

## 自动驾驶目标检测不确定性估计方法综述

赵洋, 王潇, 蔡柠泽, 程洪

(电子科技大学 自动化工程学院, 成都 611731)

**摘要:** 随着自动驾驶技术的发展, 目标检测的准确性和可靠性变得至关重要。深度学习作为自动驾驶系统中的核心组成部分, 其预测结果的不确定性估计对于系统的安全性和稳定性具有显著影响。总结了深度学习不确定性估计理论在自动驾驶目标检测中的应用, 并探讨了有效的不确定性评价体系的重要性。介绍了深度学习不确定性估计的基本理论, 包括贝叶斯神经网络、蒙特卡洛方法以及集成学习方法等。这些方法通过不同的途径量化模型预测的不确定性, 为自动驾驶系统提供了更丰富的信息。深入探讨了自动驾驶目标检测中不确定性估计的应用。通过案例分析, 展示了如何利用不确定性信息来提高目标检测的准确性, 特别是在面对复杂环境和极端条件时, 不确定性估计可以作为决策支持, 帮助系统避免潜在的风险。总结了自动驾驶目标检测不确定性估计评价指标, 同时, 考虑了模型的预测性能、不确定性估计的准确性。

**关键词:** 自动驾驶; 目标检测; 深度学习; 不确定性估计

中图分类号: TP29 文献标志码: A DOI: 10.3969/j.issn.2095-1469.2024.05.02

## A Review of Uncertainty Estimation Methods in Autonomous Driving Object Detection

ZHAO Yang, WANG Xiao, CAI Ningze, CHENG Hong

(School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China)

**Abstract:** With the advancement of autonomous driving technology, the accuracy and reliability of object detection have become increasingly crucial. Deep learning, as a core component of autonomous driving systems, significantly influences the safety and stability of these systems by estimating the uncertainty in predictive results. The paper summarizes the application of deep learning uncertainty estimation in autonomous driving object detection and discusses the significance of an effective uncertainty evaluation system. Firstly, the paper introduces the fundamental theories of deep learning uncertainty estimation, including Bayesian neural networks, Monte Carlo methods, and ensemble learning. These methods quantify model prediction uncertainty in different ways, providing autonomous driving systems with richer information. Secondly, the paper delves into the application of uncertainty estimation in autonomous driving object detection. Through case studies, it demonstrates how uncertainty information can be used to improve detection accuracy, especially in complex environments and extreme conditions. In these scenarios, uncertainty estimation provides decision support, helping the system avoid potential risks. Lastly, the paper

收稿日期: 2024-05-15 改稿日期: 2024-07-19 网络首发日期: 2024-07-30

基金项目: 国家重点研发计划项目(2022YFB2503004); 中央高校基本业务费项目(ZYGX2022J017); 机器人与智能系统国际联合研究中心开放基金项目(JQZN2023-005)

参考文献引用格式:

赵洋, 王潇, 蔡柠泽, 等. 自动驾驶目标检测不确定性估计方法综述[J]. 汽车工程学报, 2024, 14(5): 760-771.

ZHAO Yang, WANG Xiao, CAI Ningze, et al. A Review of Uncertainty Estimation Methods in Autonomous Driving Object Detection[J]. Chinese Journal of Automotive Engineering, 2024, 14(5): 760-771. (in Chinese)



summarizes the evaluation metrics for uncertainty estimation in autonomous driving object detection, considering both the model's predictive performance and the accuracy of the uncertainty estimation.

**Keywords:** autonomous driving; object recognition; deep learning; uncertainty estimation

自动驾驶技术作为现代交通系统的一次革命，其核心目标是实现安全、高效、可靠的无人驾驶。在这一愿景中，目标检测技术扮演着至关重要的角色。它负责识别和定位车辆周围的行人、车辆、交通标志以及其他障碍物，为自动驾驶车辆提供决策支持。然而，目标检测的准确性并不是绝对的，它受到多种因素的影响，包括传感器的局限性、环境的复杂性以及动态场景的不可预测性。这些因素共同构成了目标检测中的不确定性问题，对自动驾驶系统的安全性和可靠性构成了挑战。

尽管目标检测技术已经取得了显著的进展，但在实际应用中仍然面临着不确定性的挑战<sup>[1-3]</sup>。FENG Di等<sup>[4]</sup>指出不确定性可能源自传感器的噪声、目标的遮挡、快速变化的环境条件、传感器的视野限制以及目标的快速移动等。例如，在恶劣天气下，摄像头可能难以捕捉到清晰的图像；而在城市环境中，高楼大厦和其他障碍物可能会阻挡传感器的视线，导致检测结果的不完整。此外，目标的快速移动也可能导致检测系统难以准确跟踪。

为了提高目标检测的鲁棒性，相关学者开始关注如何量化和处理不确定性。量化不确定性可以帮助系统更好地理解检测结果的可信度，从而作出更加合理的决策。目前，已经有一些方法被提出来评估目标检测中的不确定性，包括基于统计的方法、基于模型的方法以及基于深度学习的方法。这些方法各有优势和局限性，如何选择合适的方法并将其有效地应用于实际的自动驾驶系统中，是一个值得深入研究的问题。

本文旨在对自动驾驶目标检测中不确定性问题的研究进展进行综述，包括不确定性的来源、评估方法以及如何利用不确定性信息来提高系统性能。首先，总结了目标检测不确定性的来源和表现形式，分析其对自动驾驶系统的影响。然后详细介绍了现有的不确定性评估和量化技术，包括它们的原

理、优势和局限性，讨论了如何将不确定性信息整合到决策过程中。最后，总结全文，并提出这一领域未来的研究方向。

## 1 不确定性估计方法

深度学习由于其突出的性能优势，在自动驾驶领域得到了广泛的使用，深度神经网络成为了自动驾驶车端环境感知常用模型之一。近年来，虽然自动驾驶技术取得了长足的进步，自动驾驶车辆依然没有放开全场景行驶，目前低阶自动驾驶系统在运行时需要驾驶人员实时监督，高阶自动驾驶系统仅限于低速行驶场景以及封闭试验路段。究其根本，自动驾驶系统仍然存在未解决的安全隐患。对于车端环境感知模块来说，当自动驾驶汽车在开放路段行驶时，会遇到各种毫无预兆的情况，比如极端天气、障碍物遮挡、传感器失灵等，这些都是无法避免的，要保证自动驾驶汽车的安全性，必须要求环境感知的准确性以及对危险情况的敏感性。出于对自动驾驶汽车安全性的考虑，势必需要感知算法同时具备精准检测和险情识别的能力，目前感知算法在深度学习框架的发展下检测性能突飞猛进，关键问题落在了如何使自动驾驶车端环境感知拥有险情识别的能力。

