验测试结果中本文融合方法相较经典D-S融合方法的提升量不够明显。后续研究可以采集恶劣天气下的交通场景数据,以扩充试验所用数据,或使用数据增强技术扩充数据集,进一步论证提出的融合方法修正分类冲突的能力。

## 参考文献 (References)

- [1] 刘法旺,徐晓庆,陈贞,等.搭载自动驾驶功能的智能网联汽车安全测试与评估方法研究[J].汽车工程学报,2022,12(3):221-227.  
LIU Fawang, XU Xiaoqing, CHEN Zhen, et al. Research on Safety Testing and Assessment Methods for Intelligent and Connected Vehicles with Autonomous Driving Functions [J]. Chinese Journal of Automotive Engineering, 2022, 12(3): 221-227. (in Chinese)
- [2] LI Yingwei, YU A W, MENG Tianjian, et al. DeepFusion: Lidar-Camera Deep Fusion for Multi-Modal 3D Object Detection[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-24, 2022, New Orleans, LA, USA. Piscataway NJ: IEEE, c2022: 17161-17170.
- [3] WU Xiaopei, PENG Liang, YANG Honghui, et al. Sparse Fuse Dense: Towards High Quality 3D Detection with Depth Completion [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-24, 2022, New Orleans, LA, USA. Piscataway NJ: IEEE, c2022: 5418-5427.
- [4] PANG Su, 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, Oct. 24 2020-Jan. 24, 2021, Las Vegas, NV, USA. Piscataway NJ: IEEE, c2020: 10386-10393.
- [5] PANG Su, MORRIS D, RADHA H. Fast-CLOCs: Fast Camera-LiDAR Object Candidates Fusion for 3D Object Detection [C]//2022 IEEE/CVF Winter Conference on Applications of Computer Vision, Jan. 3-8, 2022, Waikoloa, HI, USA. Piscataway NJ: IEEE, c2022: 187-196.
- [6] GAWLIKOWSKI J, TASSI C R N, ALI M, et al. A Survey of Uncertainty in Deep Neural Networks [Z]. arXiv Preprint: 2107.03342, 2021.
- [7] ZHAO Xujiang, OU Yuzhe, KAPLAN L, et al. Quantifying Classification Uncertainty Using Regularized Evidential Neural Networks[Z]. arXiv Preprint: 1910.06864, 2019.
- [8] HAN Zongbo, ZHANG Changqing, FU Huazhu, et al. Trusted Multi-View Classification with Dynamic Evidential Fusion [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(2): 2551-2566.
- [9] CORBIÈRE C, THOME N, BAR-HEN A, et al. Addressing Failure Prediction by Learning Model Confidence[C]//Advances in Neural Information Processing Systems, 32nd Conference on Neural Information Processing Systems, Dec. 8-14, 2020, Vancouver, Canada. 2020: 2879-2890.
- [10] DEVRIES T, TAYLOR G W. Learning Confidence for Out-of-Distribution Detection in Neural Networks [Z]. arXiv Preprint: 1802.04865, 2018.
- [11] SHAFER G. Dempster-Shafer Theory [J]. Encyclopedia of Artificial Intelligence, 1992, 1: 330-331.
- [12] 黄建根,陈祯福,裴晓飞,等.基于D-S推理算法的智能车辆轨迹预测研究[J].汽车工程学报,2018,8(1): 24-30.  
HUANG Jiangen, CHEN Zhenfu, PEI Xiaofei, et al. Research on Trajectory Prediction for Intelligent Vehicles Based on D-S Reasoning [J]. Chinese Journal of Automotive Engineering, 2018, 8(1): 24-30. (in Chinese)
- [13] 刘晓东,邓锦宇.基于冲突证据修正的DS改进方法[J].电子测量与仪器学报,2017,31(9): 1499-1506.  
LIU Xiaodong, DENG Jinyu. Improved DS Method Based on Conflict Evidence Correction [J]. Journal of Electronic Measurement and Instrumentation, 2017, 31(9): 1499-1506. (in Chinese)
- [14] MOON J, KIM J, SHIN Y, et al. Confidence-Aware Learning for Deep Neural Networks [C]//Proceedings of the 37th International Conference on Machine Learning. PMLR 119, 2020: 7034-7044.
- [15] CARION N, MASSA F, SYNNAEVE G, et al. End-to-End Object Detection with Transformers [C]//Computer Vision—ECCV 2020: 16th European Conference, Aug. 23-

- 28, 2020. Berlin: Springer International Publishing, 2020: 213–229.
- [16] Ultralytics. Ultralytics YOLO v8 Modes [EB/OL]. (2023–11–22) [2024–1–10] <https://docs.ultralytics.com/modes/>.
- [17] REN Shaoqing, HE Kaiming, 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.
- [18] DUAN Kaiwen, BAI Song, XIE Lingxi, et al. CenterNet: Keypoint Triplets for Object Detection [C]//2019 IEEE/CVF International Conference on Computer Vision, Oct. 27–Nov. 2, 2019, Seoul, Korea (South). Piscataway NJ: IEEE, c2019: 6569–6578.
- [19] SHI Shaoshuai, GUO Chaoxu, JIANG Li, et al. PV-RCNN: Point-Voxel Feature Set Abstraction for 3D Object Detection [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 13–19, 2020, Seattle, WA, USA. Piscataway NJ: IEEE, c2020: 10529–10538.
- [20] YAN Yan, MAO Yuxing, LI Bo. SECOND: Sparsely Embedded Convolutional Detection [J]. Sensors, 2018, 18(10): 1–17.
- [21] 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, June 15–20, 2019, Long Beach, CA, USA. Piscataway NJ: IEEE, c2019: 12697–12705.

#### 作者简介



程腾 (1983–), 男, 重庆梁平人, 博士, 副研究员, 主要研究方向为新一代车联网与车辆信息安全, 包括密码学、量子密码、区块链, 自动驾驶环境下的多模融合, 以及智驾感知和智能座舱交互。

E-mail: cht616@hfut.edu.cn

#### 通信作者



郭利港 (1997–), 男, 内蒙古包头人, 硕士研究生, 主要研究方向为自动驾驶环境感知和多模态融合目标检测。

E-mail: guo\_ligang@foxmail.com