Uncertainties for Online Laplace Monocular Depth Estimation

Frederik Kølby Christensen

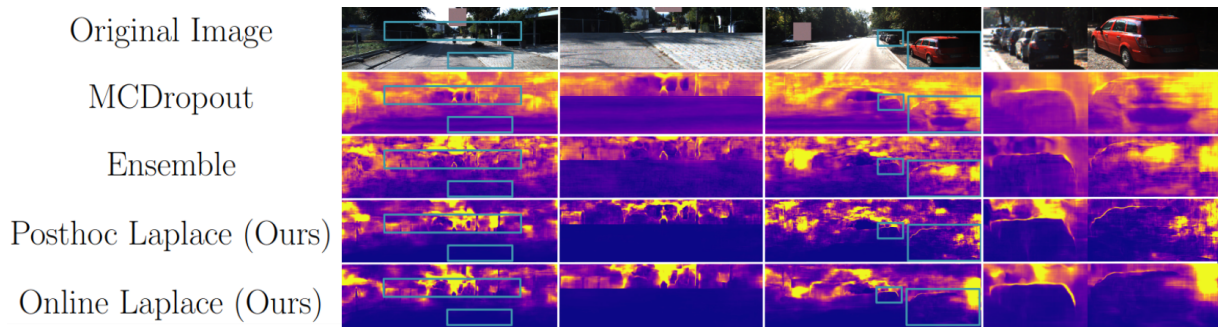

Figure 1: **Qualitative assessment of uncertainty.** Two different images, chosen at random. Brighter = higher uncertainty. A pink box is placed randomly on the images, to test models' awareness of OOD-data. It is seen that both proposed methods generate meaningful uncertainty estimates in that high uncertainties correspond to areas that are much harder for the eye to discern and low uncertainties correspond to areas that are commonly present in training data, such as roads or pavements. In comparison, MCDropout and Ensemble uncertainties are much more haphazard. Ensembles predict OOD better - however, this behaviour was inconsistent across runs. For predictions on same images, see figure 2.

Abstract

We introduce the Online and Posthoc Laplace Depth-estimator. The methods provide attractive computational complexity, and well-calibrated uncertainties. Furthermore, the Online Laplace method provides the best only monotone relationship between loss and uncertainty. The claims are verified through both quantitative experiments and qualitative observations.

1 Introduction

Reliable uncertainty quantification is of paramount importance in safety-critical or high-risk applications. As models play an increasing role in assisting medical staff or autonomous drivers, users must be able to weigh their own experience and confidence against the output from a model. This is especially crucial when the user and model disagree. Such a case might be an Oncologist believing a cancer is benevolent, while the model predicts a high probability of malignancy. Then the Oncologist must weigh their experience against the uncertainty of the model. However, such an interplay relies crucially on the model's ability to generate reliable estimates of its uncertainty. This task, however, is non-trivial as common methods either require a significant increase in computational resources (Ensembles [6]) or lead to degradation in performance (MCDropout [3]). One such safety-critical application is metric depth estimation from a single image, which is highly relevant to safe autonomous driving.

This report examines recent breakthroughs [10] in scalable, generalizable uncertainty quantification and improves on their methodology to attain feasible uncertainty estimates on large-scale datasets. Specifically, we tackle the problem of depth estimation, where we on KITTI obtain state-of-the-art uncertainties while retaining predictive power and requiring a similar or smaller training budget.

