



北京大学

本科生毕业论文

题目： 三维语义分割的不确定性问题实验分析

The Uncertainty Experimental Analysis in

3D Semantic Segmentation

姓 名： 周昊天

学 号： 1900011032

院 系： 工学院

专 业： 机器人工程

导师姓名： 赵卉菁 教授

刘畅 助理教授

二〇二三 年 五 月

北京大学本科毕业论文导师评阅表

学生姓名	周昊天	学号	1900011032
院系	工学院	专业	机器人工程
指导教师	赵卉菁 刘畅	职称	教授 助理教授
毕业论文题目	三维语义分割的不确定性问题实验分析		
导师是否同意参加 毕业论文答辩	是	建 议 成 绩 (可选填)	
导师评语	基于深度学习的三维点云语义分割是自动驾驶环境感知的关键技术。深度模型的不确定性问题给高安全性需求的自动驾驶应用带来了挑战。论文通过基于集成模型的不确定性估计算法，探究了具有代表性的三维点云语义分割模型的两类不确定性问题。论文实验分析了模型间分割性能的差异、不同类别的混淆性、不确定性与模型精度的相关性，反映了实验方法的有效性。论文表现出作者具有一定的理论基础与系统的学科知识，并具有一定的从事科研工作的能力，达到了本科学位论文的要求。 导师签名：_____年__月__日		

版权声明

任何收存和保管本论文各种版本的单位和个人，未经本论文作者同意，不得将本论文转借他人，亦不得随意复制、抄录、拍照或以任何方式传播。否则一旦引起有碍作者著作权之问题，将可能承担法律责任。

摘要

自动驾驶是智能机器人研究的重要课题，环境感知技术是自动驾驶的关键。三维点云语义分割是环境感知中的重要任务。目前三维点云语义分割的主流方法使用深度神经网络模型。然而，深度学习的不确定性问题给高安全性需求的自动驾驶实际应用带来了挑战。

本文旨在探究三维点云语义分割模型的不确定性问题。首先，本文对深度学习的不确定性问题展开文献调研，梳理了深度学习中的模型与数据不确定性及相关建模评估方法。然后，本文调研三维点云语义分割算法、数据集及评价指标。调研发现目前的不确定性评估方式仍存在模型与数据不确定性高度耦合的现象，实际应用中缺乏探究三维点云语义分割模型不确定性的研究工作。

本文进一步对三维点云语义分割的不确定性进行了实验分析。本文选取了注重分割性能的代表模型 Cylinder3D 和注重分割效率的代表模型 SqueezeSegV3。在点云分割任务表现上，本文发现相比 SqueezeSegV3，Cylinder3D 性能的提升主要出现在数据量小的类别；此外，植物类别容易被混淆，信号类别容易被错分。本文选取基于集成模型的不确定性估计方法，对两个模型进行了不确定性分析实验。实验结果表明，基于集成的不确定性度量方式能一定程度反映模型对输出的确信程度。模型不确定性与数据不确定性在数量级上有较大差距，二者有较高的相关性。实验结果表明，具有较低不确定性的类别分割精度高，总体不确定性与模型精度呈负相关趋势。最后，本文进一步探究了不同数据域分布上的不确定性表现，发现受数据分布变化，总体不确定性增加，真实精度明显下降。

关键词：三维点云语义分割，不确定性估计，集成模型，随机失活，贝叶斯深度学习

ABSTRACT

Autonomous driving is an important research topic for intelligent robots, and environmental sensing technology is the key to autonomous driving. 3D point cloud semantic segmentation is an important task in environmental sensing. Current mainstream approach in 3D point cloud semantic segmentation uses deep neural network models. However, the uncertainty problem of deep learning brings challenges to the practical application of high safety-critical autonomous driving. This article aims to explore the uncertainty problem of 3D point cloud semantic segmentation models. Firstly, this article conducts a literature review on the uncertainty problem of deep learning, and summarizes the modeling and evaluation methods of model and data uncertainty in deep learning. Then, this article investigates 3D point cloud semantic segmentation algorithms, datasets, and evaluation metrics. The investigation finds that there is still a high correlation between model and data uncertainty in the current uncertainty evaluation methods, and there is a lack of research on exploring the uncertainty of 3D point cloud semantic segmentation models in practical applications.

This thesis further analyzes the uncertainty of 3D point cloud semantic segmentation experimentally. This thesis selects a representative model for segmentation performance, Cylinder3D, and a representative model for segmentation efficiency, SqueezeSegV3. In terms of point cloud segmentation performance, this article found that compared with SqueezeSegV3, the improvement in performance of Cylinder3D mainly occurs in categories with small data volumes. In addition, in terms of categories, plants are easily confused, and signs are easily misclassified. This thesis selects the uncertainty estimation method based on ensemble to analyze the uncertainty of above two models. The experimental results show that the ensemble-based uncertainty measurement method can to some extent reflect the model's confidence in the output. Model uncertainty and data uncertainty have a large difference in magnitude, and the two have a high correlation. The experimental results show that categories with low uncertainty have high segmentation accuracy, and the total uncertainty is negatively correlated with model accuracy. Finally, this article further explores the uncertainty performance on different data domain distributions, and finds that with the change of data distribution, total uncertainty increases and real accuracy decreases significantly.

KEY WORDS: 3D Point Cloud Semantic Segmentation, Uncertainty Quantification, Ensemble, Dropout, Bayesian Deep Learning

目 录

第一章 绪论	1
1.1 研究背景与意义	1
1.2 研究现状与分析	2
1.3 研究内容与论文构成	3
第二章 深度学习不确定性综述	4
2.1 贝叶斯模型的不确定性量化	4
2.1.1 贝叶斯建模	4
2.1.2 变分推断	5
2.1.3 其它近似推断方法及面临的挑战	6
2.2 深度学习中的两类不确定性及其建模	7
2.2.1 数据不确定性	7
2.2.2 模型不确定性	9
2.3 深度学习中的不确定性估计	9
2.3.1 蒙特卡洛随机失活	9
2.3.2 分支网络估计	10
2.3.3 集成模型估计	13
2.4 深度学习的模型置信度	14
2.4.1 深度学习模型置信度的定义	14
2.4.2 置信度的校准	14
2.4.3 置信度与不确定性关系	15
第三章 三维点云语义分割任务介绍	16
3.1 问题定义	16
3.2 三维点云语义分割算法	16
3.2.1 基于图像的算法	16
3.2.2 基于体素的算法	18
3.2.3 基于点云的算法	19
3.3 经典数据集	20
3.4 评价指标	20

第四章	三维点云语义分割任务中的不确定性估计实验	22
4.1	任务目标	22
4.2	算法简述	22
4.3	实现细节	22
4.4	实验结果	25
4.4.1	点云分割任务表现	25
4.4.2	三维点云语义分割类别混淆性分析	26
4.4.3	训练过程中的不确定性变化	27
4.4.4	各类别的不确定性估计	30
4.4.5	三维点云不确定性可视化	31
4.4.6	不同数据域分布上的不确定性表现	33
第五章	总结与展望	35
5.1	总结	35
5.2	展望	35
参考文献		37
致谢		42
北京大学学位论文原创性声明和使用授权说明		44

第一章 绪论

1.1 研究背景与意义

不确定性是一个在多个学科中出现的概念，它指的是事先不能准确知道某个事件或某个决策的结果。

在深度学习被广泛使用前，人们对不确定性就已经有了一定的研究。在信息论^[1]、高斯过程^[2]、贝叶斯学习^[3]等领域中已经对不确定性进行了概率建模与探讨。

在越来越多的领域中，基于深度学习（Deep Learning-Based）的方法相比于二十年前基于规则设计（Rule-Based）的方法表现出无法想象的优越性能。近年来随着大模型在语言^[4]、视觉^[5]、游戏^[6]领域表现出令人惊叹的性能，深度学习的可行性被反复证明，并被越来越多的人所接受与使用，成为智能科学领域的基本技术。

但一切模型都有自己的适用条件。一些高度安全相关的系统要求知道模型在哪些情况下可能存在风险，以采取应急措施确保安全。遗憾的是，只要有同样的输入形式，一般的深度网络就有对应的输出，且只是一个确定的值。使用者无法确定网络究竟在随意猜测还是在输出有意义的结果。这在高度安全相关的领域是不可接受的。例如 2016 年特斯拉辅助驾驶系统发生的事故^[7]。高速公路上出现了一辆侧翻的白色卡车，显然这样的数据极为罕见，模型训练时没有考虑到。辅助驾驶系统的感知模块将白色卡车识别为天空并加速撞向了白色卡车，如图1.1蓝色实线对应流程。

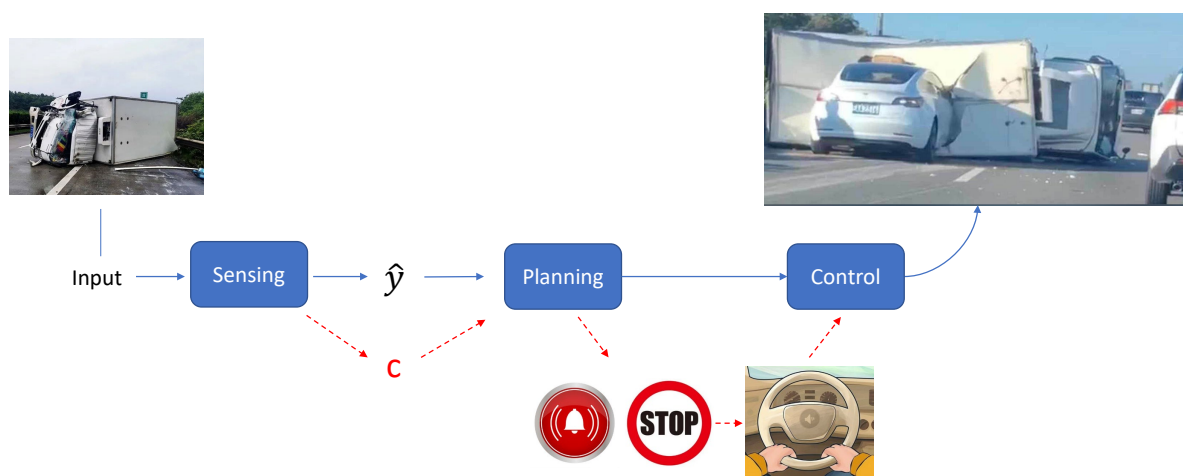


图 1.1 将不确定性引入自动驾驶的示意图

深度学习的不确定性度量就是想建模模型本身对自己输出的确信程度。如图1.1红

色虚线部分。如果感知模块在输出定值 $\hat{g}$ 的同时，输出关于这一数据或者模型本身的不确定性 c （在2.2节中，将深度学习的不确定性分为数据和模型的不确定性，这里统一记为 c ）。下游规划（Planning）模块发现感知（Sensing）模块出现了较大的不确定性，可能面临较大的危险，便发出警报采取保守措施或交由驾驶员紧急接管控制（Control）模块。以上只是一种简单的可能的不确定性应用方式，在上游模块得到了不确定性度量后，下游模块可以进行多样的处理以提升系统的鲁棒性。

自动驾驶是高度安全相关的任务，三维点云语义分割作为自动驾驶的重要感知任务，先进的（State of the Art）算法主要基于深度学习。因为数据分布的长尾性，我们不能期望所有将要见到的东西都已经见到过，即真实驾驶环境中很可能出现训练集中没有的数据。譬如侧翻的白色卡车，飘在天空的垃圾袋等。因此，我们想在获得分割结果的同时，获得模型输出的不确定性，利用这一不确定性采取更安全的策略或者进一步收集有价值的信息。

1.2 研究现状及分析

如2.1节所述，贝叶斯理论用严谨的数学语言对不确定性进行了阐释与建模。随着深度学习的发展，不少工作基于近似推断的思想将不确定性估计融入神经网络^[8-9]。随机失活（Dropout）作为贝叶斯理论与神经网络的桥梁，使不确定性在深度学习中的应用变得简便与实际^[10-11]。

但目前，在深度学习中，有较严谨的理论支撑的不确定性估计应用仍局限于简单的神经网络模型。复杂的现实任务，如三维点云语义分割对不确定性的探讨较少。

本文将不确定性估计方法应用到三维点云语义分割问题中。目前的三维点云语义分割算法主要基于深度学习，大多数方法关注最终的分割精度。据我所知，只有 SalsaNext^[12]设计融合了基于 ADF^[13]的不确定性度量方法，单独实现了一个可以输出数据不确定性和模型不确定性的部署版本。但这一版本的数据不确定性设计存在问题（详见2.3.2节）。SalsaNext 部署了可以输出不确定性和不输出不确定性的两个版本。文章的探讨重点和主要贡献还在于不具有不确定性输出的部署版本。除此之外，有很多将模型置信度作为总的预测不确定性估计的工作，通过置信度相关方法研究测试与训练数据不同分布（OOD）的问题。这类工作没有在本文详细调研。置信度相关工作以及与不确定性的关联与区别将在2.4节介绍。

总而言之，不确定性本身难以度量、没有真值，缺少统一的量化指标；现有不确定性方法难以应用到复杂的模型；三维点云本身相较于二维图片更难直观理解与可视化。虽然三维点云语义分割任务是自动驾驶的重要感知任务，但缺乏关于不确定性估计的研究。

1.3 研究内容与论文构成

本文针对三维点云语义分割任务的不确定性估计问题进行调研与实验分析。

如图1.2，本文主要内容分为五章展开。

第一章绪论。主要限定了本文探讨的不确定性范围，以及不确定性问题的意义、现状。

第二章深度学习不确定性综述。本章为调研性质工作，主要介绍了贝叶斯理论建模的不确定性，深度学习中的两类不确定性，不确定性估计方法以及模型置信度和不确定性的关联。

第三章三维点云语义分割任务介绍。本章为调研性质工作，主要介绍了三维点云语义分割任务的输入输出，对现有三维点云语义分割算法按输入数据形式进行分类调研，简述经典数据集和评价指标。

第四章三维点云语义分割任务中的不确定性估计实验。本章为实验性质工作，主要介绍了实验的目标，采用的算法，部署的细节，对三维点云语义分割模型的效果以及不确定性估计方法的有效性进行了实验。

第五章总结与展望。总结了在调研与实验中得出的结论，当前算法的局限与未来可以进一步探究的方向。

最后附有本文的参考文献与致谢等内容。

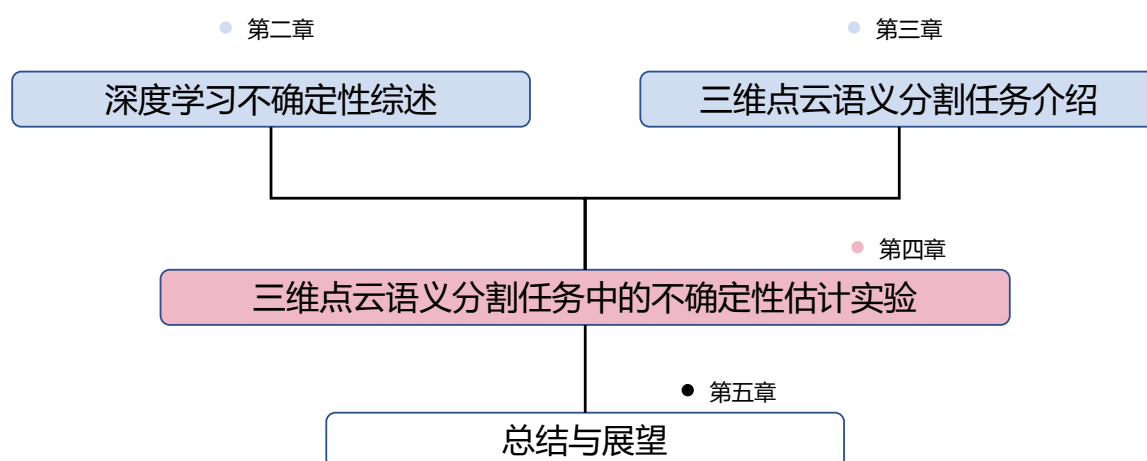


图 1.2 论文的结构图示

其中蓝色部分章节为调研性工作，红色部分章节为实验性工作

第二章 深度学习不确定性综述

本章的内容从贝叶斯建模开始，引入深度学习的两类不确定性物理概念与信息论建模，介绍深度学习中常用的两类不确定性估计方法，并比较置信度与不确定性的关联。

2.1 贝叶斯模型的不确定性量化

深度学习的不确定性建模主要分为贝叶斯方法与非贝叶斯方法。早期没有高性能的计算平台，更多的研究从贝叶斯建模的角度出发。本节主要参考 Gal 博士论文^[14]关于不确定性的贝叶斯建模。

2.1.1 贝叶斯建模

给定训练输入 $X = \{x_1, x_2, \dots, x_N\}$; 对应真值 $Y = \{y_1, y_2, \dots, y_N\}$ 。为更好理解，以下简记输入数据为 $\mathbf{x}$ ，真值为 $\mathbf{y}$ 。