2017年，AMINI等<sup>[5]</sup>提出了基于贝叶斯的深度估计方法，此方法通过近似权值的后验分布来估计算法模型的不确定性。同年，GAL<sup>[6]</sup>提出了基于类贝叶斯的变分推断深度学习不确定性估计方法，此方法所应用理论同AMINI相同，但是近似权值的后验分布手段不同，GAL使用了基于蒙特卡洛的统计学方法，使用Dropout模块对深度学习模型进行多次前向变分推断，最后通过计算方差的方式得到深度学习模型的不确定性。MICHELMORE等<sup>[7]</sup>于2018年提出了在不改变网络结构的前提下，从深度神经网络收集近似不确定性信息并且

预测崩溃率的不确定性估计方法。2019年, TAI Lei等<sup>[8]</sup>使用随机生成器在定向域风格化输入图像, 检测一系列图像中的不确定性, 不确定性最低的一帧图像用于使用。2020年, MICHELMORE等<sup>[9]</sup>及其团队开发了模拟器框架, 在贝叶斯网络的影响下, 对端到端网络不确定性进行评估。以上不确定性评价方法分别应用在各种任务背景的深度学习算法之下, 但是没有针对自动驾驶深度学习感知算法这一特定场景所应用的方法。

目前通用的不确定性估计方法主要有4种: Monte Carlo Dropout(MC-Dropout), Deep Ensembles(DE), Direct Modeling和Error Propagation。MC-Dropout和Deep Ensembles适用于对模型认知不确定性建模, 而Direct Modeling用于捕获偶然不确定性。Error Propagation既可以用于获取认知不确定性, 也可以获取偶然不确定性, 这取决于它的关注点在哪一方面。

GAL<sup>[6]</sup>提出的MC-Dropout方法将基于dropout的神经网络训练与贝叶斯神经网络(BNNs)中的变分推断(VI)联系起来。这一理论指出, 使用具有dropout的随机梯度下降(SGD)训练网络等效于优化近似于式(1)的后验分布。该方法后来得到扩展, 通过将dropout概率视为超参数来找到最佳dropout的值<sup>[10]</sup>。在测试时, 通过多次运行, 推理在启用了dropout的情况下生成来自近似后验分布的样本。设 $T$ 为具有dropout的前馈传递的总数,  $W_i$ 是dropout之后的网络权重的样本。预测概率可以从 $T$ 生成的样本中近似为:

$$p(y|x, D) = \int p(y|x, W)P(W|D)dW。 \quad (1)$$

$$p(y|x, D) \approx \frac{1}{T} \sum_{i=1}^T p(y|x, W_i)。 \quad (2)$$

对于分类任务, MC-Dropout方法中, 式(2)对应于预测分类概率的 $T$ 个数的平均, 例如softmax分数。对于回归任务, 这个方程可以看作是具有 $T$ 等权重分量分布的混合。在这种情况下, 样本均值和方差可用于描述预测概率分布。通常, MC-Dropout提供了一种在贝叶斯神经网络执行近似推理的实用方法, 并且这种方式是可以扩展到大型数

据集和大型网络构架中的。但是, MC-Dropout在测试期间需要多次随机运行, 通常为10到50次<sup>[6]</sup>。

MC-Dropout是一种深度学习中的正则化技术, 它通过在训练过程中随机丢弃神经元来增强模型的泛化能力, 同时提供预测的不确定性估计, 这对于需要模型输出不确定性信息的应用场景非常有用。作为一种变分推断方法, MC-Dropout简化了贝叶斯推断的实现, 并且可以轻松集成到现有的神经网络架构中。然而, 它也存在一些缺点, 包括可能增加的训练时间、对超参数的敏感性以及在需要高确定性的场景中可能带来的不确定性问题。此外, MC-Dropout的随机性可能导致模型预测在不同运行中有所变化, 这需要额外的步骤来确保预测的稳定性。尽管如此, MC-Dropout在适当的场景下可以显著提高模型性能, 减少过拟合, 是一种值得考虑的模型正则化策略。

LAKSHMINARAYANAN等<sup>[11]</sup>提出使用网络集合估计预测概率, 其中每个网络的输出被视为来自混合模型的独立样本。集成中的每个网络都使用相同的架构, 但使用随机打乱的训练数据进行训练, 并使用其参数的不同初始化进行训练。因为MC-Dropout也可以解释为网络的集成<sup>[11]</sup>。在实践中, 5个网络的集合已被证明足以近似预测概率<sup>[11-12]</sup>。虽然易于实现, 但深度融合的计算和内存成本在训练和推理期间都与网络数量线性扩展, 限制了其对大型网络架构的适用性<sup>[11]</sup>。

直接建模(Direct Modeling)假设网络输出具有一定的概率分布, 并使用网络输出层直接预测此类分布的参数。与对权重 $W$ 执行边缘化的贝叶斯神经网络不同, 直接建模使用这些权重的点估计值通过 $p(y|x, W)$ 来生成预测概率分布 $p(y|x, D)$ 。估计类别 $C$ 的分类概率的最常见方法是通过softmax分数 $p(y=c|x, W) = \hat{s}_c$ , 这等效于多项式质量函数。至于回归概率, 经常使用高斯分布或高斯混合模型(GMM)<sup>[13-14]</sup>。训练网络是通过最大似然估计来实现的。当使用softmax函数进行分类时,  $L(x, W)$ 被广泛称为交叉熵损失。

GAL等<sup>[10]</sup>没有使用具有交叉熵损失的标准

softmax，而是将 softmax 函数与高斯分布相结合来估计分类不确定性，方法是假设 softmax logit 向量中的每个元素都是独立的高斯分布，其均值和方差由网络输出层直接预测。此外，一些学者<sup>[15-17]</sup>提出，除了直接预测输出概率分布之外，还估计高阶共轭先验。最后，GUSTAFSSON等<sup>[18]</sup>使用深度神经网络直接预测基于能量的模型条件目标密度。由于直接建模通过单向前传递来估计不确定性，所以它比 MC-Dropout 或 Deep Ensembles 更有效。但是，需要修改网络的输出层和损失函数以进行不确定性估计。它还被证明可以在分类<sup>[6, 19]</sup>和回归<sup>[20]</sup>中产生错误校准的概率。