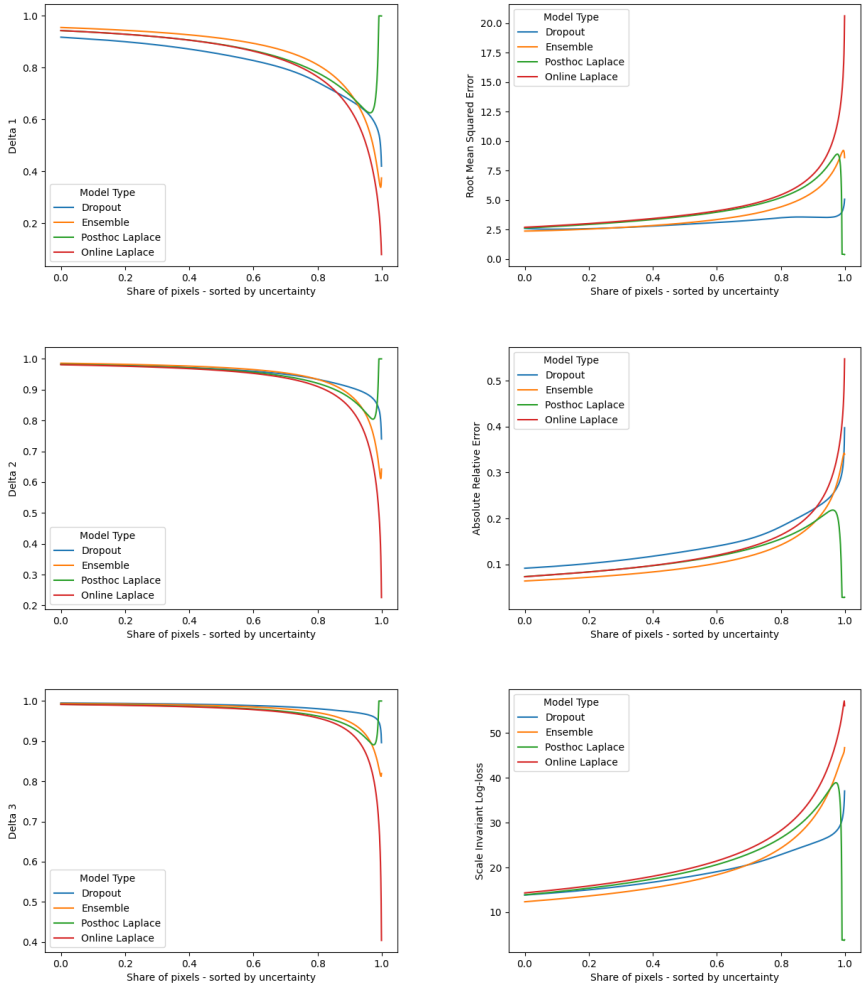

Figure 2: **Quantitative Uncertainties.** Only Online Laplace model uncertainty monotonically with error across metrics.

2 Background

In monocular depth estimation, the goal is to estimate a depth field based on a single image. As such, given an image x, a model f_θ, parametrized by the weight vector, θ, and a corresponding depth map y, we seek to minimize the loss given by the mean squared error

$$\mathcal{L}_\theta(y, x) = \|y - f_\theta(x)\|^2 \tag{2.1}$$

To obtain reliable estimates of uncertainty, we turn to a Bayesian interpretation. Rewriting the loss function using a second-order Taylor-expansion around a given weight vector, θ^*, we obtain the Laplace approximation:

$$\mathcal{L}_\theta \approx \mathcal{L}_{\theta^*} + (\theta - \theta^*)^\top \nabla_\theta \mathcal{L}_{\theta^*} + \frac{1}{2}(\theta - \theta^*)^\top \nabla^2_\theta \mathcal{L}_{\theta^*}(\theta - \theta^*) \tag{2.2}$$

