UPPSALA
UNIVERSITET

Evaluating Scalable Bayesian Deep Learning Methods for Robust Computer Vision

Fredrik K. Gustafsson[1], Martin Danelljan[2], Thomas B. Schön[1]

[1]Department of Information Technology, Uppsala University, Sweden
[2]Computer Vision Lab, ETH Zürich, Switzerland

What is the problem?

While deep learning has become the go-to approach in computer vision, these models fail to properly capture the **uncertainty** inherent in their predictions. The approach of **Bayesian deep learning** aims to address this issue in a principled manner.

What is the problem?

While deep learning has become the go-to approach in computer vision, these models fail to properly capture the **uncertainty** inherent in their predictions. The approach of **Bayesian deep learning** aims to address this issue in a principled manner.

Predictive uncertainty is then decomposed into aleatoric and **epistemic** uncertainty. Estimating epistemic uncertainty, which accounts for uncertainty in the model parameters, can mitigate model over-confidence and is thus of great importance.

What is the problem?

While deep learning has become the go-to approach in computer vision, these models fail to properly capture the **uncertainty** inherent in their predictions. The approach of **Bayesian deep learning** aims to address this issue in a principled manner.

Predictive uncertainty is then decomposed into aleatoric and **epistemic** uncertainty. Estimating epistemic uncertainty, which accounts for uncertainty in the model parameters, can mitigate model over-confidence and is thus of great importance.

Epistemic uncertainty estimation is challenging, especially for *large-scale* models used in *real-world* computer vision tasks, but **scalable** methods have recently emerged.

What is the problem?

UPPSALA UNIVERSITET

While deep learning has become the go-to approach in computer vision, these models fail to properly capture the **uncertainty** inherent in their predictions. The approach of **Bayesian deep learning** aims to address this issue in a principled manner.

Predictive uncertainty is then decomposed into aleatoric and **epistemic** uncertainty. Estimating epistemic uncertainty, which accounts for uncertainty in the model parameters, can mitigate model over-confidence and is thus of great importance.

Epistemic uncertainty estimation is challenging, especially for *large-scale* models used in *real-world* computer vision tasks, but **scalable** methods have recently emerged.

The research community however lacks a common and comprehensive **evaluation framework** for such methods. Both researchers and practitioners are currently thus unable to properly assess and compare competing methods.

Our contributions

UPPSALA UNIVERSITET

We propose a comprehensive **evaluation framework** for *scalable* epistemic uncertainty estimation methods in deep learning. It is specifically designed to test the robustness (to out-of-domain inputs) required in **real-world** computer vision applications.

Our contributions

We propose a comprehensive **evaluation framework** for *scalable* epistemic uncertainty estimation methods in deep learning. It is specifically designed to test the robustness (to out-of-domain inputs) required in **real-world** computer vision applications.

Our proposed framework employs state-of-the-art models on the tasks of **depth completion** (regression) and **street-scene semantic segmentation** (classification).

Our contributions

We propose a comprehensive **evaluation framework** for *scalable* epistemic uncertainty estimation methods in deep learning. It is specifically designed to test the robustness (to out-of-domain inputs) required in **real-world** computer vision applications.

Our proposed framework employs state-of-the-art models on the tasks of **depth completion** (regression) and **street-scene semantic segmentation** (classification).

We provide an extensive and conclusive comparison of the two current state-of-the-art *scalable* methods: **ensembling** and **MC-dropout**, demonstrating that **ensembling** consistently provides more reliable and practically useful uncertainty estimates.

Our contributions

UPPSALA UNIVERSITET

We propose a comprehensive **evaluation framework** for *scalable* epistemic uncertainty estimation methods in deep learning. It is specifically designed to test the robustness (to out-of-domain inputs) required in **real-world** computer vision applications.

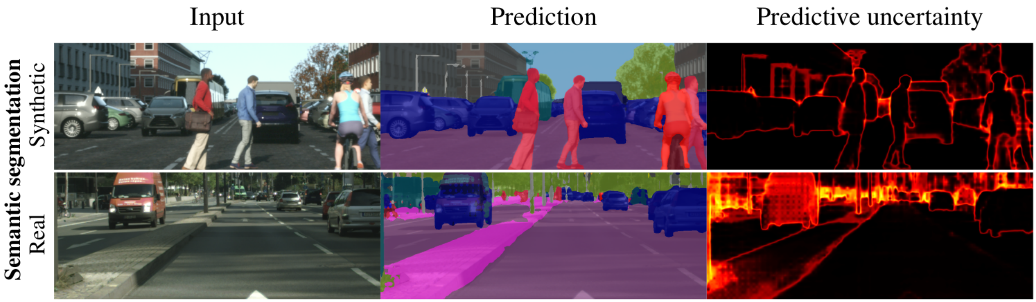
Our proposed framework employs state-of-the-art models on the tasks of **depth completion** (regression) and **street-scene semantic segmentation** (classification).

We provide an extensive and conclusive comparison of the two current state-of-the-art *scalable* methods: **ensembling** and **MC-dropout**, demonstrating that **ensembling** consistently provides more reliable and practically useful uncertainty estimates.

Publicly available code: `www.github.com/fregu856/evaluating_bdl`.