直接建模不确定性估计方法在机器学习领域提供了一种量化预测不确定性的有效手段，它通过构建专门的模型结构来学习数据与结果之间的不确定性关系，从而为决策者带来明确的风险评估和改进的决策质量。这种方法的优点包括其明确性、灵活性、易于集成到现有模型中，以及能适应数据的复杂性和不确定性。然而，它也存在一些缺点，如实现的复杂性、可能增加的计算成本、对超参数调整的高要求、对训练数据质量的依赖性，以及在模型解释性方面的挑战。尽管如此，直接建模不确定性估计方法在适当的应用场景下，可以显著提升模型的可信度和实用性。

误差传播近似于每个激活层中的方差（或不确定性），然后在整个网络中将方差从输入层传播到输出层。POSTELS等<sup>[20]</sup>将 dropout 和批量归一化视为噪声注入过程，以在训练期间学习不确定性，并提出将方差误差近似为噪声层中的协方差矩阵。然后，它们以封闭形式通过下游激活层传播误差。通过这种方式，该方法通过匹配其第一个标准激活层（如 ReLu）将耗时的 MC-Dropout<sup>[6]</sup>替换为不确定性传播层，并通过匹配其一阶和二阶中心矩将标准激活层（如 ReLu）转换为不确定性传播层。

误差传播是一种评估和量化科学实验或工程计算中不确定性的重要方法，它通过分析测量或计算过程中的误差如何传递和放大，以此估计结果的可靠性和准确性。这种方法的优点在于能提高研究的

透明度，优化试验设计，指导基于风险的决策，并促进对不确定性来源的深入理解。然而，误差传播也存在一些缺点，包括计算过程的复杂性、对额外时间和资源的需求、基于特定假设的限制，以及在某些情况下可能对其他不确定性来源的忽视。此外，误差传播的结果有时可能难以解释，特别是涉及多个相互依赖的误差源时。

MC-Dropout、深度集成、直接建模和误差传播是4种旨在提高模型泛化能力和量化不确定性的技术，它们共同关注减少过拟合和增强模型的可靠性。MC-Dropout通过在训练中引入随机性来模拟贝叶斯推断，而深度集成通过结合多个模型的预测来降低方差；直接建模专注于通过复杂模型结构来明确量化预测的不确定性；误差传播则侧重于分析不确定性如何在计算过程中传递和放大。虽然这些方法在实现机制、关注点、计算成本和适用场景上各有不同，例如 MC-Dropout 和直接建模可能增加计算负担，深度集成需要训练多个模型，而误差传播则依赖于对不确定性传递的详细分析，但它们都为提高模型的稳定性和准确性提供了有价值的工具。选择使用哪种方法应根据具体任务的需求和场景特性来决定。

## 2 目标检测不确定性估计方法

### 2.1 目标检测不确定问题分析

对于深度学习算法的不确定性研究，最早是由 GAL等<sup>[10]</sup>提出，他们从实际应用的角度入手，将深度神经网络中的预测不确定性分解为认知不确定性和偶然不确定性。其中，偶然不确定性主要来源于环境变化和传感器噪声等，认知不确定性主要来源于概念模糊、场景丰富度不足、模型架构不合适等。偶然不确定性表示模型捕获的数据噪声，可反映变化环境中的传感器限制，原则上不可消除。认知与偶然不确定性共同组成了感知算法的预测不确定性，可作为算法性能实时监测的重要指标。

### 2.1.1 偶然不确定性问题分析

深度学习偶然不确定由输入传感器数据质量差造成，搭载在自动驾驶系统上的深度学习算法遇到各种恶劣天气就会触发深度学习检测结果高偶然不确定性，如图 1 所示。

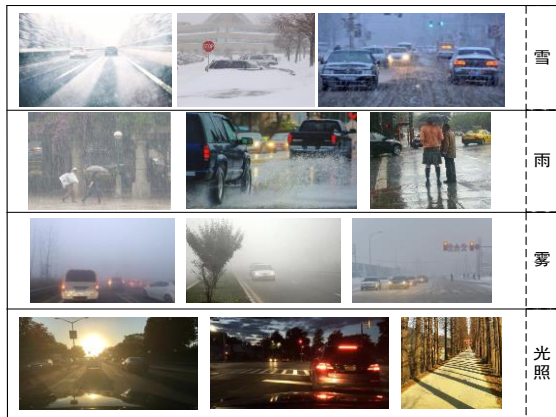


图 1 高偶然不确定触发场景示例

深度学习的偶然不确定性又可被称为数据不确定性，当传感器数据受到外界环境因素的影响时，传感器数据质量下降，统一将这种外界环境因素称为环境噪声。含有环境噪声的传感器数据输入深度学习神经网络，深度学习神经网络对含有环境噪声的数据识别不准确，即深度学习神经网络对输入含有噪声的传感器数据不确定，将这种由环境噪声造成的深度学习神经网络的不确定称为深度学习的偶然不确定性。自动驾驶常用传感器主要有：相机、毫米波雷达和激光雷达。相机和激光雷达容易受到自然环境影响，产生含有噪声的传感器数据，恶劣天气和光照是影响相机图像的主要因素，恶劣天气是影响激光雷达点云的主要因素。如图 2 所示，雪、雨、雾对车辆、行人等目标物体形成部分遮挡，相机获得含有雪、雨、雾的目标物体图像，这些含有环境噪声的图像输入深度学习神经网络会造成深度学习神经网络检测偶然不确定。同样，不良光照造成图像成像畸形，也是诱发深度学习视觉感知神经网络的感知偶然不确定的原因。作为自动驾驶常用传感器，激光雷达也会受到自然环境因素的影响。如图 2 所示，RADIATE 数据集是公开的恶劣天气数据集<sup>[21]</sup>，可以明显看出激光雷达点云生成

在雪、雨、雾天气条件下受到了影响。相比激光雷达，毫米波在雪、雨、雾天气也具有极强的穿透力，因此，在雪、雨、雾天气毫米波雷达检测效果强于激光雷达。

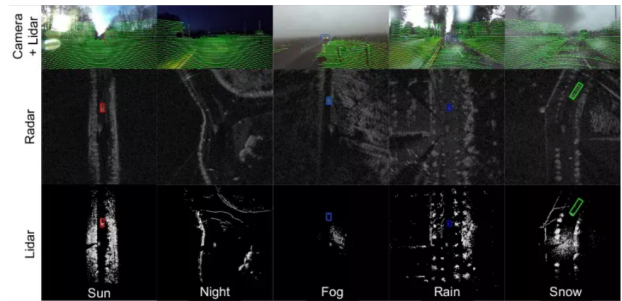


图 2 恶劣天气下各传感器数据

深度学习算法在本质上是大量数据的训练，将同一种类物体的共同特征通过学习记录到其参数中，恶劣天气的影响就像在图像和激光雷达点云中加入了环境噪声，使图像或激光雷达点云记录的本是正常的物体特征发生轻微形变或者纹理变化，进一步导致深度学习神经网络检测不确定。许多试验也证明了这一点，通过加入人工噪声模拟自然环境噪声，通过使用深度学习神经网络检测含有恶劣天气的图像，检测结果都准确说明了这一现象。图 3 展示了人工加噪制造高偶然不确定性场景时，神经网络的表现和自然噪声下神经网络的表现。