在贝叶斯回归问题中，预测模型为 $\hat{\mathbf{y}} = f^\omega(\mathbf{x})$ ，贝叶斯后验学习的目标是找到参数 ω ，使 $\hat{\mathbf{y}}$ 尽可能接近真值 $\mathbf{y}$ 。

贝叶斯建模的核心就是，这一构成模型的参数 ω 本身也服从一个分布。在输入训练数据前，我们已经对 ω 有一个先验（Prior）的估计分布， $p(\omega)$ 。

似然（Likelihood）是贝叶斯理论的核心概念。似然函数本意是给定 ω （此时不是分布，可以理解为某个采样点）和输入数据 $\mathbf{x}$ 得到真值 $\mathbf{y}$ 的概率，变量是模型参数 ω 。可以将对预测分布的先验融入到似然函数的表达式之中。极大似然就是在已有的似然函数表达式中，代入数据集真值，寻找最大化这一概率的模型参数 ω 。

在回归问题中，可以采用高斯似然函数：

$$p(\mathbf{y}|\mathbf{x}, \omega) = \mathcal{N}(\mathbf{y}; f^\omega(\mathbf{x}), \tau^{-1}I) \quad (2.1)$$

其中, τ 为模型精度，也可以理解为在原本输出上加了 τ^{-1} 的方差。

在分类问题中，通常采用软最大（Softmax）似然函数，即：

$$p(y = d|\mathbf{x}, \omega) = \frac{\exp(f_d^\omega(\mathbf{x}))}{\sum_{d'} \exp(f_{d'}^\omega(\mathbf{x}))} \quad (2.2)$$

其中 $f_d^\omega(\mathbf{x})$ 是模型对类别 d 的输出，软最大化后保证各类别输出的 0 – 1 之间的值，且全部类别的输出和为 1，可以理解为概率。

总而言之，似然函数将原本 $f^\omega(\mathbf{x})$ 的点估计变成了一种概率分布。

贝叶斯建模的两个任务分别是：

1、后验（学习）：通过不断输入的数据，得到经过修正的模型参数 ω ，换言之，得到后验分布 $p(\omega|\mathbf{x}, \mathbf{y})$ 。此处的 $\mathbf{x}, \mathbf{y}$ 理解为整个数据集的输入与输出。

此时，可以利用贝叶斯公式：

$$p(\omega|\mathbf{x}, \mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x}, \omega)p(\omega)}{p(\mathbf{y}|\mathbf{x})} \quad (2.3)$$

其中， $p(\mathbf{y}|\mathbf{x}, \omega)$ 是似然函数， $p(\omega)$ 是先验分布， $p(\mathbf{y}|\mathbf{x})$ 是似然对模型参数的积分，也称为证据因子（evidence），即：

$$p(\mathbf{y}|\mathbf{x}) = \int p(\mathbf{y}|\mathbf{x}, \omega)p(\omega)d\omega \quad (2.4)$$

2、推断（预测）：通过已有的对模型参数 ω 的估计，实现对新输入样本 $\mathbf{x}^*$ 的预测 $p(\mathbf{y}^*|\mathbf{x}^*, \mathbf{x}, \mathbf{y})$ 。

结合似然与后验分布，对模型参数积分，得到预测概率分布：

$$p(\mathbf{y}^*|\mathbf{x}^*, \mathbf{x}, \mathbf{y}) = \int p(\mathbf{y}^*|\mathbf{x}^*, \omega)p(\omega|\mathbf{x}, \mathbf{y})d\omega \quad (2.5)$$

由于贝叶斯模型的参数与预测输出天然概率形式，其模型的不确定性可以通过方差或者熵定义，如： $\text{Var}(p(\omega|\mathbf{x}, \mathbf{y}))$ ，针对每一个预测数据点的不确定性也可以类似定义，如： $\text{Var}(p(\mathbf{y}^*|\mathbf{x}^*, \mathbf{x}, \mathbf{y}))$ 。

2.1.2 变分推断

贝叶斯建模的理论非常直观和直接，在特殊的贝叶斯线性回归问题中，先验分布和似然函数都是高斯分布，共轭。证据因子 $p(\mathbf{y}|\mathbf{x})$ 可解析求解，后验和预测都能顺利得到解析形式。

然而，实际的大多数情况是，证据因子这一对 ω 的积分无法解析计算，导致无法得到后验分布解析形式。此时，我们试图寻找一个替代的有解析形式的后验分布 $q_\theta(\omega)$ ，并希望其与真实的后验分布 $p(\omega|\mathbf{x}, \mathbf{y})$ 足够接近。通过替代的后验分布，近似计算预测分布：

$$p(\mathbf{y}^*|\mathbf{x}^*, \mathbf{x}, \mathbf{y}) \approx \int p(\mathbf{y}^*|\mathbf{x}^*, \omega)q_\theta^*(\omega)d\omega \quad (2.6)$$

这一过程就是变分推断的思路。

变分推断通过 KL 散度来衡量近似分布与真实后验分布的距离：

$$KL(q_\theta(\omega)||p(\omega|\mathbf{x}, \mathbf{y})) = \int q_\theta(\omega) \log \frac{q_\theta(\omega)}{p(\omega|\mathbf{x}, \mathbf{y})} d\omega \quad (2.7)$$

证据因子下界 (ELBO, Evidence Lower BOund) 定义为:

$$ELBO = \log(\mathbf{y}|\mathbf{x}) - KL(q_{\theta}(\omega)||p(\omega|\mathbf{x}, \mathbf{y})) = \int q_{\theta}(\omega) \log \frac{p(\mathbf{y}, \omega|\mathbf{x})}{q_{\theta}(\omega)} d\omega \quad (2.8)$$

因为 KL 散度恒正, 所以 ELBO 一定不大于证据因子 $\log p(\mathbf{y}|\mathbf{x})$, 这也是其称为 ELBO (证据因子下界) 的原因。

在部分工作中, ELBO 也称作自由能损失函数。因为 ELBO 的第一部分与 θ 无关, 优化 θ 以最小化 KL 散度等价于优化 θ 以最大化 ELBO。

经过贝叶斯公式的变换, ELBO 的每一项都可以解析表示:

$$\begin{aligned} ELBO &= E_{q_{\theta}(\omega)}[\log P(\mathbf{y}|\mathbf{x}, \omega)] + E_{q_{\theta}(\omega)}[\log P(\omega)] - E_{q_{\theta}(\omega)}[\log q_{\theta}(\omega)] \\ &= E_{q_{\theta}(\omega)}[\log p(\mathbf{y}|\mathbf{x}, \omega)] - KL(q_{\theta}(\omega)||p(\omega)) \end{aligned} \quad (2.9)$$

下一步, 最大化 ELBO, 可以通过平均场、随机梯度等理论。因为较繁琐且与本文相关性不大, 不再深入介绍。

变分推断较系统的介绍与推导可以参看综述^[15], 其详细介绍了包括马尔科夫链蒙特卡洛采样 (MCMC, Markov Chain Monte-Carlo)、混合高斯 (Mixture of Gaussians)、证据因子下界 (ELBO)、平均场 (Mean-field)、坐标上升 (Coordinate Ascent) 在内的与变分推断相关的理论。

变分推断仍存在问题: 其需要假定先验分布形式, 还需要给定近似分布的形式, 这样可能导致得到的近似分布与真实后验分布有较大的偏差。另外, 后续积分不可解析仍需借助采样近似。

2.1.3 其它近似推断方法及面临的挑战

除了估计后验分布的方法, 近似推断还包括基于采样的算法, 如马尔科夫链蒙特卡洛算法 (MCMC), 这一算法可以用于不确定性的估计^[16-17]。

现代的推断方法将贝叶斯概率建模与神经网络融合, 提出了贝叶斯神经网络的优化方式。BBB^[9]假定每个神经元服从高斯分布或者混合高斯分布, 通过贝叶斯反向传播 (Bayes by Backprop), 使得每个神经元的权重不再是定值, 而是一个概率分布, PBP^[18]也采用了类似的思想, 通过高斯分布的性质对网络进行优化; 除了优化变分推断的自由能损失函数, 也有工作优化真实后验分布和替代分布的 α 散度^[19-20]。

包括变分推断在内, 这类方法不适用于大数据大模型, 并且理论推导过于复杂, 我们希望使用满足以下条件的不确定性估计方法^[14]:

- 1、可以很好地适用于大数据集
- 2、适用于复杂的模型

- 3、不需要改变现有的模型构架
- 4、可以被非专业人士简单的使用与理解

2.2 深度学习中的两类不确定性及其建模

在经典的贝叶斯理论中，由于后验分布与预测分布都是概率分布，能比较容易的估计不确定性。不过经典贝叶斯理论重点关注模型不确定性，没有对其它不确定性的来源进行建模^[14]。

在深度学习的不确定性度量中，一般认为总不确定性 (Total Uncertainty)，又称预测不确定性 (Predictive Uncertainty)，由模型不确定性 (Model Uncertainty) 与数据不确定性 (Data Uncertainty) 两部分构成^[21-23]。

以下是理想的两类不确定性概念。在实际估计中，数据不确定性由模型估计得到，往往不可避免的夹杂了模型的特性；估计模型不确定性时，也往往排除不了数据的特性。

2.2.1 数据不确定性

数据不确定性 (Data Uncertainty)，又叫偶然不确定性 (Aleatoric Uncertainty)，定义为观测中的内在的噪声，无法通过更多的数据消除^[21-23]。数据不确定性产生原因是观测误差，如计算机视觉中的传感器噪声、物体遮挡等；概念模糊，如定义类别时的重叠，类别间的相似性，类别内的多样性等；标注误差，如标注数据时的不精确，真值本身的不确定等。

数据不确定性还可以进一步细分为同方差的 (Homoscedastic) 和异方差的 (Heteroscedastic)^[21, 24]，区别在于针对不同的数据是否有不一样的不确定性，一般在不确定性的研究中假设数据不确定性是异方差的，也即针对不同输入会有不同的数据不确定性。

在分类问题中，考虑输入数据 $\mathbf{x} \in \mathcal{R}^D$ ，真实类别标签 $y \in \{\omega_1, \dots, \omega_K\}$ ，我们认为训练数据 $\mathcal{D}_{tr}$ 是从真实联合分布 $p_{tr}(\mathbf{x}, y)$ 中采样，即：

$$\begin{aligned} \mathcal{D} &= \{\mathbf{x}^{(i)}, y^{(i)}\}_{i=1}^N \\ \{\mathbf{x}^{(i)}, y^{(i)}\} &\sim p_{tr}(\mathbf{x}, y) \end{aligned} \quad (2.10)$$

在信息论中，熵 (entropy) 可以被用于度量不确定性^[1]。单个数据不确定性可以被定义为真实条件分布的熵 $\mathcal{H}[p_{tr}(y|\mathbf{x})]$ ：

$$\mathcal{H}[P_{tr}(y|\mathbf{x})] = - \sum_{c=1}^K P_{tr}(y = \omega_c|\mathbf{x}) \ln P_{tr}(y = \omega_c|\mathbf{x}) \quad (2.11)$$



图 2.1 数据不确定性大的输入



图 2.2 数据不确定小的输入

如图2.1、2.2，从数据本身出发，不妨假设图2.1有 60% 的概率是狗，40% 是猫，图2.2有 90% 是狗，10% 是猫，根据公式2.11，图2.1的数据不确定性为 0.673，图2.2的数据不确定性为 0.325。上述定义可以反映输入数据的不确定性。

整体的数据不确定性可以被定义为条件熵 $\mathcal{H}[p_{tr}(y|\mathbf{x})]$ 的期望：

$$\mathbb{E}_{p_{tr}(\mathbf{x})}[\mathcal{H}[p_{tr}(y|\mathbf{x})]] \quad (2.12)$$

也可以通过互信息 (mutual information) 来衡量整体的数据不确定性：

$$\begin{aligned} \mathcal{I}[y, \mathbf{x}] &= KL[p_{tr}(\mathbf{x}, y) || p_{tr}(\mathbf{x})p_{tr}(y)] \\ &= \mathcal{H}[p_{tr}(y)] - \mathbb{E}_{p_{tr}(\mathbf{x})}[\mathcal{H}[p_{tr}(y|\mathbf{x})]] \end{aligned} \quad (2.13)$$

其中，边缘分布 $p_{tr}(y)$ 定义为：

$$p_{tr}(y) = \int_{\mathcal{R}^D} p_{tr}(\mathbf{x}, y) d\mathbf{x} \quad (2.14)$$

互信息在信息论中定义为联合分布 $p_{tr}(\mathbf{x}, y)$ 与边缘分布 $p_{tr}(\mathbf{x}), p_{tr}(y)$ 的乘积的 KL 散度，如果 $\mathbf{x}$ 与 y 独立性越大，即输入无法给出更多的信息增益，数据的不确定性越大，互信息越小，反之亦然。

在回归问题中， y 并非离散的值，而是连续的变量，此时式2.11中熵的求和变为微分熵 (differential entropy) 的积分：

$$\mathcal{H}[p_{tr}(y|\mathbf{x})] = - \int p_{tr}(y|\mathbf{x}) \ln p_{tr}(y|\mathbf{x}) dy \quad (2.15)$$

回归问题中的整体数据不确定性也可以类似分类问题进行计算，只不过熵的非负性在微分熵中不再存在，但不确定性的相对关系依然保持。

总而言之，数据不确定性的物理含义是与模型无关，内在于数据中的不确定性，在简单的理想条件下，可以用数学公式表达，不过实际问题中我们无法知道真实分布 p_{tr} ，只能用2.3节中的方式来估计数据不确定性。

2.2.2 模型不确定性

模型不确定性 (Model Uncertainty), 又叫认知不确定性 (Epistemic Uncertainty), 知识不确定性 (Knowledge Uncertainty), 定义为对哪一物理模型产生了当下观测的无知性, 可以通过更多的数据消除^[21]。模型不确定性产生的原因是模型中认知与知识的缺乏^[22-23]。

模型不确定性也称分布不确定性 (Distributional Uncertainty)^[22], 借用域适应 (Domain Adaptation) 领域的术语, 训练数据来自于一个潜在的分布 (Distribution), 而在复杂任务 (如自动驾驶的感知任务) 中, 测试数据往往服从不同的潜在分布, 模型容易过拟合于训练数据分布, 而对分布外 (OOD, Out Of Distribution) 的数据手足无措。因此, OOD 数据是模型不确定性产生的重要原因。

一般认为模型不确定性比预测不确定性更难估计^[22-23], 原因之一就是不像数据不确定性那样, 可以由数据和信息论直接给出理论定义, 只有通过2.3节的方式实际估计。

2.3 深度学习中的不确定性估计

如第一章所述, 对神经网络输出的不确定性估计非常重要, 将不确定性理论融入深度学习成为高度安全相关的人工智能系统不可避免的课题。本节重点关注输出 2.2 节中两类不确定性的算法。部分以模型最后一层输出为置信度, 从而表现总预测不确定性的方法将在 2.4 节介绍。神经网络的不确定性估计主要分为以 MC Dropout^[11]为代表的贝叶斯方法和以集成^[25] (Ensemble) 为代表的非贝叶斯方法。

在实际使用时, 回归问题更多采用2.3.2节方法, 因为集成模型概率分布的期望无法实际计算^[22], 需要用方差替代^[26]。分类问题更多采用2.3.3节方法, 因为分支网络的模型不确定性并没有给出公认的计算方式^[27]。

2.3.1 蒙特卡洛随机失活

随机失活 (以下称 Dropout)^[10]是一种解决神经网络过拟合的常用正则化手段。它在训练时以概率 p 随机丢失全连接层的神经元, 测试时正常开启并乘以统一尺度的系数 $\frac{1}{1-p}$ 。目的是确保网络中的部分神经元组合也能正常工作, 防止网络部分神经元权重过大, 导致其它神经元无效。

Dropout 是贝叶斯理论与现代神经网络结合的重要桥梁。对于深度学习不确定性量化问题, Gal 在 16 年的博士论文^[14]证明了带有 dropout 的神经网络等价于某种变分推断的优化过程, 通过 dropout 优化的权重, 已经具备变分推断、贝叶斯神经网络 (BNN, Bayesian Neural Network) 的特性; Gal 在同年的 ICML 会议论文^[11]证明了一个任意深度和非线性激活函数的神经网络, 只要每一个权重层都应用了 dropout, 就在

数学上等价于一个深度高斯过程。Gal 在 17 年的 ICML 会议论文^[28]又指出：Dropout 和其它一些随机正则化手段可以被认为是一种近似推断。

这里进行不确定性量化的 Dropout 方式称为蒙特卡洛 Dropout (MC Dropout, Monte-Carlo Dropout)，与一般的 Dropout 不同，MC Dropout 要求在测试时开启 Dropout 层，这让网络在测试时也具有了随机性。仅使用 MC Dropout，可以将网络近似后验分布的方差 ($\text{Var}_{q(y^*|x^*)}(y^*)$)，作为神经网络总不确定性的一种度量^[11]。