Street-scene semantic segmentation

Given an image $x \in \mathbb{R}^{h\times w\times 3}$, the task is to predict y of size $h \times w$, in which each pixel is assigned to one of C classes (road, car, etc.). Models are trained on **synthetic data** and evaluated on **real-world data**, testing robustness to **out-of-domain** inputs.



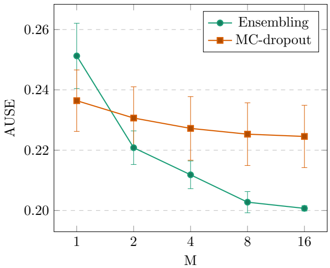
Metrics for evaluation of uncertainty estimation quality:

- **AUSE:** *relative* measure that reveals how well the estimated uncertainty can be used to sort predictions from worst (large true prediction error) to best.
- **ECE:** *absolute* measure in terms of calibration. A well-calibrated model is not over-confident (highly confident but incorrect predictions) nor over-conservative.

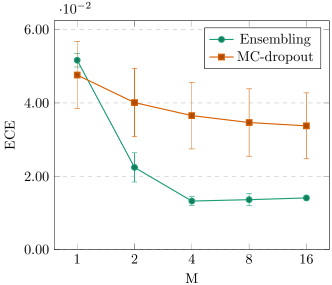
Street-scene semantic segmentation - Results

Metrics for evaluation of uncertainty estimation quality:

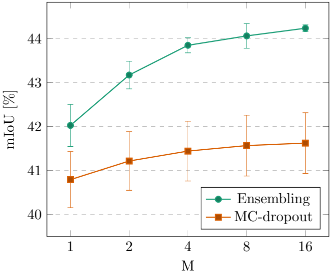
- **AUSE:** *relative* measure that reveals how well the estimated uncertainty can be used to sort predictions from worst (large true prediction error) to best.
- **ECE:** *absolute* measure in terms of calibration. A well-calibrated model is not over-confident (highly confident but incorrect predictions) nor over-conservative.



(a) AUSE.



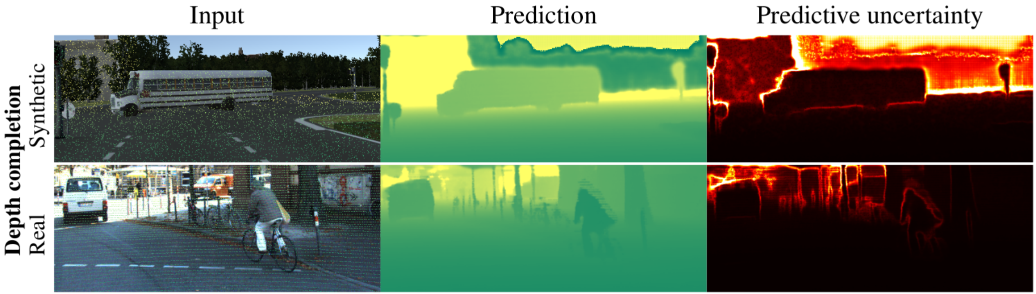
(b) ECE.



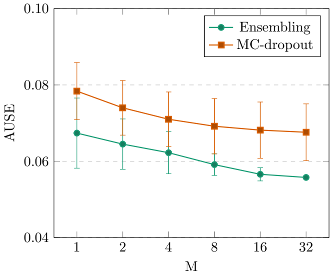
(c) mIoU.

Depth completion

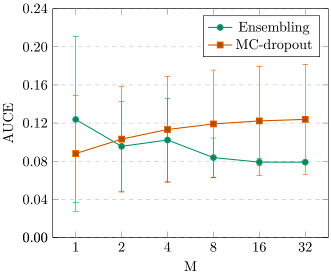
Given an image $x_{\text{img}} \in \mathbb{R}^{h \times w \times 3}$ and an associated *sparse* depth map, the task is to predict a *dense* depth map $y \in \mathbb{R}^{h \times w}$ of the scene. Models are trained on **synthetic data** and evaluated on **real-world data**, testing robustness to **out-of-domain** inputs.



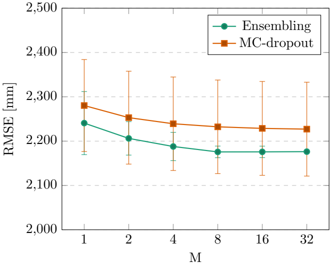
Metrics for evaluation of uncertainty estimation quality: **AUSE** and **AUCE** (generalization of ECE to the regression setting).



(a) AUSE.



(b) AUCE.



(c) RMSE.

Discussion & conclusion

Required **training** scales linearly with M for ensembling, but this is not a major concern in most safety-critical applications, such as automotive.

The main drawback of both ensembling and MC-dropout is instead the computational cost at **test time** that scales linearly with M, affecting real-time applicability.

Discussion & conclusion

UPPSALA UNIVERSITET

Required **training** scales linearly with M for ensembling, but this is not a major concern in most safety-critical applications, such as automotive.

The main drawback of both ensembling and MC-dropout is instead the computational cost at **test time** that scales linearly with M, affecting real-time applicability.

Our work suggests that **ensembling** should be considered the new go-to method for *scalable* epistemic uncertainty estimation. We attribute the success of ensembling to its ability to capture **multi-modality** in the posterior distribution $p(\theta|\mathcal{D})$.

Contact

UPPSALA UNIVERSITET

Fredrik K. Gustafsson

Uppsala University

fredrik.gustafsson@it.uu.se

www.fregu856.com