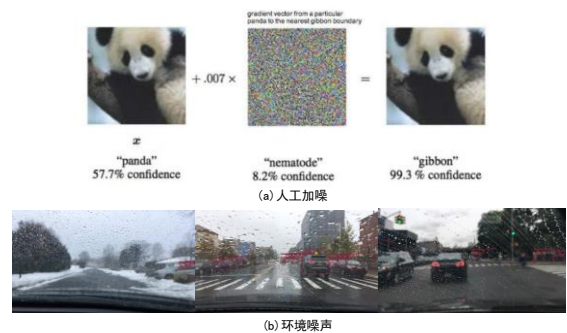


图 3 人工噪声和环境噪声影响试验

如图 3 所示，将加了噪声的图像输入到分类神经网络中，分类神经网络对图像中的种类识别错误，同样将雨天交通道路场景图输入到目标检测神经网络中，神经网络无法通过已学习过的记忆准确识别出车辆位置。上述试验说明由于环境因素影响

或者人工加噪，都会影响深度学习神经网络，使神经网络识别不准确，也表明了环境因素会导致深度学习神经网络产生高偶然不确定性。

### 2.1.2 认知不确定性问题分析

认知不确定是指经过训练的基于深度学习的神经网络对于超出训练数据集之外的同一类物体特征识别不确定，将这种不确定称为深度学习神经网络的认知不确定性。触发深度学习神经网络认知不确定的自动驾驶场景可以简述为以下几种情况：发生局部遮挡、发生形状变化、发生目标物体重叠，如图4所示。深度学习认知不确定常发生于基于视觉的深度学习神经网络检测中，激光雷达和毫米波雷达获取的信息不同于图像信息，因此，上述基于视觉欺骗的手段对雷达类传感器不起作用，也无法影响基于雷达点云的神经网络检测。

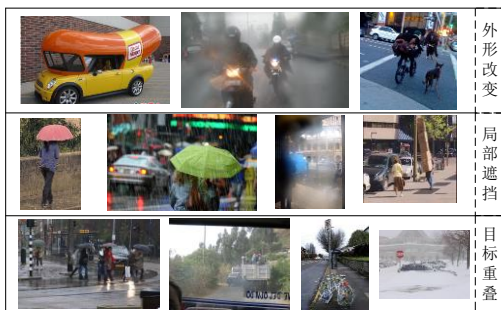


图4 认知不确定触发场景示例

如图4所示，在有些光照良好、传感器数据中环境噪声较小的情况下，由于目标物体自身原因和与其他物体的位置关系就会导致深度学习算法在检测时产生认知不确定。一般来讲，图像可能同时导致深度学习神经网络产生认知和偶然两种不确定。深度学习算法的认知不确定越明显，越表明深度学习算法泛化能力不足，学习能力不够。这同样也解释了为什么深度学习一定存在认知不确定，现有神经网络经过训练后的神经网络参数值都是固定，这表明神经网络在进行检测任务时不会主动去学习迭代参数。如果将神经网络训练比作人类学习，神经网络在学习了指定数据集上的类别特征后，在后续投入应用的阶段中，神经网络所有认知都是基于指定数据集得到。神经网络性能优劣便由两个因素直接决定：神经网络自身结构优劣、数据集对于类别特征

泛化性强弱。由于神经网络结构固定，数据集数据量的局限性，无论是多么强的神经网络，它无法做到对真实世界所有事物的检测完全正确，这就意味着确定性的深度学习神经网络的检测一定带有或高或低的认知不确定。

以基于视觉的目标检测神经网络为例，将触发深度学习认知不确定的图像和没有触发深度学习认知不确定的图像各两张输入到YOLOv5s目标检测网络中，获得的检测结果如图5所示。在图5中，未触发认知不确定的场景中，车辆没有相互遮挡，形状改变的情况，而在触发认知不确定的场景中，车辆之间有遮挡情况，行人姿势有改变。而目标检测网络对两组图片的检测表明，其对未触发认知不确定场景有准确检测，对触发认知不确定场景检测不准确且存在高不确定性。



图5 认知不确定场景触发试验结果

偶然不确定性是目标检测过程中不可避免的一部分，而认知不确定性则是可以通过技术进步和数据积累来减少的。在自动驾驶目标检测中，降低偶然不确定性可以通过改进传感器技术和数据处理算法来实现，而减少认知不确定性则需要更复杂的模型训练和更丰富的环境数据。两者共同作用于目标检测的准确性和可靠性，是自动驾驶系统性能提升的关键因素。在实际应用中，自动驾驶系统需要同时考虑这两种不确定性，通过传感器融合、多模态学习、模型优化等手段来提高目标检测的准确性，从而确保自动驾驶的安全性和效率。

## 2.2 目标检测不确定性估计研究现状

### 2.2.1 偶然不确定性问题研究方法

在 2.1.1 节中提到, 偶然不确定性是数据因素造成, 输入神经网络的数据出现不稳定或者不正确的情况时, 就会导致目标检测结果无法确定对错。直接建模法是对可能出现偶然不确定性的目标检测结果进行侧面监督的一种方式, 通过构建神经网络来拟合目标检测结果的概率分布, 得到噪声分布来显性表达目标检测结果的偶然不确定性的

高低。如图 6 所示, FENG Di 等<sup>[4]</sup>提出了一个基于激光雷达和图像融合的两阶段目标检测网络, 并使用直接建模的方式估计了激光雷达第一阶段检测结果的不确定性, 使用了不确定性采样的方式提升了整个网络的鲁棒性。该方法提出一种概率深度神经网络, 用于结合激光雷达点云和 RGB 图像, 以实现鲁棒和准确的 3D 目标检测。该方法明确地对分类和回归任务中的不确定性进行建模, 并利用这些不确定性通过采样机制来训练融合网络。此方法使用直接建模法对第一阶段的激光雷达检测网络中的偶然不确定性进行拟合。