总而言之，带有 Dropout 的模型在特定条件下，可以理解作为一种贝叶斯神经网络^[11]，多次采样近似后验分布的过程等价于多次输入样本前传的过程^[28]。而在实际使用时，每一权重层都要加 Dropout 的约束条件可以被适当放宽^[29]，加入了 Dropout 的神经网络可以通过传统的贝叶斯理论得到不确定性估计。

同时，带有 Dropout 层的神经网络，也可以看做一种共享参数的特殊网络集成，可以用 2.3.3 节的集成模型估计算法来得到模型和数据两类不确定性。

2.3.2 分支网络估计

在文献^[21]中，给出了将两类不确定性融合于一个模型的方法。该方法基于贝叶斯神经网络，即带有 Dropout 的神经网络。

因为这类方法需要简单修改原任务的网络，添加预测方差（理解为数据不确定性）的分支网络，所以将这类方法称为分支网络估计方法。

对于回归问题，我们记输入为 $\mathbf{x}$ ，贝叶斯网络模型 f ，参数 $\hat{W}$ 采样于近似后验分布 $q(W)$ ，即开启了 MC Dropout 的网络参数。我们设计分支结构，要求网络在输出预测均值 $\hat{y}$ 的同时输出预测方差 $\hat{\sigma}^2$ ：

$$[\hat{y}, \hat{\sigma}^2] = f^{\hat{W}}(\mathbf{x}) \quad (2.16)$$

此时网络优化的损失函数为：

$$\mathcal{L}_{BNN}(\theta) = \frac{1}{D} \sum_i \frac{1}{2} \hat{\sigma}_i^{-2} \|y_i - \hat{y}_i\|^2 + \frac{1}{2} \log \hat{\sigma}_i^2 \quad (2.17)$$

其中， θ 为产生 $\hat{y}_i, \hat{\sigma}_i^2$ 的网络参数， D 为每个预测 y 的维度。如图片逐像素深度估计问题中， D 为图片像素总数， i 代表第 i 个像素， y_i 为第 i 个像素的真值， $\hat{y}_i$ 为第 i 个像素的预测值， $\hat{\sigma}_i^2$ 是第 i 个像素的预测方差。

该损失函数可以理解为极大高斯似然，认为输出 y 服从均值为 $\hat{y}$ ，方差为 $\hat{\sigma}^2$ 的高斯分布：

$$p(y_i|x_i, \theta) = \frac{1}{\sqrt{2\pi\hat{\sigma}_i^2}} \exp\left(-\frac{(y_i - \hat{y}_i)^2}{2\hat{\sigma}_i^2}\right) \quad (2.18)$$

取负对数，极大 2.18 中 y_i 为真值的概率等价于最小损失函数 2.17。

2.17也可以理解为：预测方差 $\hat{\sigma}_i^2$ 较大的地方，对损失函数的贡献较小，更容许预测 $\hat{y}_i$ 与真值有偏差，同时为了防止优化时处处都输出较大方差，还加入了正则项 $\frac{1}{2}\log\hat{\sigma}_i^2$ 。

此时的数据不确定性 AU 由分支网络输出。更具体地，在测试时，打开 Dropout，前传 T 次，取分支网络输出的均值作为针对某一输入数据的数据不确定性：

$$AU(x_i) = \frac{1}{T} \sum_{t=1}^T \hat{\sigma}_{t,i}^2 \quad (2.19)$$

此时的模型不确定性 EU 由主干网络输出的方差得到。更具体地，在测试时，打开 Dropout，前传 T 次，计算得到的方差作为针对某一输入数据的模型不确定性：

$$EU(x_i) = \frac{1}{T} \sum_{t=1}^T \hat{y}_{t,i}^2 - \left(\frac{1}{T} \sum_{t=1}^T \hat{y}_{t,i} \right)^2 \quad (2.20)$$

其中，T 为前传次数， $\hat{y}_{t,i}, \hat{\sigma}_{t,i}^2$ 是模型输出，即 $\hat{y}_{t,i}, \hat{\sigma}_{t,i}^2 = f^{\hat{W}_t}(x_i)$ ， $\hat{W}_t \sim q(W)$ 。

总的不确定性就是 AU 与 EU 的和。

对于分类问题，思路和回归问题类似。

记 $\hat{z}_i$ 是 logits 输出（即没有经过 Softmax 的 $f_i^W(x_i)$ ），记每一类别 c 的 logits 为 $\hat{z}_{c,i}$ ：

$$\begin{aligned} (\hat{z}_{c,i}|W) &\sim \mathcal{N}(f_i^W(x_i), (\sigma_i^W(x_i))^2) \\ \hat{p}_i(x_i) &= \text{Softmax}(\hat{z}_i) \end{aligned} \quad (2.21)$$

损失函数形式由高斯似然函数变为在 logits 加入了方差为 $(\sigma_i^W(x_i))^2$ 的高斯噪声的软最大（Softmax）似然函数：

$$\mathcal{L}(x_i) = -\log E_{\mathcal{N}(\hat{z}_i; f_i^W(x_i); (\sigma_i^W(x_i))^2)} [\hat{p}_{i, c_{true}}(x_i)] \quad (2.22)$$

通过采样的方式计算期望，对主网络和分支网络分别优化：

$$\begin{aligned} \hat{z}_{i,t} &= f_i^W(x_i) + \sigma_i^W(x_i) \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, I) \\ \mathcal{L}(x) &= - \sum_i \log \frac{1}{K} \sum_{k=1}^K \frac{\exp(\hat{x}_{i,k, c_{true}})}{\sum_{c=1}^C \exp(\hat{x}_{i,k, c})} \end{aligned} \quad (2.23)$$

其中， c_{true} 为真实类别， $\hat{p}_{i, c_{true}}$ 为预测向量在真实类别上分量，K 为采样次数，C 为类别总数，k 为第 k 次采样，c 为第 c 个类别。

关于不确定性的估计原文没有给出，本文按照自己的理解补全。

针对单个输入的数据不确定性，与回归问题类似：

$$AU(x_i) = \frac{1}{T} \sum_{t=1}^T (\sigma_{i,t}^W(x_i))^2 \quad (2.24)$$

模型不确定性本应取主干网络输出的方差，但针对预测概率 $p_{i,c_{pred}}$ 求方差并不合理，可能的思路有：

1、针对 logits 多次前传求方差

$$\begin{aligned} EU(x_i) &= \frac{1}{C} \sum_{c=1}^C \mathbb{V}(\hat{z}_{c,i}) \\ &= \frac{1}{C} \sum_{c=1}^C \left(\frac{1}{T} \sum_{t=1}^T \hat{z}_{c,t,i}^2 - \left(\frac{1}{T} \sum_{t=1}^T \hat{z}_{c,t,i} \right)^2 \right) \end{aligned} \quad (2.25)$$

2、针对平均概率求熵：

$$\begin{aligned} EU(x_i) &= H\left(\frac{1}{T} \sum_{t=1}^T \hat{p}_{i,t}(x_i)\right) \\ &= - \sum_{c=1}^C \left(\frac{1}{T} \sum_{t=1}^T p_{i,t,c}(x_i) \right) \ln \left(\frac{1}{T} \sum_{t=1}^T p_{i,t,c}(x_i) \right) \end{aligned} \quad (2.26)$$

其实，思路 2 得到的 EU 是集成模型估计 2.3.3 节中的总不确定性。因为原文^[21]并没有给出分类问题下的两类不确定性计算方式，本文也没有对这种方式得到的两类不确定性进行实验，不作进一步评价。

值得一提的是，SalsaNext^[12]是针对三维点云语义分割任务（分类问题）的网络模型，与本文研究方向类似。SalsaNext 设计融合了基于 ADF 构架^[13]的 MC Dropout 方法，单独实现了一个可以输出数据不确定性和模型不确定性的部署版本。

但是，SalsaNext 假设的是经过 softmax 层后的概率向量满足高斯分布，此时损失函数变为：

$$\mathcal{L} = \frac{1}{2\hat{\sigma}^2} \|y_i - p_i\|^2 + \frac{1}{2} \log(\hat{\sigma}^2) \quad (2.27)$$

违反直观的是，这里 y_i 是真值的单热点（one-hot）向量，这一形式类似最小均方误差（MSE, Mean Square Error）损失函数（回归问题的 MSE 损失函数可以由极大高斯似然函数推得）。但在分类问题中通常不使用 MSE 函数进行优化（原因包括损失函数非凸，真值属于多项式分布，概率真值属于多项式分布，优化方向不确定等）。因此，相较于 2.23，这一损失函数的设计存疑。

基于以上损失函数，SalsaNext 的不确定性取 Dropout 前传 T 次，C 个类别的平均：

$$AU = \frac{1}{T} \sum_{t=1}^T \frac{1}{C} \sum_{c=1}^C (\hat{\sigma}_{t,c}^2) \quad (2.28)$$

模型不确定性取 logits 向量的方差在各类别的平均，这里即以上思路一，式 2.26。

2.3.3 集成模型估计

集成 (Ensemble) 模型顾名思义, 就是将多个完成相同任务的模型并联, 通过某种融合手段将这些模型集成, 以得到比单模型更好的输出。这类似人类集体决策的过程, 常用的融合方式有共享参数、投票等。一直以来, 集成模型被认为可以得到更好的预测性能^[30]。不过也一直存在需要几倍于单模型参数和测试时间的问题。随着大模型时代的到来, 加之其理论上可以方便并行, 集成的思想有望受到更多关注。

通常集成模型的训练采用同一套网络结构, 进行不同的初始化, 或者不同的数据增强方式^[25]。单个采用 Dropout 训练的模型, 可以理解为一种共享权重的集成模型^[25]。

以下参考 Malinin 博士论文^[22]中关于集成模型不确定性的建模。

记 $P(y|x^*, \mathcal{M}^{(m)})_{m=1}^M$ 是一组集成模型。在集成模型中, 模型不确定性可以建模为多个模型之间的分歧, 即输出 y 与模型 $\mathcal{M}$ 之间的互信息 $\mathcal{I}[y, \mathcal{M}|x^*, \mathcal{D}]$:

$$\mathcal{I}[y, \mathcal{M}|x^*, \mathcal{D}] = \mathcal{H}[\mathbb{E}_{p(\mathcal{M}|\mathcal{D})}[P(y|x^*, \mathcal{M})]] - \mathbb{E}_{p(\mathcal{M}|\mathcal{D})}[\mathcal{H}[P(y|x^*, \mathcal{M})]] \quad (2.29)$$

其中, 等号右边第一项可以理解为总不确定性, 等号右边第二项可以理解为数据不确定性。总不确定性项是模型最终输出的交叉熵, 是两类不确定性共同作用的结果; 数据不确定性项由单个模型的交叉熵的期望组成, 与模型间的分歧无关。

在分类问题中, 分类网络经过 Softmax 层后的输出可以被理解为网络预测的概率向量 $p(y|x^*, \mathcal{M})$, 此时 $\mathbb{E}_{p(\mathcal{M}|\mathcal{D})}[\mathcal{H}[P(y|x^*, \mathcal{M})]]$ 可以通过采样模型参数, 平均输出近似得到:

$$\mathbb{E}_{p(\mathcal{M}|\mathcal{D})}[\mathcal{H}[P(y|x^*, \mathcal{M})]] = \frac{1}{M} \sum_m [p(y|x^*, \mathcal{M}^{(m)})] \quad (2.30)$$

但在分类问题中, 虽然公式2.29形式不变, 但其中熵是积分形式的微分熵, 对于简单的高斯分布可以解析的计算出微分熵, 但对无法给出解析形式的分布计算微分熵并不容易。由于公式2.29中的总不确定性项很难计算, 这一公式在回归问题的应用受到了限制。

一个替代的方式是脱离信息论的不确定性度量, 采用方差分解公式 (Law of Total Variance)^[26]:

$$\mathbb{V}[y|x^*] = \mathbb{V}_{p(\mathcal{M}|\mathcal{D})}[\mathbb{E}_{p(y|x^*, \mathcal{M})}[y]] + \mathbb{E}_{p(\mathcal{M}|\mathcal{D})}[\mathbb{V}_{p(y|x^*, \mathcal{M})}[y]] \quad (2.31)$$

其中, 总方差 $\mathbb{V}[y|x^*]$ 可以理解为总不确定性, 被分解为期望方差 (数据不确定性) 和集成模型中每个模型预测的方差的均值 (模型不确定性)。

2.4 深度学习的模型置信度

2.4.1 深度学习模型置信度的定义

模型置信度和不确定性在物理意义上是同一件事的两面，也代表模型对输出的确信程度。在分类问题中，不少研究直接将经过 Softmax 后的概率向量的最大分量作为网络对于输出的置信度^[31-32]。这样的好处是能轻松得到置信度，但问题是神经网络总是过于自信^[25, 31, 33]。Gal 也在论文^[11]中列举了这种方式输出置信度可能存在的问题。

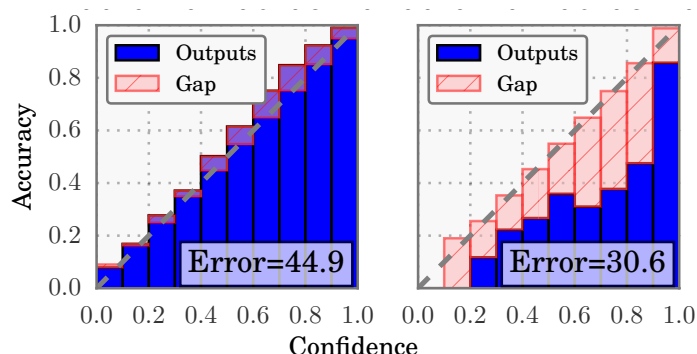


图 2.3 预测置信度与真实准确率校准图，图片摘自^[31]

2.4.2 置信度的校准

不过输出的置信度可以在统计上校准 (Calibration) 到真实置信度。如图2.3，横坐标是网络对某样本的预测置信度，将置信度在 $[0,1]$ 均匀划分为 10 格，纵坐标是预测置信度属于某格区间范围的样本的真实准确率。图2.3左边是校准的较好的模型，虽然预测精度 (右下角 Error) 不如2.3右边，但我们能根据它的置信度判断模型对输出的把握。常用的校准方式有保续回归 (Isotonic Regression, IR)^[34]，其核心思路是将原本未校准的置信度分段映射到与真实准确率校准的置信度，如图2.4。

另一种常用校准方式是温度归一 (Temperature Scaling, TS)^[31]，其核心思路是将网络最后层置信度除以温度系数 T 。一般在验证集上优化负对数似然 (NLL) 指标，网格搜索 (Grid Search) 找到 T 的最佳取值。因为 TS 的简单有效，许多后续工作基于 TS 和集成模型的 TS 展开^[33, 35]。

一般置信度校准的研究会与数据集迁移 (Dataset Shift) 以及数据分布外 (OOD) 问题结合。在进行了大量简单数据集 (Toy Dataset) 和浅层模型的实验后,^[36]得到了以下结论：

- 1、几乎所有方法在数据集迁移的过程，不确定性性能和精度都面临下降。
- 2、在独立同分布的测试集上表现好的校准方法并不一定在分布外 OOD 数据集表现好。

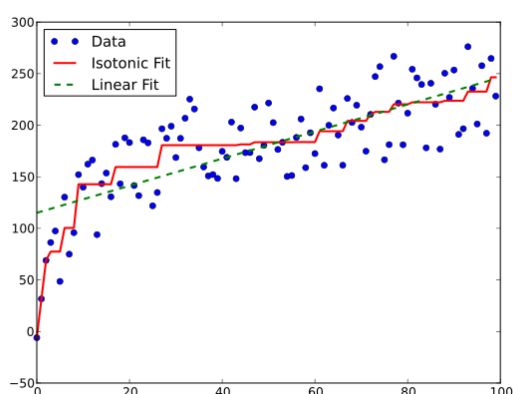


图 2.4 保序回归示意图

3、当数据集的偏移增大的时候，后验的置信度校准方法不如评估模型不确定性 (EU) 的方式。

4、深度集成的方法对数据集迁移最鲁棒。

2.4.3 置信度与不确定性关系

目前多数文章仍将模型置信度的研究归在不确定性估计 (Uncertainty Quantification) 的关键词下。两者除了在获取方式上不同之外，可以进行直接比较。在对比了置信度和不确定性相关文献后，本文初步整理了两者之间的优劣势，详见表2.1。

表 2.1 不确定性与置信度的优缺点

	不确定性 (Uncertainty)	置信度 (Confidence)
优点	1、有贝叶斯理论体系支撑 2、视觉问题中的两类不确定性有物理含义	1、可校准，得到的数值物理含义强，可以直接与真实准确率关联验证 2、测试时的计算开销少，结果稳定
缺点	1、得到的数值可解释性弱 2、目前没有公认的校准方法和度量指标	1、缺乏理论支撑 2、校准于验证集，要求与测试集的分布偏移不大