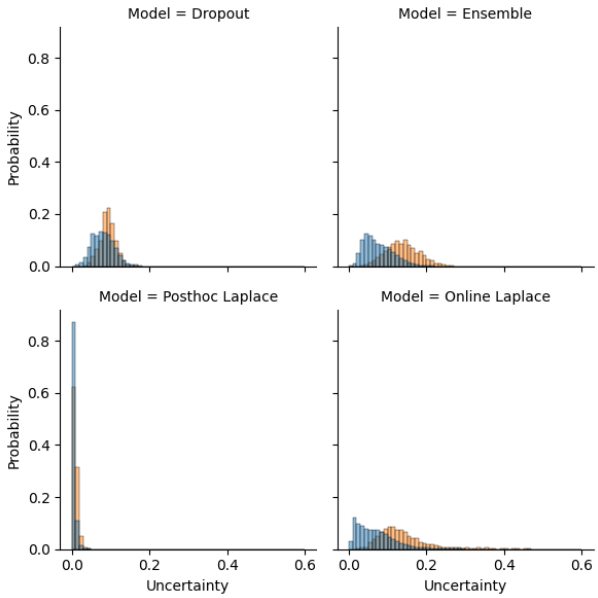

Figure 3: **Detection of OOD**: Only Online Laplace and Ensembles separate in distribution data and out-of-distribution data.

If we interpret our loss, $\mathcal{L}_\theta$, as the unnormalized negative log-posterior, then - if θ^* is a MAP-estimate - the second term vanishes, and the third is positive semi-definite, since $\mathcal{L}_\theta$ is convex. Since the Laplace Approximation is a second-order polynomial, this implies the assumption that the posterior is, in fact, Gaussian. Applying this approximation after training the model, one obtains Posthoc Laplace approximation [8]. More computationally demanding is the approach by Miani et al. (2022)[10], where they approximate the hessian at every training step, using this approximation to sample parameters and thus networks - they name this approach Online Laplace. Since they find a well-behaved posterior to be the result and state-of-the-art uncertainties - we build upon the Online Laplace to obtain intuitive, meaningful uncertainties in the realm of single image depth estimation.

3 Methodology

Given a dataset D consisting of images $x \in \mathbb{R}^{C\times H\times W}$ and depth labels $y \in \mathbb{R}^{1\times H\times W}$, we train a model, $f_\theta, \theta \in \Theta$ to obtain a local optimum w.r.t. parameters - denote these by θ^*. We then seek a posterior distribution over the model, $f_.$'s weights $\theta \in \Theta$, $p(\theta|D)$. We then get, by e.q. 2.2, that the posterior is given by

$$p(\theta|D) = \mathcal{N}\left(\theta, \mu = \theta^*, \Sigma = \left(\nabla^2_\theta \mathcal{L}_{\theta^*} + \sigma^{-2}_{\text{prior}}\mathbb{I}\right)^{-1}\right) \tag{3.1}$$

This is the Posthoc Laplace approximation. The approach is enticing - after obtaining a (pre-)trained model - one can just fit the covariance, and the model is now Bayesian. Unfortunately,

stochastic gradient-based training ends up exploring areas around local minima, rather than staying close to such minima, and θ^* ends up not being a MAP estimate. Even worse - the Hessian (and thus the covariance matrix) may change drastically during such exploration, leading to brittleness. [13]

To combat this, Miani et. al (2022) instead sample from the (at time step t best estimate of the) posterior,

$$q_{t,\theta} = \mathcal{N}\left(\theta, \mu = \theta_t, \Sigma = (H_{\theta_t} + \sigma_{prior}^{-2}\mathbb{I})^{-1}\right) \tag{3.2}$$

Where

$$\theta_{t+1} = \theta_t + \lambda \nabla_\theta \mathbb{E}_{\theta \sim q_t}[\mathcal{L}_\theta] \tag{3.3}$$

updates the parameters θ and

$$H_{\theta_{t+1}} = (1-\alpha)H_{\theta_t} + \nabla_\theta^2 \mathbb{E}_{\theta \sim q_t}[\mathcal{L}_\theta] \tag{3.4}$$

updates the hessian. The initialization follows the prior, $q^0 = \mathcal{N}(0, \sigma_{prior}^2)$. Instead of the actual hessian, we approximate it in two parts. First, we use the Generalized-Gauss-Newton approximation (GGN) [5, 8]. GGN approximates the hessian of a neural network, composed by $f \circ g$, where f is the neural network, and g is the loss function, by linearizing f. Letting J_θ denote the Jacobian of f w.r.t. θ, we get the hessian, as the loss is convex, gives the positive semi-definite matrix;

$$J_\theta^\top \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} J_\theta \tag{3.5}$$

To avoid letting the hessian scale quadratically, we approximate it with its diagonal, following [13]. Even though this approximation scales linearly, we find that approx. 80% of the time per step is spent calculating the hessian, and thus, inspired by [12], we update the matrix only once every 10 steps, as we find further updates unnecessary.