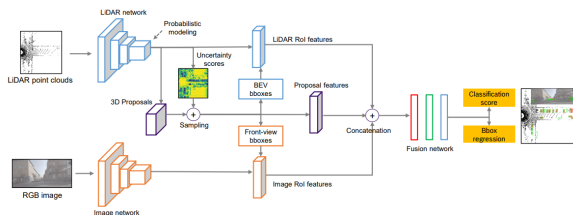


图6 阶段性不确定性估计在目标检测中的应用<sup>[4]</sup>

使用直接建模的方式估计目标检测的偶然不确定性是一种常用偶然不确定性估计方式。ZHAO Yang等<sup>[22]</sup>提出一个基于毫米波雷达和图像融合的一阶段网络, 并使用直接建模的方式估计了检测结果的不确定性, 之后使用此不确定性提升了检测精度。如图 7 所示, 该方法对融合了毫米波雷达特征的目标检测结果进行不确定性建模, 使用高斯分布拟合目标检测回归和分类结果的概率分布, 使用线性神经网络模型输出目标检测结果对应的方差值, 并将输出的方差作为目标检测结果的偶然不确定性估计值。

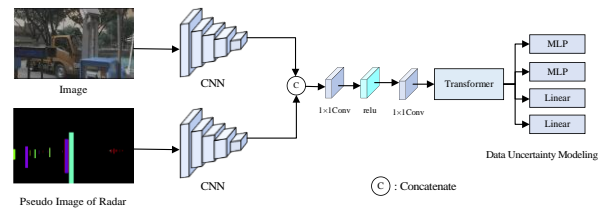


图7 结果不确定性估计在目标检测中的应用<sup>[22]</sup>

同样, YANG Fan等<sup>[23]</sup>提出一种方法, 使用直接建模的思想来估计纯视觉目标检测模型的不确定性, 此方法将概率表示模型与Transformer相结合, 在不确定条件下进行显式推理, 即不确定引导Transformer推理(UGTR), 用于伪装目标检测。其核心思想是首先学习中枢输出的条件分布, 得到初始估计和相关的

不确定性, 然后利用注意机制对这些不确定区域进行推理, 产生最终的预测。上述应用均是针对目标检测分类和边界框回归进行概率建模, 通常使用同样的概率分布, 例如单变量高斯分布, 拉普拉斯分布, 或者两者组合的分布。但是对于回归的概率建模, 上述方法忽略了回归变量之间的相关性, 并且可能无法完全反映边界框中的复杂不确定性, 特别是在物体被遮挡时。

基于此, PAN Hujie等<sup>[24]</sup>将回归变量转换回三维边界框的8个角的表示, 并分别直接学习每个角的不确定性。而HARAKEH等<sup>[25]</sup>学习一个具有完整协方差矩阵的多变量高斯分布, 用于回归变量。HE Yihui等<sup>[26]</sup>提出了一个更通用的概率分布, 通过使用多变量高斯混合来描述二维边界框, 考虑了相关性和多模态行为。通常通过最小化负对数似然(NLL)来训练具有随机不确定性的概率目标检测器。相比之下, 一些研究工作<sup>[27-29]</sup>引入了一个与真实边界框参数相关的先验分布, 并最小化了预测概率分布和先验分布之间的Kullback-Leibler散度(KLD)<sup>[30]</sup>。通过这种方式, 预测模型的不确定性时, 可以通过其先验分布进行规范化, 稳定训练该模型并提高检测性能。

综上所述, 使用直接建模的方式估计目标检测的偶然不确定性应用广泛, 其核心思想是假设预测结果服从概率分布, 使用神经网络来拟合这一概率分布, 通过训练得到概率模型, 使用此概率模型来

估计目标检测结果的偶然不确定性。使用直接建模的方式估计自动驾驶目标检测模型只需要一次前向推理就得到所需要的结果，因此，不占用网络额外推理时间，实时性较好。

### 2.2.2 认知不确定性问题研究方法

目标检测的认知不确定性多表现为目标检测神经网络的认知不确定性，由于目标检测神经网络所含参数的有限性和结构的局限性，会导致目标检测结果漏检错检。众多研究致力于通过 MC-Dropout 来获取目标检测中的不确定性。MILLER 等<sup>[31]</sup>对 SSD 检测器的后端进行了调整，以便在测试阶段通过 dropout 推理过程产生多个样本，并以此来估计分类的不确定性。在后续处理中，他们收集了多轮随机 dropout 推断的输出样本，并在每轮推断后独立应用 NMS。通过空间关联性，将多轮推断得到的冗余检测结果进行聚类，并利用特定公式为每个聚类分配一个对象的分类概率分布。他们还探讨了不同聚类技术对于分类不确定性质量的影响，并发现简单的聚类方法如 BSAS 在提供不确定性值方面优于复杂的匈牙利算法。在后续的研究中，MILLER 等<sup>[32]</sup>开发了基于 FasterRCNN<sup>[33]</sup>和 SSD<sup>[34]</sup>架构的认知不确定性估计框架，这些框架能在特定锚点位置生成多个样本，并通过计算 softmax 分类输出的平均值和熵来模拟认知不确定性。最终，通过标准的 NMS 操作合并了重复的检测结果。试验结果显示，在使用深度集成方法时，预测不确定性的估计在两个研究的架构上均优于单一模型的 MC-Dropout 方法<sup>[35]</sup>。FENG Di 等<sup>[36]</sup>则利用 MC-Dropout 和深度集成技术来评估分类的不确定性，并将这种不确定性应用于 3D 激光雷达目标检测器的主动学习训练中。仅在 3D 检测架构的检测头部使用 dropout 层，并在后处理阶段采用标准的 NMS 操作。ZHAO Yang 等<sup>[37]</sup>提出一种金字塔贝叶斯的方法对 MC-Dropout 进行了优化，并用于语义分割的不确定性研究。PENG Liang 等<sup>[38]</sup>利用 MC-Dropout 理论分析基于纯视觉的目标检测模型 YOLO 的边界框不确定性，建立了极端天气和不良照明条件下的道路图像数据集，在此基础上进行试

验，分析目标检测结果边界框不确定性和置信度之间的关系。

综上所述，使用 MC-Dropout 方法估计目标检测模型认知不确定性主要步骤如下。首先，在模型中加入或者利用模型中原有的 Dropout 层，其次，在推理阶段打开网络某层的 Dropout 层，最后，进行  $N$  次推理后得到  $N$  组相似的检测结果。在结果后处理阶段使用 NMS 进行相似结果过滤，得到最终结果后通过聚类算法得到相似的结果，最后每组相似的结果计算其均值、边界框方差、类别方差以及类别熵。使用 MC-Dropout 方法估计目标检测模型的认知不确定性需要多次前向推理，根据蒙特卡洛统计理论，当前向推理的次数趋近于无穷时，其检测结果均值趋向于真实值。在实际应用中，一般需要将前向推理次数控制在 30~50 次区间范围内。