第三章 三维点云语义分割任务介绍

感知环境是自动驾驶汽车上路的第一环任务，三维点云语义分割是自动驾驶中的关键感知任务之一^[37]。自动驾驶汽车通过各种传感器来感知周围环境。GPS 采集由数值向量构成的位置信息，摄像头采集由二维图像构成的周围环境信息，而激光雷达采集点云格式的周围环境信息。相较于普通相机，三维激光雷达（3D Lidar）抗光线干扰能力强^[38]，感知范围大^[39]，已经是自动驾驶问题中广泛使用的传感器^[37]。除了激光雷达以外，RGB-D 相机也可以采集带有颜色信息的稠密点云数据，但激光雷达采集的点云更稀疏，精度更高^[40]。在本文的语境中，三维点云语义分割任务专指针对激光雷达采集的点云格式的数据进行分析与场景理解的核心任务。

本章主要介绍三维点云语义分割任务的问题定义，对这一任务的现有算法、数据集和评价指标进行总结。

3.1 问题定义

如图3.1，三维点云数据格式是由 N 个点，每个点 d 维特征组成的。我们以自动驾驶感知任务涉及的三维点云为例，激光雷达在室外大场景一帧采集数万个点，每个点一般只有空间位置信息，即特征是 3 维的 x, y, z 坐标，部分数据集给出了反射率等其它特征信息。三维点云语义分割问题每一帧的输入就是这样 N 个点云，输出是针对每个点的分类标签，即事先定义好如汽车、道路、行人等类别标签，将每个点归类到某一类别。

3.2 三维点云语义分割算法

在深度学习普及之前，就已经有三维点云语义分割的相关研究。这些传统方法通常使用手工设定的特征提取方法，如支持向量机、随机森林算法^[41]等。这些算法具有很好的可解释性。但是，随着深度学习的广泛应用，基于深度学习的算法取得了传统算法无法企及的效果。近年来的最先进（State of the Art）方法都是基于深度学习的。下面以输入点云的形式进行分类，对近年来的点云分割算法进行总结。

3.2.1 基于图像的算法

激光雷达扫描点云时有多根激光线，每根线扫描一周 360° 。以 HDL-64E 为例，其为 64 线激光雷达，最高分辨率为 0.35° ，即单线最多扫描到 1028 个点。利用激光雷达

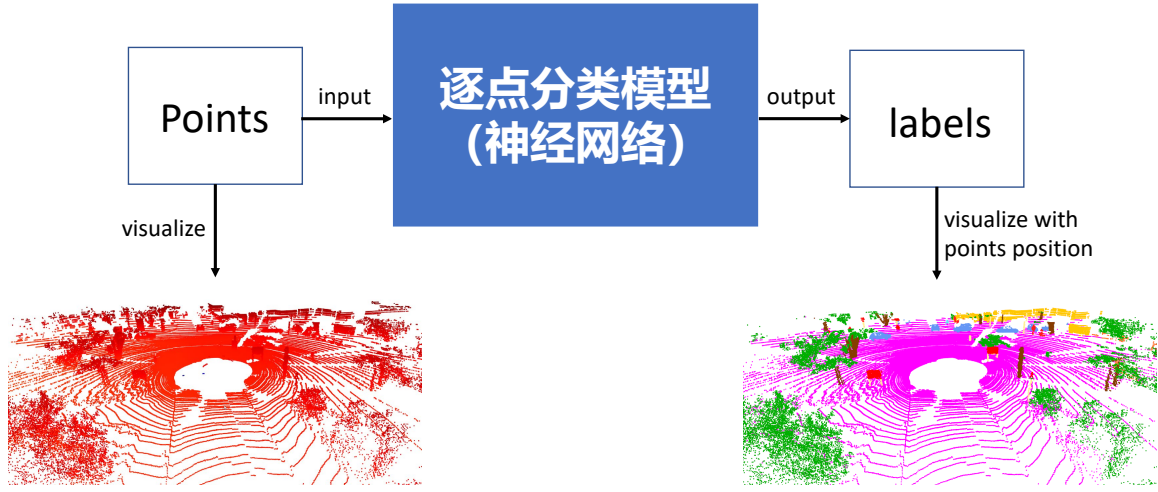


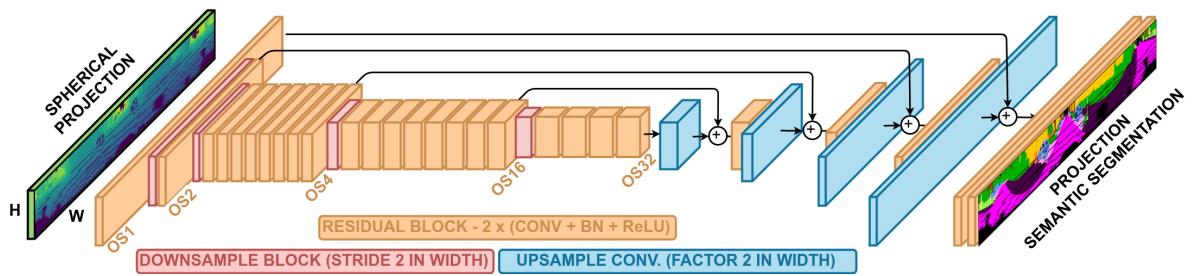
图 3.1 三维点云语义分割任务基本流程示意图

这一性质，理论上可以将点云无损的映射到一张很扁的图片。记图片宽度为 W ，高度为 H ，原始点云坐标 (x, y, z) 与图片像素位置 (u, v) 的转换关系为^[42]:

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \frac{1}{2} [1 - \arctan(y, x)/\pi] W \\ [1 - (\arcsin(z/\sqrt{x^2 + y^2 + z^2}) + FOV_{down})/FOV] H \end{pmatrix} \quad (3.1)$$

其中， FOV_{down} 为激光雷达的最大下视角， FOV 为最大上下全视角。

这样转化后的图片被称为深度图像 (Range Image)。当然这样转化到图片像素后，原始点云的深度信息丢失了，为保留这一信息，会为每个投影后的点记录 (x, y, z) 等原始信息，于是点云输入 (N, d) 就转化成了带有蒙版 (Mask) 的深度图像输入 (H, W, d') 。


 图 3.2 基于图像的代表算法 RangeNet++^[43]流程图

接下来，用神经网络进行类似 U-net^[44] 的卷积下采样反卷积上采样操作，以提取特征，进行逐像素的分类，得到基于深度图像的预测。利用这一预测设计交叉熵等损失函数，进行网络的学习。预测时，只要将得到的深度图像预测映射回原始点云即可。

在这一过程中, SqueezeSegV3^[42]注意到深度图像与普通图像信息量分布的不同。因为经过投影映射, 图片中只有部分对应原始点云有物体的像素有效, 再用同样的卷积算子对整张图片进行计算就变得不合适了, 所以 SqueezeSegV3 给卷积算子加上空间注意力机制, 对不同位置的像素进行不同的卷积计算。

之后的实验中, SqueezeSegV3 作为本文的一种三维点云语义分割基线算法。虽然这一算法的效果与先进的点云分割算法有差距, 但计算效率较高。

目前, 已有不少基于图像的点云分割算法^[12, 38, 43]。最新的研究 RangeFormer^[45]第一次证明了基于深度视角的方法也可以超过基于体素和基于点云的方法。其在特征提取的时候利用了 Transformer^[5]模块, 还提出了基于深度图像的数据增强和后处理方式, 目前在 SemanticKITTI 数据集上达到了 73.3% 的 mIoU。

3.2.2 基于体素的算法

除了利用投影到图像的方式以外, 还可以将整个空间划分为体素网格, 进行三维的卷积操作。具体而言, 将 (N, d) 的点云转化为 (X, Y, Z, d') 的体素格式, 进行空间卷积操作。需要注意的是, 点云相比于全空间来说是稀疏的, 大部分体素中其实不含有点云, 此时三维卷积的运算量远大于二维卷积, 因此需要引入稀疏卷积的操作^[46]。三维稀疏卷积的整体流程还是类似二维卷积, 只是计算方式有所不同。

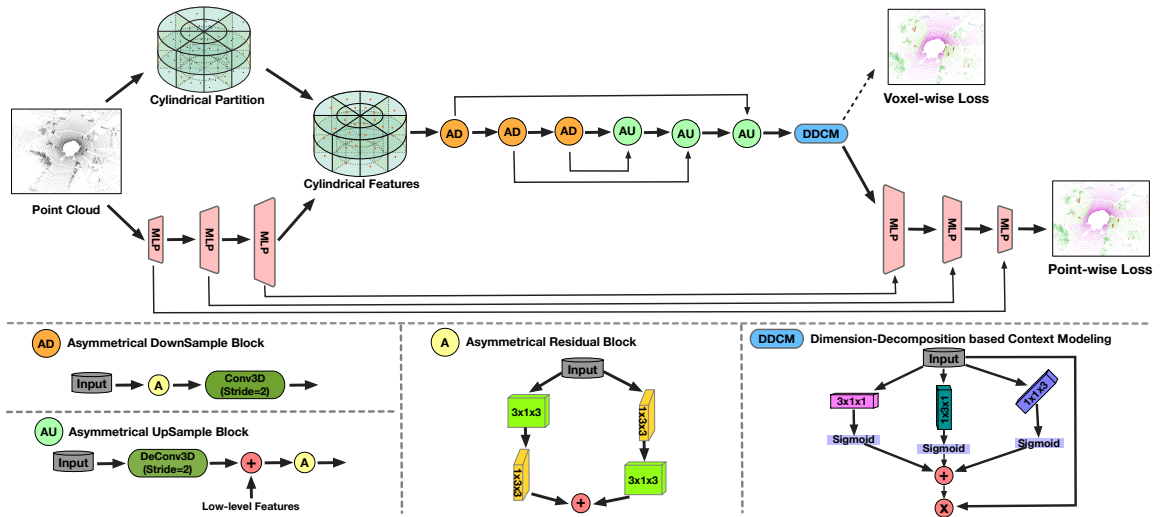


图 3.3 基于体素的代表算法 Cylinder3D^[47]流程图

在这一过程中, Cylinder3D^[47]注意到了激光雷达采集点云近密远疏的特点^[41]。相比于方格式的划分, Cylinder3D 采用圆柱坐标系, 按角度、平面距离和高度划分空间, 符合了近密远疏的特征。同时, Cylinder3D 加入了非对称残差模块, 让空间的水平垂直响应更突出, 更适于描述真实世界的物体特征。之后的实验中, Cylinder3D 作为主

要的基线算法，截止目前，其在效果上仍然是先进的。

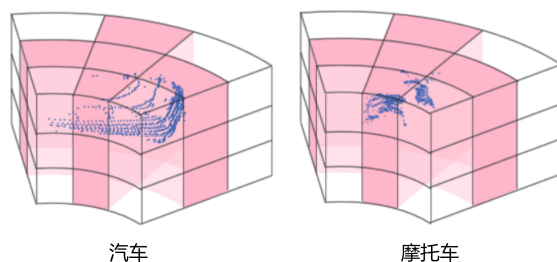


图 3.4 Cylinder3D 空间划分方式，图摘自^[47]

目前，已有不少基于体素的点云分割算法^[48-51]。最新的研究 SphereFormer^[52]是基于空间体素的算法，它的空间划分方式是雷达窗口和立方体窗口融合，相比于其它算法，更关注了远处物体的分类准确率。目前在 SemanticKITTI 数据集上达到了 74.8% 的 mIoU。

3.2.3 基于点云的算法

除了将点云进行特定的处理，化归到图像或者体素，也有算法大量直接对点云进行处理^[53-56]。

这类算法的先驱是著名的深度学习在点云分割的开山之作 Pointnet，重点关注点云的无序性，平移不变性。同时，点云的位置信息被反复利用，此类方法会更聚焦于点云与邻居的特征联结。在这一过程中，RandLA^[57]提出了基于空间注意力的局部特征聚合和局部特征编码模块。经过部署测试，其效果和效率与 SqueezeSegV3 和 Cylinder3D 相比，都处于中间水平，本论文没有对这一算法进行过多实验，未来的研究中可以进一步探究。

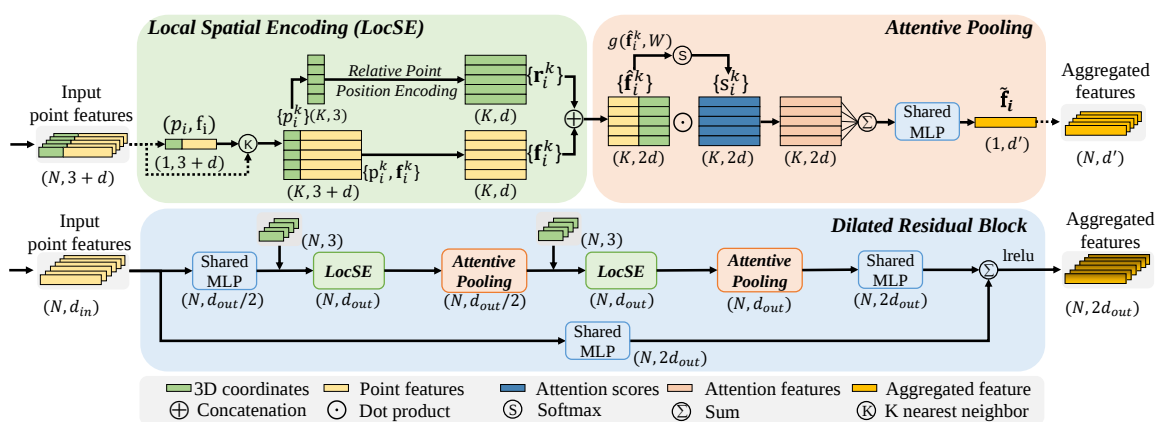


图 3.5 基于点云的代表算法 RandLA^[57]流程图

3.3 经典数据集

深度监督学习模型依赖大量带有真值的数据集。正如最近由 Meta 发布的针对二维图像分割任务的 Segment Anything 模型，就是针对超过 1 千万张图像，超过 10 亿个掩码进行学习的大模型^[58]。



图 3.6 Segment anything 二维图片分割示例图

不过相较于二维图片，点云数据本身不容易直观理解，标注成本大^[37, 41]。这里介绍三个开源的自动驾驶场景的三维点云语义分割数据集。

SemanticKITTI^[40]是由德国波恩大学采集并标注的室外大场景点云数据集，目前仍有大量方法在该数据集上进行测试与比较。该数据集一共包含 22 条序列，其中序列 0-10 将真值开放，一般序列 8 用作验证集，其余 10 条用于训练。在公开的竞赛中划分了 20 个类别标签。

SemanticPOSS^[59]是由北京大学赵卉菁课题组采集的北京大学校园内的道路场景，数据格式与 SemanticKITTI 完全一致，但校园道路场景与城市道路场景的差异较大，如校园中行人、自行车的占比大幅上升。

nuScenes^[60]是由 Motional 团队开发的在波士顿和新加坡采集的自动驾驶数据集，除了激光雷达数据以外，还有 GPS、摄像头等传感器数据，因此还开展了目标检测、轨迹规划等多种竞赛。在点云分割竞赛中划分了 17 个类别标签。

3.4 评价指标

平均交并比 (mIoU, mean Intersection of Union)^[61]:

$$\begin{aligned} IoU(c) &= \frac{|y_{pd} = c \cap y_{gt} = c|}{|y_{pd} = c \cup y_{gt} = c|} \\ mIoU &= \frac{1}{C} \sum_c IoU(c) \end{aligned} \quad (3.2)$$

其中， c 为某一类别， C 为类别总数， y_{pd} 为预测标签， y_{gt} 为真实标签。

针对每一个类别都可以计算类别对应的 IoU，然后对类别进行平均。这一指标是各个方法进行性能比较时首要关注的指标。

帧率 (Fps, Frames per second):

1 秒钟时间可处理的平均帧数。因为自动驾驶任务要求实时性，部分研究会致力于在保持准确性或牺牲少量准确性的前提下，提升处理效率。这项指标和硬件设备高度相关，车载移动计算平台的性能自然不如固定服务器，因此在部署算法到实车时，这一指标也需要关注。

全局准确率 (OA, Overall Accuracy):

$$OA = \frac{|y_{pd} = y_{gt}|}{N} \quad (3.3)$$

其中，N 为点云总数。OA 即预测正确点数占比。

频率加权交并比 (FwIoU, Frequency weighted Intersection of Union):

$$FwIoU(c) = \frac{|y_{gt} = c| * |y_{pd} = c \cap y_{gt} = c|}{N * |y_{pd} = c \cup y_{gt} = c|} \quad (3.4)$$

相比于 mIoU，FwIoU 只是为每个类别添加了权重，在真值中出现越多的类别权重越大。但因为往往模型在小类别上学习的不好，较先进的模型在这一指标上都表现的不错，所以这一指标的应用并不多。

查准率 (Pre, Precision):

$$Pre(c) = \frac{|y_{gt} = c \cap y_{pd} = c|}{|y_{pd} = c|} \quad (3.5)$$

查全率 (Rec, Recall):

$$Rec(c) = \frac{|y_{gt} = c \cap y_{pd} = c|}{|y_{gt} = c|} \quad (3.6)$$

加权查准率 (wPre, weighted Precision)^[62]:

$$wPre(c) = \frac{|y_{gt} = c \cap y_{pd} = c| / |y_{gt} = c|}{\sum_r (|y_{gt} = r \cap y_{pd} = c| / |y_{gt} = r|)} \quad (3.7)$$

第四章 三维点云语义分割任务中的不确定性估计实验

4.1 任务目标

三维点云语义分割是自动驾驶的重要感知任务，在这一任务中引入两类不确定性的估计，可以传播到后续规划控制模块，以提升自动驾驶的安全性。但如前所述，目前在该任务中引入两类不确定性评估的研究稀缺。本文在基于体素系的 Cylinder3D 算法和基于投影系的 SqueezeSegV3 算法基础上，采用集成模型的不确定性评估算法，分析现有三维点云语义分割模型的不确定性评估效果。

4.2 算法简述

本次实验面向的是三维点云的语义分割任务，属于分类问题，因此本次实验采用2.3.3节的分类问题不确定性评估算法。更具体地，我们采取3.2节介绍的 Cylinder3D 模型和 SqueezeSegV3 模型，统一在模型最后加入丢失率为 0.2 的 Dropout 层。为了比较集成模型多样性带来的不确定性评估差异，本次实验尝试集成两种不同的模型以评估不确定性。

认知不确定性 (AU, Aleatoric Uncertainty):

$$\begin{aligned} AU &= E_{p(M|D)}(H(p(\hat{y}|x^*, M_i))) \\ &\approx \frac{1}{T} \sum_{i=1}^T H(p(\hat{y}|x^*, M_i)), M_i \sim p(M|D) \end{aligned} \quad (4.1)$$

模型不确定性 (EU, Epistemic Uncertainty):

$$\begin{aligned} EU &= H(E_{p(M|D)}(p(\hat{y}|x^*, M_i))) - E_{p(M|D)}(H(p(\hat{y}|x^*, M_i))) \\ &\approx H\left(\frac{1}{T} \sum_{i=1}^T p(\hat{y}|x^*, M_i)\right) - \frac{1}{T} \sum_{i=1}^T H(p(\hat{y}|x^*, M_i)), M_i \sim p(M|D) \end{aligned} \quad (4.2)$$

4.3 实现细节

本次实验采用的三维点云语义分割模型主要是 Cylinder3D 和 SqueezeSegV3。SqueezeSegV3 采用 21 层版本，不加入 KNN 后处理。

训练采用的数据集是 SemanticKITTI。SemanticKITTI 开源有真值的训练集是序列 0-10，我们选取序列 0-4 作为训练集，序列 8 作为验证集。即两模型独立在序列 0-4 上训练 30 代，预测阶段在序列 8 上每 10 帧抽取一帧，进行模型表现与不确定性的分

析。观察模型表现时，两模型关闭 Dropout 层输出；单模型 Dropout 评估不确定性时，打开 Dropout 层多次前传，如图4.1；两模型集成评估不确定性时，关闭 Dropout，如图 4.2。

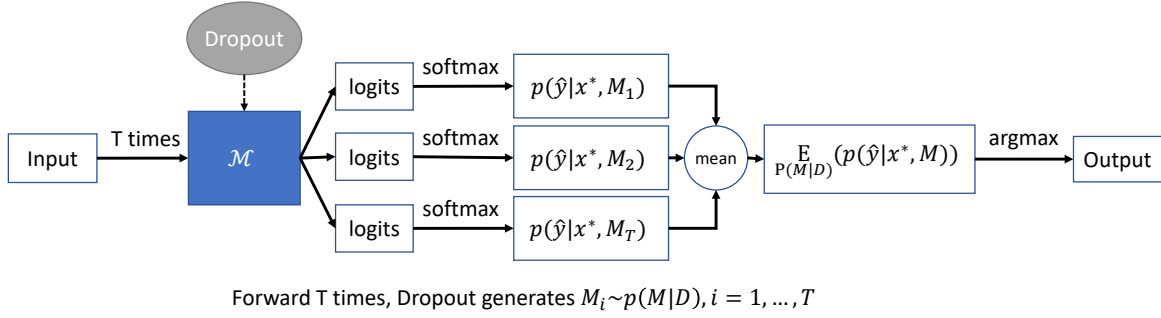


图 4.1 单模型 Dropout 算法流程

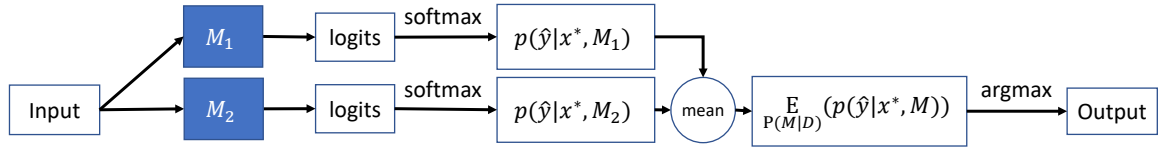


图 4.2 两模型 Ensemble 算法流程

一般来说，语义分割数据集在标注时会存在一些不太好定义的点，将其归为 0 类。为了保证一帧点云的空间数据结构不变，属于 0 类的点在训练、测试时都会正常输入，但是不会计入 loss 的计算，也不会计入 IoU 的计算。除去特殊的 0 类以外，因为一些类别的相似和冗余，同时也方便跨数据集比较，如图4.3，本次实验将 SemanticKITTI 数据集原始的 26 类合并为 11 类，划分方式和数据量如表4.1所示。具体地，我们将 11 个类别按数据量划分为大、中、小类别。大类别为：道路 (road, ro)，植被 (plants, pl)，建筑 (building, bu)，栅栏 (fence, fe)，车辆 (car, ca)；中类别为：树干 (trunk, tr)，杆 (pole, po)；小类别为：信号 (sign, si)，自行车 (bike, bi)，行人 (people, pe)，骑手 (rider, ri)。

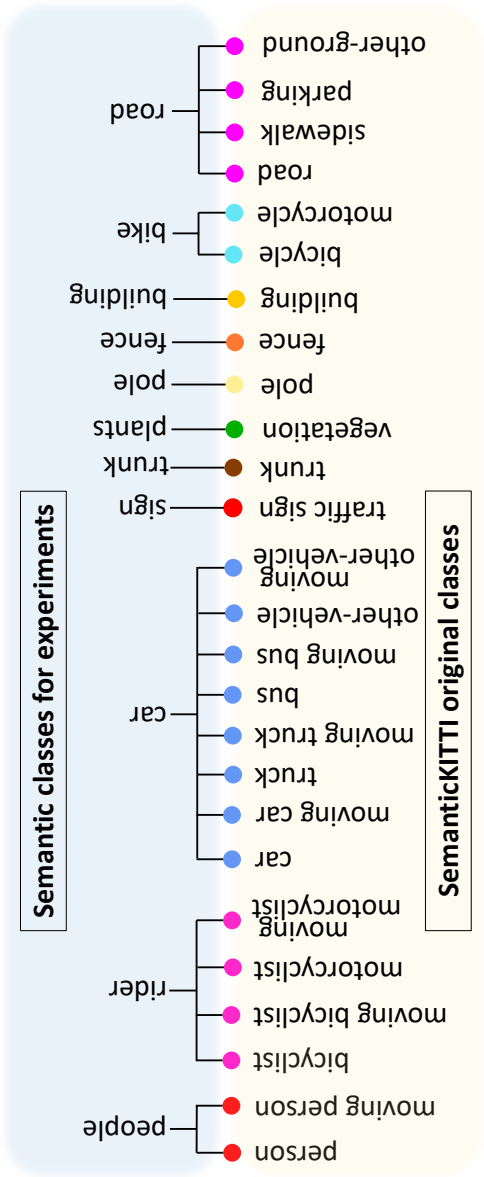


图 4.3 本次实验类别定义

表 4.1 SemanticKITTI 数据集训练集和测试集各类别数据量

scale	Large					Middle		Small			
class	road(ro)	plants(pl)	building(bu)	fence(fe)	car(ca)	trunk(tr)	pole(po)	sign(si)	bike(bi)	people(pe)	rider(ri)
SemanticKITTI (train)	730.27M	522.39M	268.33M	143.17M	98.19M	12.43M	5.78M	1.19M	1.12M	0.569M	0.386M
SemanticKITTI (test)	154.65M	145.94M	56.88M	12.64M	33.59M	5.51M	1.67M	0.381M	0.594M	0.477M	0.329M
Training weights	1.00	1.00	1.00	1.00	1.00	1.36	2.19	8.48	8.98	17.19	25.09

4.4 实验结果

4.4.1 点云分割任务表现

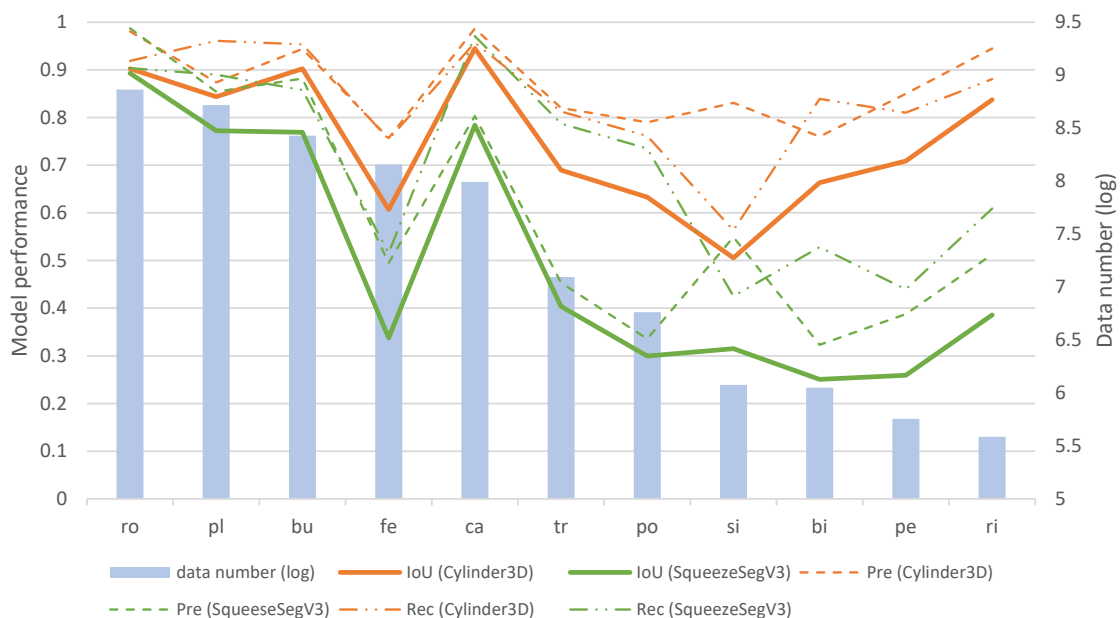


图 4.4 Cylinder3D 和 SqueezeSegV3 点云分割性能表现

从图4.4可以看出, Cylinder3D 和 SqueezeSegV3 两个模型在各个类别上的 IoU (交并比), pre (准确率), rec (召回率) 趋势都是相近的。大类别除了栅栏以外, 两个模型的 IoU 都达到了 75% 以上, pre 和 rec 都达到了 80% 以上。两模型在中类别上的性能有所下降, 在小类别上的性能进一步下降。比较反常的是栅栏这一类别, 虽然数据量较大, 但因为类别内的多样性较大, 点云分割的难度较大, 模型在该类的性能下降。可以从混淆矩阵4.7, 4.8看出, 栅栏类别主要容易被错分成植被、建筑、汽车等类别。Cylinder3D 和 SqueezeSegV3 模型性能差距比较大, 前者 mIoU 为 74.9%, 后者为 49.7%。

如图4.5所示, 两模型在大类别上的性能相近, 但是中类别性能差距较大, 小类别的性能差距更大。在处理速度上, 部署的 SqueezeSegV3 帧率是 Cylinder3D 的两倍。可见, SqueezeSegV3 是一个注重效率的模型, Cylinder3D 是一个注重性能的模型。

如图4.6所示, 按照距离传感器的距离对每帧的点云进行划分, 由于激光雷达点云近密远疏的特性, 每帧点云远距离的点云个数很少。同时可以看到两个模型在远距离的点云分割性能上都有所下降, 这也是大多数模型的共性。

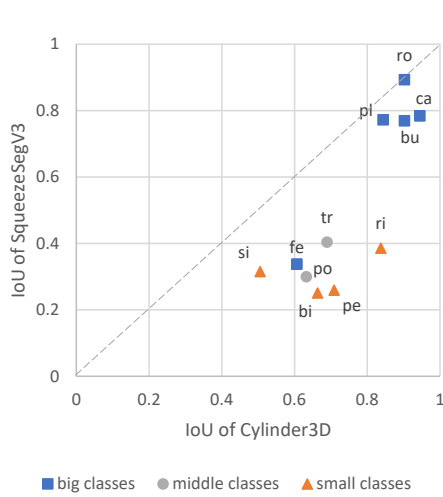


图 4.5 1 两模型性能比较

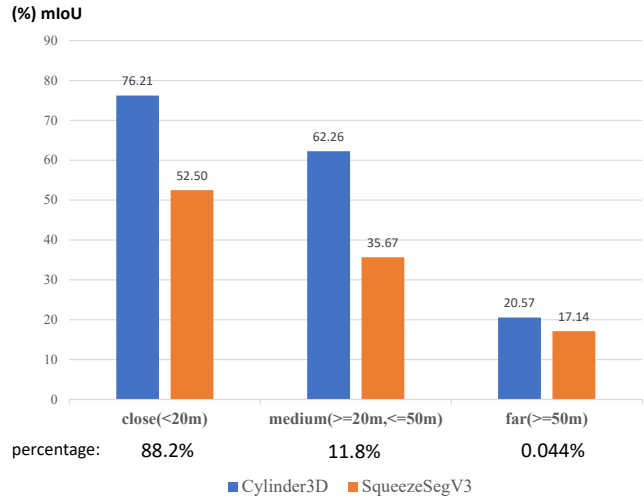


图 4.6 模型性能与距离的关系

4.4.2 三维点云语义分割类别混淆性分析

如图4.7, 4.8, 混淆矩阵的行代表真值为某类别的点, 列代表预测为某类别的点。理想的情况下, 混淆矩阵对角线全为 1, 其它区域为 0, 此时每帧点云的预测完全与真值相同。实际情况是对角线数值较大, 但仍有一些非对角元具有不小的数值。混淆矩阵的行还反映了真值为某类的点云被错分为其它类别的情况, 是一种错分性, 此时可以计算每一类别的错误预测率 (WPR, Wrong Prediction Rate):

$$WPR(c, i) = \frac{|y_{pd} = i \cap y_{gt} = c|}{y_{gt} = c} \quad (4.3)$$

其中 $WPR(c, i)$ 代表真值为 c 类的点被预测成 i 类的概率。

gt \ pd	ro	pl	bu	fe	ca	tr	po	si	bi	pe	ri	Rec
ro	0.92	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.92
pl	0.02	0.96	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.96
bu	0.00	0.02	0.95	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.95
fe	0.02	0.10	0.12	0.75	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.75
ca	0.01	0.01	0.01	0.01	0.96	0.00	0.00	0.00	0.00	0.00	0.00	0.96
tr	0.01	0.16	0.01	0.00	0.00	0.81	0.01	0.00	0.00	0.00	0.00	0.81
po	0.02	0.09	0.02	0.01	0.01	0.07	0.76	0.01	0.00	0.00	0.00	0.76
si	0.00	0.07	0.03	0.04	0.00	0.02	0.27	0.56	0.00	0.00	0.00	0.56
bi	0.02	0.06	0.01	0.02	0.05	0.00	0.00	0.00	0.84	0.00	0.00	0.84
pe	0.02	0.06	0.05	0.01	0.03	0.00	0.00	0.00	0.01	0.81	0.01	0.81
ri	0.01	0.01	0.00	0.00	0.02	0.00	0.00	0.00	0.01	0.06	0.88	0.88
wPre	0.87	0.60	0.78	0.86	0.89	0.89	0.73	0.97	0.98	0.92	0.99	