## 3 不确定性评价指标

在目标检测领域存在许多评价指标，可以从不同的角度来反映或比较感知算法的性能。除了常用的检测精度指标外，学术界也开始对算法不确定性的评价指标展开研究。

### 3.1 精度、召回率和 F1 得分

在目标检测的背景下，仅考虑高于预定义的分类分数阈值  $\delta_{cls}$  的预测进行评估。它们进一步分为真阳性 (TP) 和误报 (FP)，基于它们与真值目标的交集 (IOU) 分数。如果检测到的检测值高于某个预定义的 IOU 阈值  $\delta_{iou}$ ，则当它们与真值目标共享相同的类标签时，它们将被视为真值。错误分类或 IOU 低于  $\delta_{iou}$  的检测将被视为误报。假阴性 (FN) 进一步定义为真值边界框，由于 IOU 分数较低，这些框要么丢失，要么与预测无关。有研究使用了 TP 和 FP 的计数来评估他们提出的概率目标检测器。TP 和 FP 评分也用于构建研究<sup>[39-40]</sup>中的 Receiver Operating Characteristic (ROC) 曲线。

### 3.2 平均精度 (mAP)

平均精度 (AP) 是由 EVERINGHAM 等<sup>[41]</sup>提出的，用于评估物体探测器的性能。它被定义为连续精度召回 (PR) 曲线下的面积，通过有限数量的

采样点上的数字积分近似<sup>[41]</sup>。平均 AP (mAP) 是为测试数据集中找到的每个目标类别计算的所有 AP 值的平均值。它也可以像文献 [42] 中提出的跨多个 IOU 阈值  $\delta_{iou} \in [0.5, \dots, 0.95]$  进行平均, 该度量称为 COCO mAP。虽然 mAP 是目标检测中的标准评估指标, 但它没有考虑预测的不确定性。事实上, 如果两个概率目标检测器预测出的边界框具有相同的平均值, 但协方差矩阵却截然不同, 那么它们将具有相同的平均精度均值 (mAP) 性能。关于概率物体检测的大多数工作<sup>[4, 28, 41, 43-46]</sup> 仍然使用 mAP 作为对其提出的方法进行定量评估的唯一指标, 强调在集成概率检测方法时精度提高的次要影响, 而不是关注输出分布的正确性。

### 3.3 基于概率的检测质量 (PDQ)

HALL 等<sup>[47]</sup> 提出了基于概率的检测质量 (PDQ) 作为衡量图像上 2D 概率目标检测质量的指标。PDQ 后来被一些研究<sup>[25, 32]</sup> 用于不确定性评估。PDQ 旨在共同评估基于图像的物体检测中的语义不确定性和空间不确定性。通过将预测的分类分数与图像中每个对象实例的真实标签相匹配来评估语义不确定性。空间不确定性由协方差矩阵编码, 假设边界框的右上角或左下角为高斯分布。当预测的概率与预测误差相关联时, 例如, 当较大的空间不确定性与不准确的边界框预测相关时, 可实现最佳 PDQ。

PDQ 使用匈牙利算法为每个基本事实分配最佳对应检测, 从而消除了 mAP 所需的对 IOU 阈值的依赖性。此外, PDQ 测量检测器分配给真阳性检测结果的概率质量, 并在单个分类分数阈值下进行评估, 要求被测模型在评估之前过滤低分输出检测结果。PDQ 假设 2D 高斯角分布, 并且无法评估假设拉普拉斯分布的方法。最后, 由于空间质量的定义方式, PDQ 只能针对在图像空间中定义的 2D 概率检测结果进行计算, 而对于 3D 概率目标检测器没有直接的扩展可用。

### 3.4 最小不确定度误差

不确定性误差 (UE) 最初是由 MILLER 等<sup>[39]</sup> 提出的, 用于评估概率对象检测器。UE 可以被认为是基于阈值的简单分类器在使用其预测的空间不

确定性估计值将输出检测分类为真阳性和误报时出错的概率。UE 的范围介于 0~0.5 之间, 当不确定性误差接近 0.5 时, 使用预测的不确定性估计值将真阳性与误报分开并不比随机分类器更好。检测器在所有可能的阈值上可达到的最佳不确定度误差称为最小不确定度误差 (MUE), 用于比较文献 [25] 和 [39] 中的概率目标检测器。

### 3.5 Jaccard IoU

WANG Zining 等<sup>[40]</sup> 提出 Jaccard IoU (JIoU) 作为 IoU 的概率推广。与仅比较两个边界框之间的确定性几何重叠的 IoU 不同, JIoU 测量其空间分布的相似性。这种分布可以通过概率对象检测网络或通过推断真实标签 (作为“概率分布的参考”) 中固有的不确定性来预测<sup>[40]</sup>。事实上, 当边界框假定遵循简单的均匀分布时, JIoU 简化为 IoU。与 IoU 类似, JIoU 范围在  $[0, 1]$  内。仅当两个边界框具有相同的位置、相同的范围以及相同的空间分布时, 才会最大化该值。通常, JIoU 提供了 IoU 的自然扩展来评估概率对象检测, 也考虑了真实标签中固有的模糊性和不确定性, 这些被其他评估指标 (如 mAP 和 PDQ) 所忽略。然而, 需要一个单独的模型来近似标记过程中的真实值空间不确定性, 这仅在 LiDAR 点云中提出<sup>[40, 48]</sup>。因此, 迄今为止, JIoU 的使用仅限于评估基于 LiDAR 的概率目标检测。

以上总结了目标识别里常用的评价指标, 以及不再仅根据确定性感知结果来评判感知算法性能, 而是考虑感知结果服从的概率分布。随着对不确定性研究的深入, 将会出现更多维度的评价指标。

## 4 总结与展望

在自动驾驶目标检测领域, 不确定性是一个不可忽视的问题, 它关系到系统的安全性和可靠性。本文综述了目标检测不确定性的多个方面, 包括其来源、量化方法以及如何利用不确定性信息提升系统性能。

### 4.1 总结

传感器的局限性、环境的复杂性以及动态场景的不可预测性是导致目标检测不确定性的主要原

因。这些因素导致检测结果的不稳定性，增加了自动驾驶系统在决策时的难度。已有多种方法被提出用来量化目标检测中的不确定性，包括基于统计的方法、模型驱动的方法和基于机器学习的技术。这些方法各有优势和局限性，选择合适的量化方法需要考虑具体的应用场景和系统要求。将不确定性信息整合到决策过程中，可以提高目标检测系统的鲁棒性。例如，通过集成不确定性估计的决策框架，系统能更精确地评估检测结果的可信度，从而作出更合理的决策。利用不确定性反馈来优化检测算法，可以进一步提升系统的性能。