图 4.7 Cylinder3D 混淆矩阵

gt \ pd	ro	pl	bu	fe	ca	tr	po	si	bi	pe	ri	Rec
ro	0.90	0.07	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.90
pl	0.01	0.89	0.02	0.03	0.01	0.02	0.01	0.00	0.00	0.00	0.00	0.89
bu	0.00	0.08	0.86	0.02	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.86
fe	0.01	0.07	0.16	0.52	0.23	0.00	0.01	0.00	0.00	0.00	0.00	0.52
ca	0.00	0.01	0.01	0.00	0.97	0.00	0.00	0.00	0.00	0.00	0.00	0.97
tr	0.01	0.15	0.02	0.01	0.01	0.79	0.01	0.00	0.00	0.00	0.00	0.79
po	0.01	0.06	0.02	0.01	0.02	0.14	0.74	0.01	0.00	0.00	0.00	0.74
si	0.00	0.11	0.02	0.11	0.02	0.09	0.20	0.43	0.00	0.03	0.00	0.43
bi	0.01	0.06	0.03	0.05	0.26	0.00	0.01	0.00	0.53	0.01	0.03	0.53
pe	0.01	0.11	0.11	0.07	0.05	0.12	0.05	0.00	0.01	0.44	0.03	0.44
ri	0.01	0.03	0.00	0.01	0.06	0.05	0.03	0.00	0.03	0.18	0.61	0.61
wPre	0.92	0.54	0.68	0.62	0.59	0.64	0.69	0.96	0.92	0.67	0.92	

图 4.8 SqueezeSegV3 混淆矩阵

混淆矩阵的每列反映了预测为某类的点可能由哪些类别混入, 是一种被混淆性, 可以计算每一类别的被混淆率 (BCR, Be Confused Rate):

$$BCR(c, i) = \frac{|y_{pd} = c \cap y_{gt} = i|}{y_{pd} = c} \quad (4.4)$$

其中 $BCR(c,i)$ 代表预测为 c 类的点原本真值是 i 类的概率。

WPR 和 BCR 关联性较强, 通过可视化这两个指标 (图4.9) 可以看到, Cylinder3D 在性能上总体优于 SqueezeSegV3。Cylinder3D 的栅栏 (fe) 类别容易被错分到植被 (pl) 和建筑 (bu), 这也是 Cylinder3D 在栅栏这一大类别上表现不佳的原因; 树干 (tr) 类别容易错分到植被 (pl) 类别, 这两个类别的本身语义有一定的重合; 信号牌 (si) 类别容易被错分到电线杆 (po) 类别, 这两个类别本身比较相近。SqueezeSegV3 的栅栏和四个小类别都容易被错分。两个模型也有一些共性: 植被类别 (pl) 虽然本身不容易被错分, 但其它类别的点容易混入这一类别, 信号牌 (si) 类别总是容易被错分。

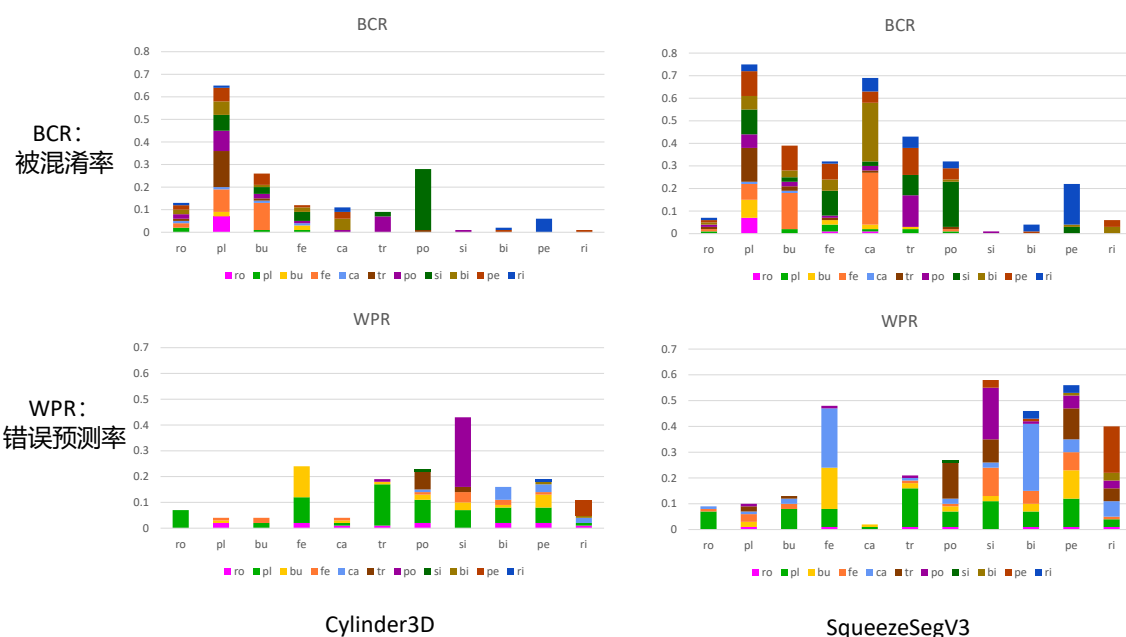


图 4.9 Cylinder3D 和 SqueezeSegV3 的 BCR 与 WPR

4.4.3 训练过程中的不确定性变化

这里观察性能较好的 Cylinder3D 模型的两类不确定性在训练过程中的变化。不确定性的估计方式是集成模型 Dropout, 即开启网络最后的 Dropout 层, 前传 10 次通过4.1, 4.2计算。

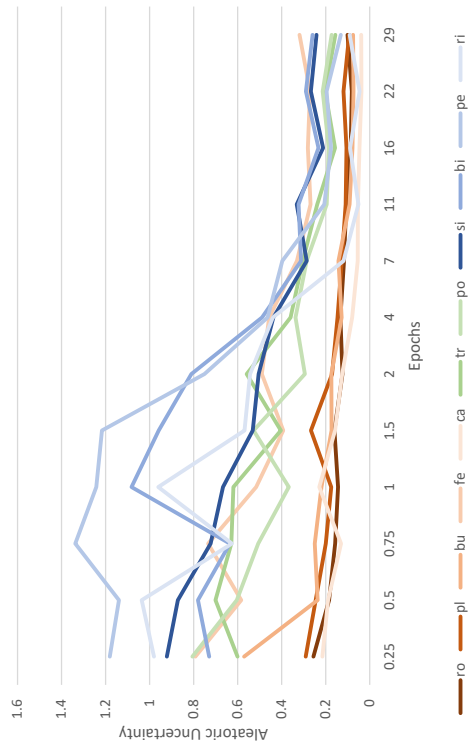


图 4.10 Cylinder3D 在训练过程中的数据不确定性变化曲线

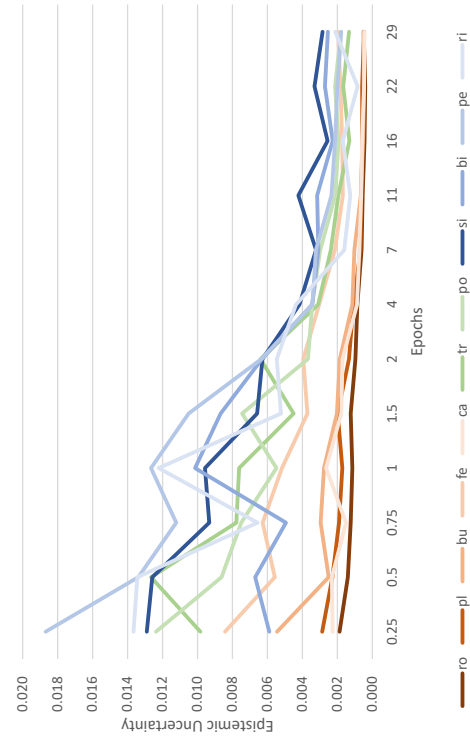


图 4.11 Cylinder3D 在训练过程中的模型不确定性变化曲线

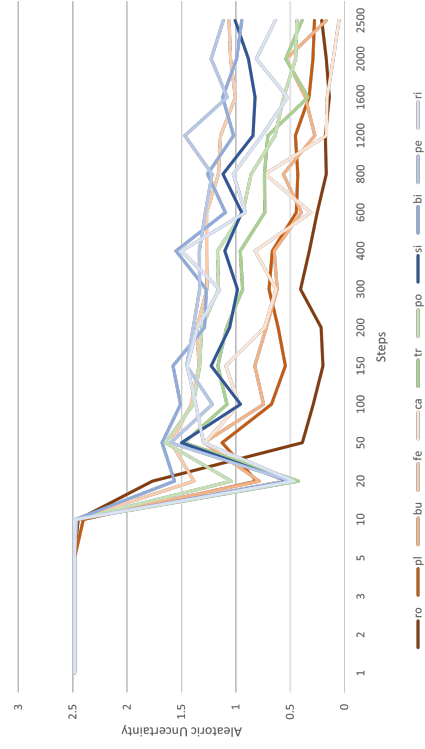


图 4.12 Cylinder3D 训练最初几步的数据不确定性变化图

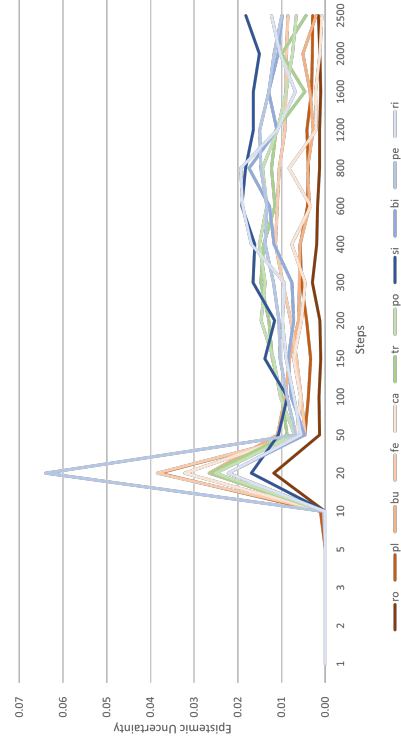


图 4.13 Cylinder3D 训练最初几步的模型不确定性变化图

从图4.10、4.11可以看出，Cylinder3D 在训练过程中通过这种方式得到的两类不确定性估计已经差了两个数量级，从公式出发，总体数量级的差距反映的是在随机丢失最后一层 20% 的参数时，模型多次前传几乎对预测结果没有影响，也即模型已经具备某种程度上的鲁棒性。

如2.2节所述，数据不确定性与模型本身无关，反映的是来自于数据的噪声等，模型不确定性随着训练数据的增加和模型性能的提升应该逐渐降低。随着模型的不间断训练，理想的数据不确定性应该保持不变，理想的认知不确定性应该不断下降，而事实是前期的训练数据不确定性和模型不确定性都在快速下降，后期波动稍有降低。可以看到如此建模获取的不确定性与物理直观仍有差距。造成这一现象主要有三个原因：1、Dropout 的不确定性估计方式与模型本身高度耦合，数据不确定性由模型估计，不可避免的含有模型的成分；2、集成模型估计公式2.29成立的前提是：模型已经训练完备，模型能够描述参数空间；3、在最初几代，随着模型性能的快速上升，对自己的输出也逐渐自信，反映在最后的概率向量，两类不确定性都在快速下降。

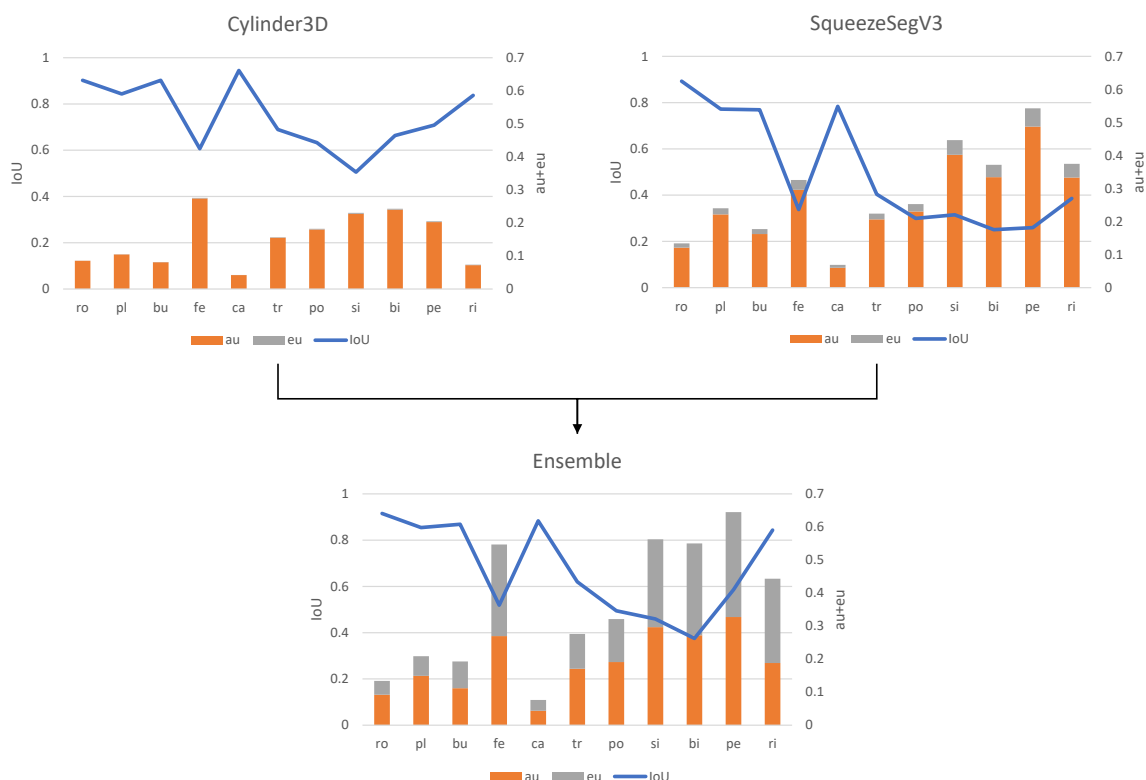


图 4.14 两类不确定性与模型性能在各类别上的表现曲线图

虽然绝对值与物理直观有差距，但是同一时刻各个类别的相对水平还是符合直观的。图中橙色系的曲线代表训练中的大类别，绿色系是训练中的中类别，蓝色系是训练中的小类别。在训练中的趋势是大类别的不确定性是相对最小的。除了栅栏 (fe) 类

别的数据不确定性一直较大。这反映的正是栅栏类别内的多样性，让模型在训练时就难以判断准确，也最终导致了在栅栏类上的 IoU 指标表现并不好。小类别的不确定性是相对最大的，只有骑手 (ri) 类别的数据不确定性在最后降到了和大类别同一水平这反映的是 Cylinder3D 在 SemanticKITTI 数据集上学习到了显著的骑手类别特征，这也最终导致了在骑手类上的 IoU 指标表现很好。

两类不确定性在训练到 10 代之后就下降到一个稳定水平，并上下波动，为了更明显地观察趋势，图4.10、4.11的横坐标并不均匀。本实验也尝试观察训练最初几步 (step) 的不确定性变化。如图4.12、4.13，模型训练在训练初期 AU、EU 表现出未定义行为 (实验过程中一个 epoch 有 2776 个 step，这里是最初 200 个 step)。此现象的解释是早期网络还处于随机化状态，模型是一致预测不好的，也即 EU 趋于 0，AU 趋于最大。随着训练到一定时期，网络具备表征能力。此时的 AU、EU 的估计才有意义。并且集成模型本身要求模型参数属于真实物理模型参数空间，即 $M_i \sim P(M|\mathbb{D})$ ，因此初期的 AU、EU 评估意义不大。

4.4.4 各类别的不确定性估计

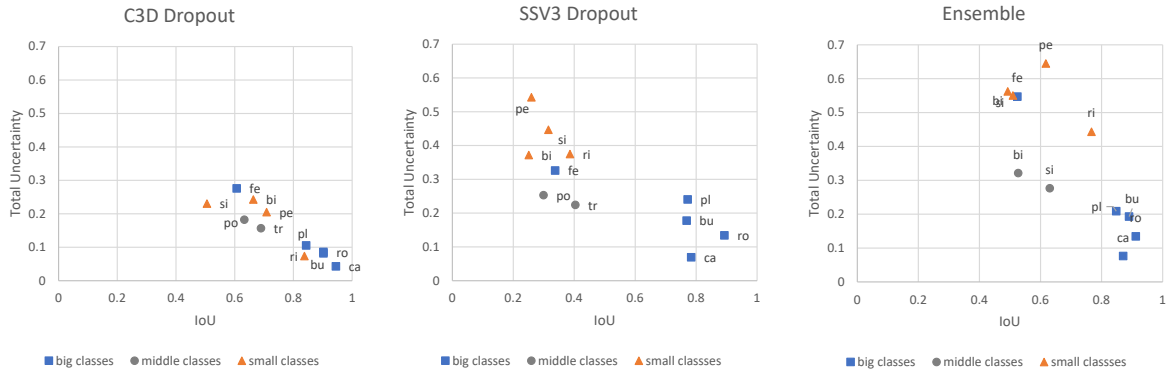


图 4.15 总不确定性与模型性能上的表现散点图

在模型训练完后，我们可以将两类不确定性和模型的交并比 (IoU) 性能进行对比。如图4.14，Cylinder3D 两种不确定性数量级的差距较明显。在 IoU 表现较差的信号牌、栅栏类别上，总的不确定性高，在 IoU 表现较好的汽车、道路、建筑、植物和骑手类别上，总的不确定性低。性能相对差一些的 SqueezeSegV3 模型对各个类别的模型不确定性都要高一些。在大类别上，两个模型的不确定性估计非常相似，栅栏类的不确定性都较高。两个模型最大的差别在四个小类别上：Cylinder3D 在四个小类别上的平均 IoU 比 SqueezeSegV3 高出一倍多，SqueezeSegV3 在四个小类别上的平均总不确定性也比 Cylinder3D 高出一倍多。