#### 4.2 研究展望

未来的研究可以探索如何更有效地利用不确定

性信息来指导算法的改进。建立更加全面和具有挑战性的数据集，以更好地模拟真实世界中的不确定性情况，为算法训练和评估提供支持。将不同的不确定性量化方法和目标检测算法进行有效融合，以实现更准确和鲁棒的检测。在保证检测准确性的同时，提高算法的实时性，以满足自动驾驶系统对快速响应的需求。探索如何利用多种传感器数据（如摄像头、雷达、激光雷达等）来提高目标检测的准确性和鲁棒性。将不确定性估计运用到自动驾驶的其他技术领域中去，包括路径规划、行为预测等。同时，提高目标检测模型的可解释性，可以帮助研究人员和工程师更好地理解模型的行为，从而更有效地预防和处理不确定性。

#### 参考文献 (References)

- [1] WANG Ke, WANG Yong, LIU Bingjun, et al. Quantification of Uncertainty and Its Applications to Complex Domain for Autonomous Vehicles Perception System [J]. IEEE Transactions on Instrumentation and Measurement, 2023, 72: 1-17.
- [2] PENG Liang, LI Boqi, YU Wenhao, et al. SOTIF Entropy: Online SOTIF Risk Quantification and Mitigation for Autonomous Driving [J]. IEEE Transactions on Intelligent Transportation Systems, 2024, 25(2): 1530-1546.
- [3] YANG Kai, TANG Xiaolin, LI Jun, et al. Uncertainties in Onboard Algorithms for Autonomous Vehicles: Challenges, Mitigation, and Perspectives [J]. IEEE Transactions on Intelligent Transportation Systems, 2023, 24(9): 8963-8987.
- [4] FENG Di, CAO Yifan, ROSENBAUM L, et al. Leveraging Uncertainties for Deep Multi-Modal Object Detection in Autonomous Driving [C]//IEEE Intelligent Vehicles Symposium (IV), Oct. 19-Nov. 13, 2020, Las Vegas, NV, USA. Piscataway NJ: IEEE, c2020: 877-884.
- [5] AMINI A, SCHWARTING W, SOLEIMANY A, et al. Deep Evidential Regression [Z]. arXiv: 1910.02600, 2019.
- [6] GAL Y. Uncertainty in Deep Learning [D]. Cambridge: University of Cambridge, 2016.
- [7] MICHELMORE R, KWIATKOWSKA M, GAL Y. Evaluating Uncertainty Quantification in End-to-End Autonomous Driving Control [Z]. arXiv: 1811.06817, 2018.
- [8] TAI Lei, YUN Peng, CHEN Yuying, et al. End-to-End Driving Deploying through Uncertainty-Aware Imitation Learning and Stochastic Visual Domain Adaptation [Z]. arXiv: 1903.00821, 2019.
- [9] MICHELMORE R, WICKER M, LAURENTI L, et al. Uncertainty Quantification with Statistical Guarantees in End-to-End Autonomous Driving Control [C]. 2020 IEEE International Conference on Robotics and Automation (ICRA), May 31-Aug. 31, 2020, Paris, France. Piscataway NJ: IEEE, c2020: 7344-7350.
- [10] GAL Y, HRON J, KENDALL A. Concrete Dropout [C]//Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS). 2017: 3584-3593.
- [11] LAKSHMINARAYANAN B, PRITZEL A, BLUNDELL C. Simple and Scalable Predictive Uncertainty Estimation Using Deep Ensembles [Z]. arXiv: 1612.01474, 2017.
- [12] OVADIA Y, FERTIG E, REN Jie, et al. Can You Trust Your Model's Uncertainty? Evaluating Predictive Uncertainty Under Dataset Shift [C]//Proceedings of the 33rd International Conference on Neural Information Processing Systems (NIPS), 2019: 14003-14014.
- [13] BLUNDELL C, CORNEBISE J, KAVUKCUOGLU K, et al. Weight Uncertainty in Neural Networks [C]//Proceedings of the 32nd International Conference on International Conference on Machine Learning. 2015, 37: 1613-1622.
- [14] GRAVES A. Practical Variational Inference for Neural Networks [C]//Proceedings of the 24th International

- Conference on Neural Information Processing Systems (NIPS).2011:2348–2356.
- [15] SENSOY M, KAPLAN L, KANDEMIR M. Evidential Deep Learning to Quantify Classification Uncertainty [C]//Proceedings of the 32nd International Conference on Neural Information Processing Systems(NIPS).2018: 3183–3193.
- [16] MALININ A, GALES M. Predictive Uncertainty Estimation Via Prior Networks [C]//Proceedings of the 32nd International Conference on Neural Information Processing Systems (NIPS).2018:7047–7058.
- [17] AMINI A, SCHWARTING W, SOLEIMANY A, et al. Deep Evidential Regression[Z].arXiv:1910.02600,2019.
- [18] LIU Xixi, LIN C T, ZACH C. Energy-Based Models for Deep Probabilistic Regression [C]//2022 26th International Conference on Pattern Recognition (ICPR), Aug. 21–25, 2022, Montreal, QC, Canada. Piscataway NJ: IEEE, c2022:2693–2699.
- [19] KULESHOV V, FENNER N, ERMON S. Accurate Uncertainties for Deep Learning Using Calibrated Regression [C]//Proceedings of the 35th International Conference on Machine Learning (ICML). 2018: 2796–2804.
- [20] POSTELS J, FERRONI F, COSKUN H, et al. Sampling-Free Epistemic Uncertainty Estimation Using Approximated Variance Propagation [C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), Oct. 27, Seoul, Korea (South). Piscataway NJ: IEEE, c2019:2931–2940.
- [21] SHEENY M, DE PELLEGRIN E, MUKHERJEE S, et al. RADIATE: A Radar Dataset for Automotive Perception in Bad Weather [C]//2021 IEEE International Conference on Robotics and Automation (ICRA), May 30–June 5, Xi'an, China. Piscataway NJ: IEEE, c2021:1–7.
- [22] ZHAO Yang, WANG Xiao, CHENG Hong. Radar-Image Fusion Transformer for Object Detection with Uncertainty Estimation [C]//2023 IEEE 26th International Conference on Intelligent Transportation Systems(ITSC), Sept. 24–28, Bilbao, Spain. Piscataway NJ: IEEE, c2023: 814–819.
- [23] YANG Fan, ZHAI Qiang, LI Xin, et al. Uncertainty-Guided Transformer Reasoning for Camouflaged Object Detection [C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV), Oct. 10–17, Montreal, QC, Canada. Piscataway NJ: IEEE, c2021:4126–4135.
- [24] PAN Hujie, WANG Zining, ZHAN Wei, et al. Towards Better Performance and More Explainable Uncertainty for 3D Object Detection of Autonomous Vehicles [C]//2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC), Sept. 20–23, 2020, Rhodes, Greece. Piscataway NJ: IEEE, c2020:1–7.
- [25] HARAKEH A, SMART M, WASLANDER S. BayesOD: A Bayesian Approach for Uncertainty Estimation in Deep Object Detectors [C]//2020 IEEE International Conference on Robotics and Automation (ICRA), May 31–Aug. 31, 2020, Paris, France. Piscataway NJ: IEEE, c2020:87–93.
- [26] HE Yihui, WANG Jianren. Deep Mixture Density Network for Probabilistic Object Detection [C]//2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Oct. 24, 2020–Jan. 24, 2021, Las Vegas, NV, USA. Piscataway NJ: IEEE, c2020: 10550–10555.
- [27] FENG Di, ROSENBAUM L, TIMM F, et al. Labels are Not Perfect: Improving Probabilistic Object Detection via Label Uncertainty[Z].arXiv:2008.04168,2020.
- [28] HE Yihui, ZHU Chenchen, WANG Jianren, et al. Bounding Box Regression with Uncertainty for Accurate Object Detection [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15–20, 2019, Long Beach, CA, USA. Piscataway NJ: IEEE, c2019:2888–2897.
- [29] MEYER G P, THAKURDESAI N. Learning an Uncertainty-Aware Object Detector for Autonomous Driving[Z].arXiv:1910.11375,2019.
- [30] KULLBACK S, LEIBLER R A. On Information and Sufficiency [J]. The Annals of Mathematical Statistics, 1951, 22(1):79–86.
- [31] MILLER D, NICHOLSON L, DAYOUB F, et al. Dropout Sampling for Robust Object Detection in Open-Set Conditions [C]//2018 IEEE International Conference on Robotics and Automation (ICRA), May 21–25, 2018, Brisbane, QLD, Australia. Piscataway NJ: IEEE, c2018: 3243–3249.
- [32] MILLER D, SÜNDERHAUF N, ZHANG Haoyang, et al. Bench Marking Sampling-Based Probabilistic Object Detectors [C]//2019 IEEE/CVF IEEE Conference on Computer Vision and Pattern Recognition Workshops.2019: 1–6.
- [33] REN Shaoqing, HE Kaiming, GIRSHICK R, et al. Faster R-CNN: Towards Realtime Object Detection with Region Proposal Networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137–1149.

- [34] LIU Wei, ANGUELOV D, ERHAN D, et al. SSD: Single Shot Multibox Detector [C]//European Conference on Computer Vision.2016:21–37.
- [35] ZHANG Zheng, LIU Qi, LI Yanjie, et al. Safe Reinforcement Learning in Autonomous Driving with Epistemic Uncertainty Estimation [J]. IEEE Transactions on Intelligent Transportation Systems (Early Access), 2024: 1–14.
- [36] FENG Di, WEI Xiao, ROSENBAUM L, et al. Deep Active Learning for Efficient Training of a Lidar 3D Object Detector [C]//2019 IEEE Intelligent Vehicles Symposium, June 6–12, 2019, Paris, France. Piscataway NJ:IEEE, c2019:667–674.
- [37] ZHAO Yang, TIAN Wei, CHENG Hong. Pyramid Bayesian Method for Model Uncertainty Evaluation of Semantic Segmentation in Autonomous Driving [J]. Automotive Innovation, 2022, 5(1):70–78.
- [38] PENG Liang, WANG Hong, LI Jun. Uncertainty Evaluation of Object Detection Algorithms for Autonomous Vehicles [J]. Automotive Innovation, 2021, 4(3):241–252.
- [39] MILLER D, DAYOUB F, MILFORD M, et al. Evaluating Merging Strategies for Sampling-Based Uncertainty Techniques in Object Detection [C]//2019 IEEE International Conference on Robotics and Automation, May 20–24, 2019, Montreal, QC, Canada. Piscataway NJ: IEEE, c2019:2348–2354.
- [40] WANG Zining, FENG Di, ZHOU Yiyang, et al. Inferring Spatial Uncertainty in 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, c2020:5792–5799.
- [41] EVERINGHAM M, VAN GOOL L, WILLIAMS C K I, et al. The Pascal Visual Object Classes (VOC) Challenge [J]. International Journal of Computer Vision, 2009, 88(2):303–338.
- [42] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: Common Objects in Context [C]//European Conference on Computer Vision.2014:740–755.
- [43] FENG Di, ROSENBAUM L, TIMM F, et al. Leveraging Heteroscedastic Aleatoric Uncertainties for Robust Real-Time Lidar 3D Object Detection [C]//2019 IEEE Intelligent Vehicles Symposium, June 9–12, 2019, Paris, France. Piscataway NJ:IEEE, c2019:1280–1287.
- [44] CHOI J W, CHUN D Y, KIM H, et al. Gaussian Yolov3: An Accurate and Fast Object Detector Using Localization Uncertainty for Autonomous Driving [C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), Oct. 27–Nov. 2, 2019, Seoul, Korea (South). Piscataway NJ:IEEE, c2019:502–511.
- [45] DONG Xu, WANG Pengluo, ZHANG Pengyue, et al. Probabilistic Oriented Object Detection in Automotive Radar [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), June 14–19, 2020, Seattle, WA, USA. Piscataway NJ: IEEE, c2020:102–103.
- [46] LEE Y W, HWANG J W, KIM H I, et al. Localization Uncertainty Estimation for Anchor-Free Object Detection [Z]. arXiv:2006.15607, 2020.
- [47] HALL D, DAYOUB F, SKINNER J, et al. Probabilistic Object Detection: Definition and Evaluation [C]//2020 IEEE Winter Conference on Applications of Computer Vision (WACV), Mar. 1–5, 2020, Snowmass, CO, USA. Piscataway NJ:IEEE, c2020:1031–1040.
- [48] FENG Di, WANG Zining, ZHOU Yiyang, et al. Labels are Not Perfect: Inferring Spatial Uncertainty in Object Detection [J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(8):9981–9994.

## 作者简介



王潇 (2000–), 女, 甘肃白银人, 硕士研究生, 主要研究方向为智能网联汽车预期功能安全评估。

Tel: 028-61830797

E-mail: m17609494597@163.com

## 通信作者



赵洋 (1988–), 男, 山西运城人, 博士, 助理研究员, 主要研究方向为智能网联汽车感知与决策。

Tel: 028-61830797

E-mail: yzhao@uestc.edu.cn