Cylinder3D 的 mIoU 为 74.9%，SqueezeSegV3 的 mIoU 为 49.7%，在将两种模型集成后，mIoU 为 69%。此时的模型不确定性在栅栏类别和小类别上都大幅增加，与数据不确定性达到同一量级。总体的不确定性趋势与 SqueezeSegV3 表现类似。

如图4.15, 可以直观地看出模型输出的不确定性与模型的性能呈负相关，集成模型度量的不确定性确实与模型对自己输出的把握有一定相关性。

4.4.5 三维点云不确定性可视化

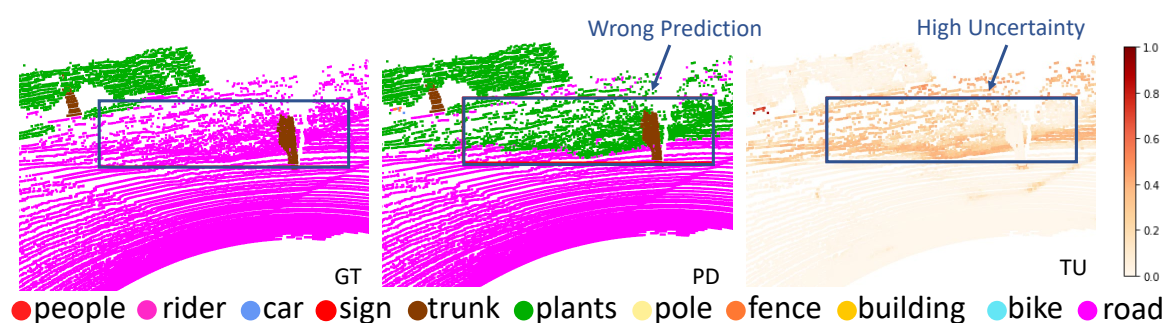


图 4.16 第 100 帧某视角的点云二维可视化图

其中，左为真值 (GT, Groundtruth); 中为预测 (PD, Prediction); 右为总不确定性 (TU, Total Uncertainty), 其值归一化并可视化到每一点上, 颜色越深代表不确定性越高。

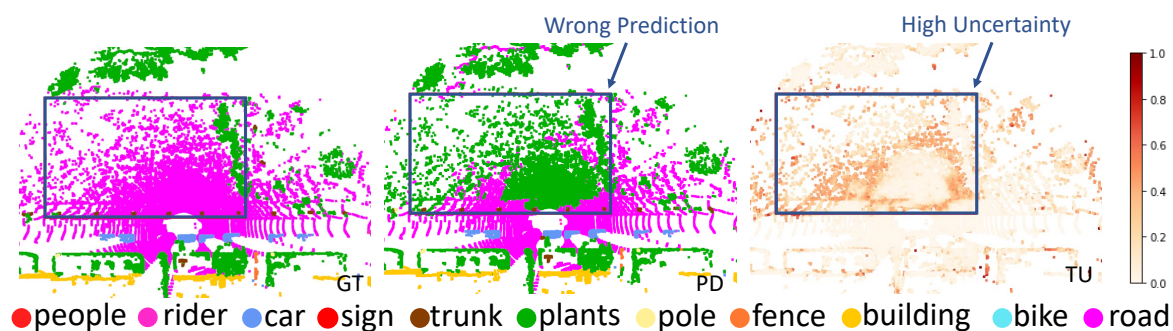


图 4.17 第 150 帧点云俯视图

为直观可视化不确定性在点云上的表现，论文展开个例研究 (Case Study)。如图4.16, 是 Cylinder3D 模型对第 100 帧某视角点云的二维可视化。可以看到在红框中的点云虽然真值是道路，但因为和植被的形态表现较接近，网络错误的将其中一些点预测为植被，此时网络自知有猜测成分，总不确定性较高。

同样的例子如图4.17, 在画红框区域，真值掺杂了道路与植被。此时仅依靠点云数据的形态两类区分确实较为困难。模型在猜测两类点的同时也给出了较高的总不确定性。

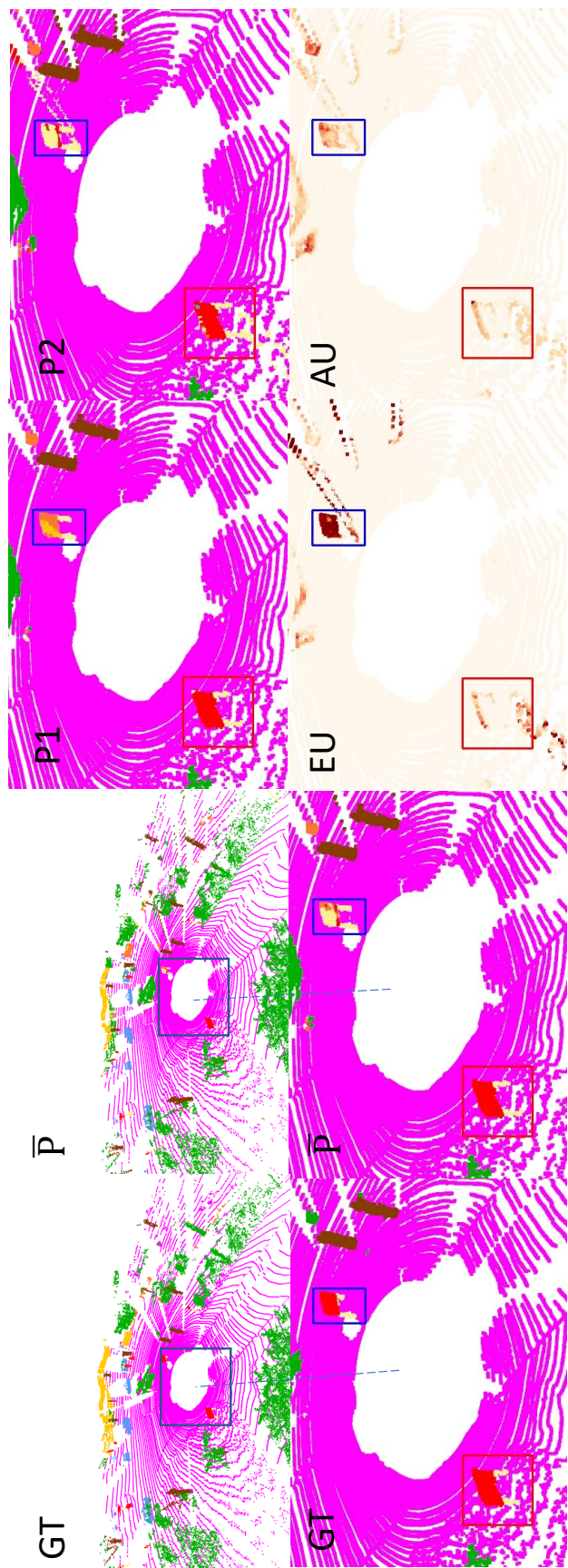


图 4.18 第 0 帧某视角的点云二维可视化图
其中， $\bar{P}$ 为两模型集成后的预测结果，GT 为真值。重点关注图片中两个信号牌的部分，P1 为 Cylinder3d 的预测结果，P2 为 SqueezeSegV3 的预测结果，EU 为集成后的模型不确定性，AU 为集成后的数据不确定性。

本论文接着关注单模型 Dropout 和两模型集成的结果。如图4.18，对比场景中的两个信号灯部分。靠左边的信号灯离传感器更近，两模型分割精度都较高。而右边的信号灯在两个模型与集成后的预测结果都表现不好。在不确定性指标上，两类不确定性相关性高，集成后左边信号灯的不确定性小，但边界处的不确定性大。右边信号灯整体的模型不确定性非常大，表现模型间分歧大；数据不确定性大，表现数据本身较难理解。

4.4.6 不同数据域分布上的不确定性表现

我们将在 KITTI 数据集上训练好的模型迁移到 POSS 数据集，因为数据集本身特性的不同，语义分割的性能出现了大幅下降。Cylinder3D 的 mIoU 从 72% 降至 34%，SqueezeSegV3 从 50% 降至 24%，两种模型的 Ensemble 从 69% 降至 34%。数据集原文^[59]使用的基线模型是 Pointnet++，迁移时从 30% 的 mIoU 降到了 13%。对比这一降幅，迁移的表现是合理的。随机选取一帧 Cylinder3D 的预测结果进行可视化并与真

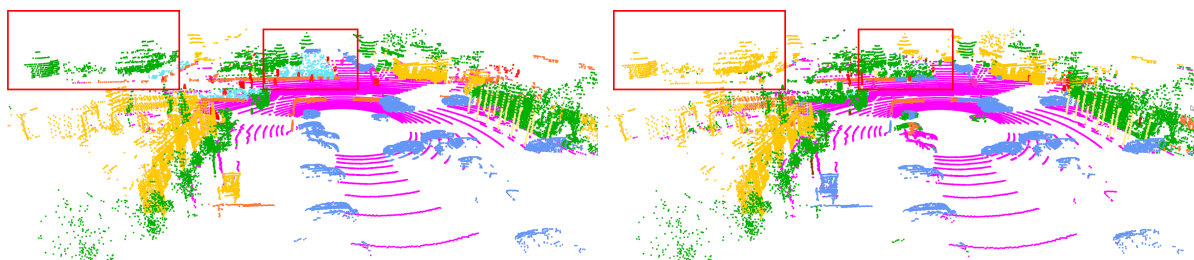


图 4.19 Cylinder3D 在 SemanticKITTI 训练后在 SemanticPOSS 上的表现

值对比发现，虽然 mIoU 大幅下降，但全场景的分割结果还是可观的，大部分的道路、车辆和植被还是分类正确。图4.19中两个画红框的区域展现了模型分类错误的地方：一个是远处的植被，模型错误地分类成了建筑；另一个是天蓝色的自行车，模型分类成了植被。事实上，自行车这一类别在两个数据集上的差异非常明显：KITTI 数据集不会有大片停放的自行车，而 POSS 中很容易出现。

在这样的新场景中，我们观察各类别的不确定性。如图4.20，不论是 Dropout 还是 Ensemble，不论是数据不确定性还是模型不确定性，绝对数值相较图4.14都有大幅增加。三种模型在两类不确定性的相对数量级与 KITTI 上的结果基本一致。三种模型在栅栏和四个小类别上表现相对较差，不确定性也相对较高。

如图4.21，各类别的总体不确定与模型性能仍呈负相关趋势。

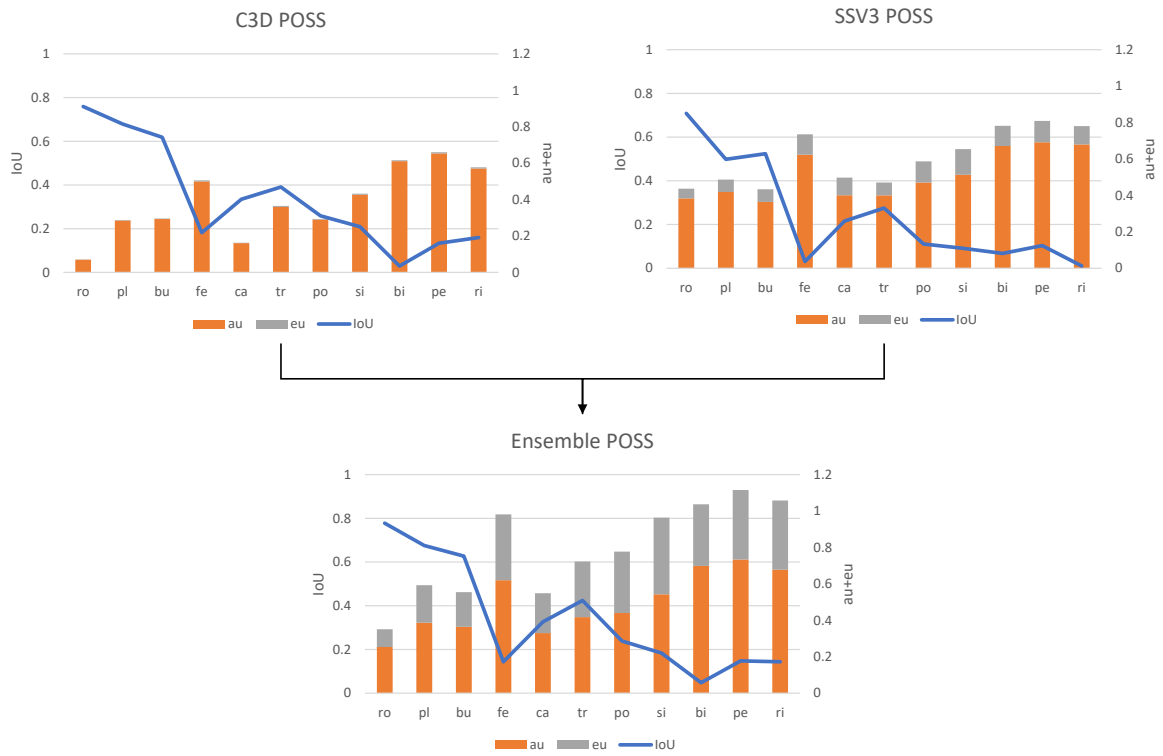


图 4.20 数据域变化后，两类不确定性与模型性能在各类别上的表现曲线图

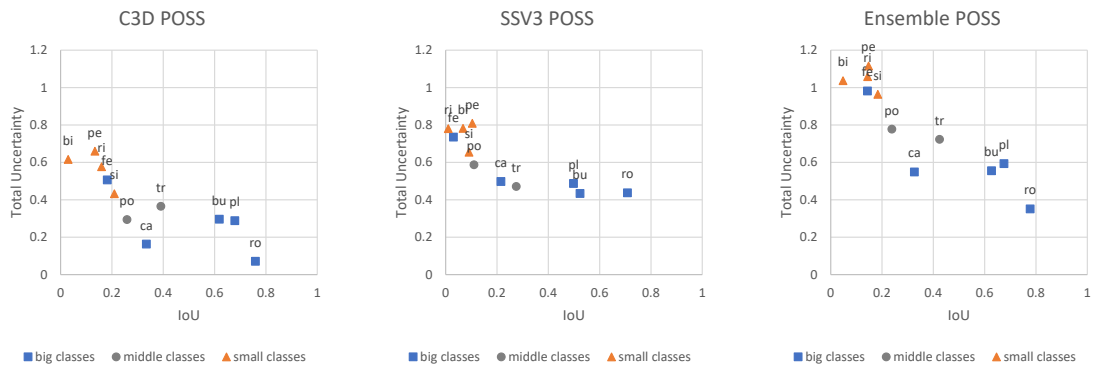


图 4.21 数据域变化后，总不确定性与模型性能上的表现散点图

第五章 总结与展望

5.1 总结

室外场景的三维点云语义分割是自动驾驶感知环节的重要任务。深度模型在特征不突出的或者数据量较少的类别上表现出性能下降。深度模型普遍在距离较远处的点上出现性能下降。深度模型容易将其它类别混入植物类别，信号类别的点因为类内多样性容易被错分。在数据集迁移的时候，深度模型容易出现性能下降，表明对训练数据域的过拟合问题依旧严重。

在不确定性度量问题上，由于三维点云语义分割任务模型相比简单的分类任务要复杂，数据量更大。在充分训练后，随机失活（Dropout）方式产生的集成模型结构相似，模型不确定性相对数据不确定性小。总的不确定性与模型性能负相关，即在不确定性大的数据上，模型性能往往较差。集成不同种类模型的数据不确定性和模型不确定性数量级相同，总不确定性与模型性能也呈负相关。两种方式在面对未见过的数据时，数据不确定性和模型不确定性都会增加。

但同时，目前的不确定性估计方法有以下局限：一是获取的数据不确定性和模型不确定性与物理直观差距大，二是数据不确定性和模型不确定性的耦合，三是无法检验不确定性衡量的准确，从而无法知道如何校准。

针对第二个问题，数据噪声与模型认知从概念上也有相关性。譬如，色盲对颜色信息差距大的输入不敏感，对尖锐形状敏感的人对棱角特征敏感。有文献^[27]也认为模型不确定性与数据不确定性本应耦合：以输入为图片，输出为方向控制的问题为例，在原本模型充分学习的图片输入上加噪声，此时的图片变为了未见过的稀有数据，模型不确定性自然也随数据不确定性上升。

5.2 展望

根据已有调研与实验结果，未来关于不确定性的研究可以从以下几个方面进行：

(1) 不确定性统一量化评估校准问题。有研究认为 NLL 指标可以一定程度反映不确定性的性能^[27, 31]。进一步的，也可以通过不确定性对后续任务如 OOD 任务的影响^[36]，即通过不确定性是否能有效提升后续任务表现来衡量得到的不确定性的有效性。

(2) 三维点云语义分割任务尝试集成更多种类的模型。在简单任务上，通过不同初始化，不同数据训练顺序得到的模型集成能有较好的效果^[25]。但在不同数据任务上同样 Ensemble 方式表现效果有很大区别，点云分割相对简单分类任务数据量大。如不同

初始化的 SqueezeSegV3 在训练中几乎有一样的损失函数变化曲线，最后训练的性能也几乎一致，很难相信几乎一致的模型集成有好的不确定性度量。相较之下，在不同训练集上训练的相同模型，在 loss 曲线，mIoU 曲线上的差异会大一些。未来可以进一步消融实验，探究 Ensemble 方式与数量变化时，不确定性估计的区别。

(3) 换用其它不确定性评估方法。因为集成模型需要有差异的多个模型，这让统一部署难度增大，训练测试时间变长。一类基于 ADF 的不确定性评估算法^[27]不需要修改原有算法，在部署贝叶斯版本的模型框架基础上，可以度量两类不确定性。可以进一步调研这一算法，并与更多的不确定性获取方式比较。

(4) 探究获取模型置信度的相关算法。本文更关注不确定性本身，对关于模型置信度的算法未进行深入分析。2.4节阐述了置信度与不确定性的关系，置信度相关算法在度量校准方式上比不确定性有天然优势，未来也可以尝试从置信度的角度探讨模型对输出的确信程度。

参考文献

- [1] Cover T M. Elements of information theory[M]. John Wiley & Sons, 1999.
- [2] Williams C K, Rasmussen C E. Gaussian processes for machine learning[M]. MIT press Cambridge, MA, 2006.
- [3] Neal R M. Bayesian learning for neural networks[M]. Springer Science & Business Media, 2012.
- [4] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017, 30.
- [5] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[J]. ArXiv preprint arXiv:2010.11929, 2020.
- [6] Silver D, Schrittwieser J, Simonyan K, et al. Mastering the game of go without human knowledge[J]. Nature, 2017, 550(7676):354-359.
- [7] Tesla crash preliminary evaluation report[R]. U.S. Department of Transportation, National Highway Traffic Safety Administration, 2017.
- [8] Graves A. Practical variational inference for neural networks[J]. Advances in neural information processing systems, 2011, 24.
- [9] Blundell C, Cornebise J, Kavukcuoglu K, et al. Weight uncertainty in neural network[C]//International conference on machine learning. 2015:1613-1622.
- [10] Srivastava N, Hinton G, Krizhevsky A, et al. Dropout: a simple way to prevent neural networks from overfitting[J]. The journal of machine learning research, 2014, 15(1):1929-1958.
- [11] Gal Y, Ghahramani Z. Dropout as a bayesian approximation: Representing model uncertainty in deep learning[C]//International conference on machine learning. 2016:1050-1059.
- [12] Cortinhal T, Tzelepis G, Erdal Aksoy E. Salsanext: Fast, uncertainty-aware semantic segmentation of lidar point clouds[C]//Advances in Visual Computing: 15th International Symposium, ISVC 2020, San Diego, CA, USA, October 5–7, 2020, Proceedings, Part II 15. 2020:207-222.
- [13] Gast J, Roth S. Lightweight probabilistic deep networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018:3369-3378.
- [14] Gal Y. Uncertainty in deep learning[D]. Department of Engineering University of Cambridge, 2016.
- [15] Blei D M, Kucukelbir A, McAuliffe J D. Variational inference: A review for statisticians[J]. Journal of the American statistical Association, 2017, 112(518):859-877.
- [16] Welling M, Teh Y W. Bayesian learning via stochastic gradient Langevin dynamics[C]//Proceedings of the 28th international conference on machine learning (ICML-11). 2011:681-688.

- [17] Chib S, Greenberg E. Understanding the metropolis-hastings algorithm[J]. The american statistician, 1995, 49(4): 327-335.
- [18] Hernández-Lobato J M, Adams R. Probabilistic backpropagation for scalable learning of bayesian neural networks[C]//International conference on machine learning. 2015: 1861-1869.
- [19] Depeweg S, Hernández-Lobato J M, Doshi-Velez F, et al. Learning and policy search in stochastic dynamical systems with bayesian neural networks[J]. ArXiv preprint arXiv:1605.07127, 2016.
- [20] Hernandez-Lobato J, Li Y, Rowland M, et al. Black-box alpha divergence minimization[C]//International conference on machine learning. 2016: 1511-1520.
- [21] Kendall A, Gal Y. What uncertainties do we need in bayesian deep learning for computer vision?[J]. Advances in neural information processing systems, 2017, 30.
- [22] Malinin A. Uncertainty estimation in deep learning with application to spoken language assessment[D]. Department of Engineering University of Cambridge, 2019.
- [23] Abdar M, Pourpanah F, Hussain S, et al. A review of uncertainty quantification in deep learning: Techniques, applications and challenges[J]. Information Fusion, 2021, 76: 243-297.
- [24] Kendall A, Gal Y, Cipolla R. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7482-7491.
- [25] Lakshminarayanan B, Pritzel A, Blundell C. Simple and scalable predictive uncertainty estimation using deep ensembles[J]. Advances in neural information processing systems, 2017, 30.
- [26] Depeweg S, Hernandez-Lobato J M, Doshi-Velez F, et al. Decomposition of uncertainty in Bayesian deep learning for efficient and risk-sensitive learning[C]//International Conference on Machine Learning. 2018: 1184-1193.
- [27] Loquercio A, Segu M, Scaramuzza D. A general framework for uncertainty estimation in deep learning[J]. IEEE Robotics and Automation Letters, 2020, 5(2): 3153-3160.
- [28] Gal Y, Islam R, Ghahramani Z. Deep bayesian active learning with image data[C]//International conference on machine learning. 2017: 1183-1192.
- [29] Kendall A, Badrinarayanan V, Cipolla R. Bayesian segnet: Model uncertainty in deep convolutional encoder-decoder architectures for scene understanding[J]. ArXiv preprint arXiv:1511.02680, 2015.
- [30] Dietterich T G. Ensemble methods in machine learning[C]//Multiple Classifier Systems: First International Workshop, MCS 2000 Cagliari, Italy, June 21–23, 2000 Proceedings 1. 2000: 1-15.
- [31] Guo C, Pleiss G, Sun Y, et al. On calibration of modern neural networks[C]//International conference on machine learning. 2017: 1321-1330.
- [32] Hendrycks D, Gimpel K. A baseline for detecting misclassified and out-of-distribution examples in neural networks[J]. ArXiv preprint arXiv:1610.02136, 2016.

- [33] Rahaman R, et al. Uncertainty quantification and deep ensembles[J]. *Advances in Neural Information Processing Systems*, 2021, 34: 20063-20075.
- [34] Zadrozny B, Elkan C. Transforming classifier scores into accurate multiclass probability estimates[C]//*Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*. 2002: 694-699.
- [35] Zhang J, Kailkhura B, Han T Y J. Mix-n-match: Ensemble and compositional methods for uncertainty calibration in deep learning[C]//*International conference on machine learning*. 2020: 11117-11128.
- [36] Ovadia Y, Fertig E, Ren J, et al. Can you trust your model’s uncertainty? evaluating predictive uncertainty under dataset shift[J]. *Advances in neural information processing systems*, 2019, 32.
- [37] Li Y, Ibanez-Guzman J. Lidar for autonomous driving: The principles, challenges, and trends for automotive lidar and perception systems[J]. *IEEE Signal Processing Magazine*, 2020, 37(4): 50-61.
- [38] Wu B, Wan A, Yue X, et al. Squeezeseg: Convolutional neural nets with recurrent crf for real-time road-object segmentation from 3d lidar point cloud[C]//*2018 IEEE international conference on robotics and automation (ICRA)*. 2018: 1887-1893.
- [39] Broggi A, Grisleri P, Zani P. Sensors technologies for intelligent vehicles perception systems: A comparison between vision and 3D-LIDAR[C]//*16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*. 2013: 887-892.
- [40] Behley J, Garbade M, Milioto A, et al. Semantickitti: A dataset for semantic scene understanding of lidar sequences[C]//*Proceedings of the IEEE/CVF international conference on computer vision*. 2019: 9297-9307.
- [41] Gao B, Pan Y, Li C, et al. Are we hungry for 3D LiDAR data for semantic segmentation? A survey and experimental study[J]. *ArXiv preprint arXiv:2006.04307*, 2020.
- [42] Xu C, Wu B, Wang Z, et al. Squeezesegv3: Spatially-adaptive convolution for efficient point-cloud segmentation[C]//*Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXVIII 16*. 2020: 1-19.
- [43] Milioto A, Vizzo I, Behley J, et al. Rangenet++: Fast and accurate lidar semantic segmentation[C]//*2019 IEEE/RSJ international conference on intelligent robots and systems (IROS)*. 2019: 4213-4220.
- [44] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]//*Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. 2015: 234-241.
- [45] Kong L, Liu Y, Chen R, et al. Rethinking Range View Representation for LiDAR Segmentation[J]. *ArXiv preprint arXiv:2303.05367*, 2023.
- [46] Tang H, Liu Z, Li X, et al. TorchSparse: Efficient point cloud inference engine[J]. *Proceedings of Machine Learning and Systems*, 2022, 4: 302-315.

- [47] Zhu X, Zhou H, Wang T, et al. Cylindrical and asymmetrical 3d convolution networks for lidar segmentation[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021:9939-9948.
- [48] Çiçek Ö, Abdulkadir A, Lienkamp S S, et al. 3D U-Net: learning dense volumetric segmentation from sparse annotation[C]//Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19. 2016:424-432.
- [49] Graham B, Engelcke M, Van Der Maaten L. 3d semantic segmentation with submanifold sparse convolutional networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018:9224-9232.
- [50] Meng H Y, Gao L, Lai Y K, et al. Vv-net: Voxel vae net with group convolutions for point cloud segmentation[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2019:8500-8508.
- [51] Tchapmi L, Choy C, Armeni I, et al. Segcloud: Semantic segmentation of 3d point clouds[C]//2017 international conference on 3D vision (3DV). 2017:537-547.
- [52] Lai X, Chen Y, Lu F, et al. Spherical Transformer for LiDAR-based 3D Recognition[J]. ArXiv preprint arXiv:2303.12766, 2023.
- [53] Qi C R, Yi L, Su H, et al. Pointnet++: Deep hierarchical feature learning on point sets in a metric space[J]. Advances in neural information processing systems, 2017, 30.
- [54] 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.
- [55] Li J, Chen B M, Lee G H. So-net: Self-organizing network for point cloud analysis[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018:9397-9406.
- [56] Wu W, Qi Z, Fuxin L. Pointconv: Deep convolutional networks on 3d point clouds[C]//Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition. 2019:9621-9630.
- [57] Hu Q, Yang B, Xie L, et al. Randla-net: Efficient semantic segmentation of large-scale point clouds[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020:11108-11117.
- [58] Kirillov A, Mintun E, Ravi N, et al. Segment anything[J]. ArXiv preprint arXiv:2304.02643, 2023.
- [59] Pan Y, Gao B, Mei J, et al. Semanticpos: A point cloud dataset with large quantity of dynamic instances[C]//2020 IEEE Intelligent Vehicles Symposium (IV). 2020:687-693.
- [60] Caesar H, Bankiti V, Lang A H, et al. Nuscenes: A multimodal dataset for autonomous driving[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020:11621-11631.

- [61] Everingham M, Van Gool L, Williams C K, et al. The pascal visual object classes (voc) challenge[J]. International journal of computer vision, 2010, 88: 303-338.
- [62] Pan Y, Zhao H. Re-Thinking classification confidence with model quality quantification[Z]. 2023.

致谢

终于有一天，我居然也写到了这一页。

想想致谢应该从什么时候写起吧。既然这里是一片随我发挥的地方，那就从十六年前，小学一年级。关于小学的记忆不多了，但我清楚从一年级开始我就住校了。也许是这份独立，让我甚至直到四年前，当别人问起“你觉得最要感谢的人是谁”这样的问题时，我会不假思索的回答是我自己。是自己咬牙坚持下来的苦难。只因当时的“任务”只有学。每每看到采访时别人回答“我首先感谢的是我的父母，是我的亲人”，我内心就在嘀咕：装什么装。

这里我首先感谢的是我的父母，是我的亲人。这四年来我才意识到父母对自己的爱是真正无条件的。那两个真正在全世界都背叛你的时候支持你的人，对我来说是我的父母。当步入社会，人的目标不再只有单纯地学，身边最稳定的因素就是父母，然后是亲人。我总能在心情低落的时候找他们倾诉，得到慰藉。

小的时候觉得长大就轻松了。初中觉得高中就自由了，高中觉得大学就自由了，大一觉得大四就自由了，现在才发现，人总是越来越忙。天天轻松不劳而获是不可能的。短暂的快乐可能带来长久的痛苦，而短暂的痛苦才能带来长久的快乐。浑浑噩噩混日子，在梦醒时分会感到无限的恐惧与痛苦；踏踏实实追梦想，才会在闲暇之际感到充实的安心与快乐。

这里我接着感谢我的导师赵卉菁老师。在这里我有了可以踏踏实实追梦想的条件。感谢赵老师给我提供的资源与帮助。感谢赵老师的宽容、耐心与真诚。人总是贪婪地对已经拥有的东西不以为意，羡慕别人的条件与成就。在与同龄人的谈心中才会发现自己身在福中不知福的傲慢。

在赵老师的组里我认识到了优秀的师兄，这里感谢潘彦成、朱泽宇、胡少驰、丁哲章、谢睿、姚子炀、王帅、高飙。万能的师兄们总是放下手头的事毫无保留地答疑解惑。不论是 bug 体质的我踩进什么坑时的专家会诊，还是心态上遇到挫折时的真心开导，我都受益匪浅。赵老师组里的每一位同学都是一座宝库，我甚至在担心没有了宝库该怎么办。师门聚餐时我也在怀疑我怎么可能在若干年后变成这样优秀的师兄。

从学习转变为科研后，工作上常常和师兄们交流，生活中也许未来是单人单间。我有时觉得单人单间并不比多人宿舍优越。推开门时的热情会化作空荡荡的寒舍。时事与消息的来源会变得单一，心态的交流与认同会受到阻碍。

这里感谢 37 楼 201 的室友周孙鹏，刘玮明，解博杰。两南两北五湖四海的组合总是需要包容与互助。虽然我不喜欢待在寝室，但每天回寝室时都充满了热情。不论是

夜谈，推销，还是送伞，拼水果捞。虽然人与人的关系不可能一直高位运行，但让记忆过滤掉不欢，留下美好与念想。

每个人都由他的朋友、他交流的人、他遇见的人定义。很幸运我的周围能一直围绕着优秀的人。

这里感谢我认识的所有老师，不论是否教过我，教过我时间的长短。感谢我的大学同学，不论是同院同班同课。还有我的高中初中小学同学，尤其是常常见面约饭约快乐的，以后还要继续啊。

一位教过我的数学老师、班主任、共产党员在班会的时候跟我们说，人最重要的是回馈社会。你回馈给社会的永远比不过社会提供给你的。

这里感谢所有的后勤人员，不论是学一早上记住我“哦，鸡汤面？”的大伯，还是理一偶然听到跟门卫说“反正我们都是服务同学嘛”的保洁，还是理教那位 11 点催人“早点回去休息”却还是留了门的保安，还是在宿舍、图书馆，校园的各个角落工作的螺丝钉。

最近深感我的记忆衰退的厉害，本次致谢我也尽量避免举例，而是用一种更抽象的概念去概括。总是担心有一天也许我认不出哪位故人了。感情总会随时间淡化，友情还是需要常常相见或交流来加深。打个广告，欢迎朋友们随时来约，我很乐意说走就走。

最后，感谢每一位帮助过我的人，不论这种帮助是大到解决了一个什么难题，还是小到一句鼓励的话。

还是用我最喜欢的一句话作为结尾：一切都结束了吗？不，这只是新的开始。

北京大学学位论文原创性声明和使用授权说明

原创性声明

本人郑重声明：所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不含任何其他个人或集体已经发表或撰写过的作品或成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本声明的法律结果由本人承担。

论文作者签名：周昊天

日期：2023年5月12日

学位论文使用授权说明

本人完全了解北京大学关于收集、保存、使用学位论文的规定，即：

- 按照学校要求提交学位论文的印刷本和电子版本；
- 学校有权保留学位论文的印刷本和电子版，并提供目录检索与阅览服务，在校园网上提供服务；
- 学校可以采用影印、缩印、数字化或其它复制手段保存论文；

论文作者签名：周昊天

导师签名：

刘畅

日期：2023年5月12日