4 Experimental setup

4.1 Datasets

KITTI [7] is a depth-prediction dataset containing stereo images paired with corresponding LIDAR-depth maps captured by a car moving through daily traffic in Karlsruhe, Germany. Images are of size 376x1241, which we crop down to 352x1216 (following the official benchmark crop).

4.2 Model

The base model used is a convolutional UNet[2] with 5.5M parameters, implemented in the pytorch-based nnj library [9] (which implements efficient diagonal hessians), where the second-to-last layer is the sigmoid function. To this we add a final scale-and-shift layer rescaling the sigmoid output interval from $(0, 1)$ to $(0, 80)$. Our main comparisons are Ensembles [6] of which we train M=5, as well as MCDropout [3] (MCDropout p=0.05 - any higher and the model's predictive power severely deteriorated). MCDropout, Posthoc Laplace and Online Laplace are all evaluated with M=100 passes in evaluation (to compensate for the approximately 5 times longer training run of Ensembles.

4.3 Metrics

We evaluate based on the metric depth space d by computing the relative error. Denoting pixel number with $i \in \{1, ..., I\}$, model number by $m \in \{1, ..., M\}$, output from model m on pixel i $\tilde{d}_{i,m}$ and mean prediction $\hat{d}_i = \frac{1}{M} \sum_{m=1,...,M} \tilde{d}_{i,m}$, we compute the absolute relative error (REL) $= \frac{1}{I} \sum_{i=1,...,I} \frac{|d_i - \hat{d}_i|}{d_i}$. , the root mean squared error $(RMSE)$ $= \left(\frac{1}{M} \sum_{m=1,...,M} (d_i - \hat{d}_i)^2 \right)^{1/2}$, the scale-invariant error (SIL)[1] $= \frac{1}{M} \sum_{m=1...M} log(\hat{d}_i/d_i)^2 - \left(\frac{1}{M} \sum_{m=1...M} log(\hat{d}_i/d_i) \right)^2$, and the threshold accuracy $\delta_n :=$ % of pixels s.t. $\max(\hat{d}_i/d_i, d_i/\hat{d}_i) < 1.25^n$ for $n \in \{1, 2, 3\}$. Quantities without subscript i denote averages over pixels.

In evaluation we follow the literature and cap depth at 80m and use a region-of-interest-crop (Garg-crop) [4] as well as the KITTI-benchmark crop [7]. Furthermore, we report uncertainty in two versions, calculated as such; using the standard deviation of predictions across models, e.g. for MCDropout it would be $\sigma_i^2 = \frac{1}{M} \sum_{m=1...M} \tilde{d}_{i,m}^2 - \hat{d}_i^2$, we obtain pixel-wise uncertainty, defined as $\frac{\sigma_{i,modeltype}}{\mu_{i,modeltype}}$. This measure of uncertainty intuitively penalizes mistakes based on relative scale - an estimate being 1m off in depth estimation might in some respect be much worse when a target is 5m away than being 5 meters off a 80m target. As such, the metric is better suited for OOD-detection, compared to e.g. standard deviation, as the model can use clues in the image (e.g. vertical position and horizon lines) to output sensible predictions of depth in the OOD area, while still showing uncertainty in those predictions. Furthermore — and crucial for OOD-detection — it matters not for uncertainty whether distance is measured in meters or feet.

5 Results

Looking at figure 1, two things stand out - Posthoc and Online Laplace attribute uncertainty to the visually difficult and high loss-inducing pixels, such as edges or dark, murky areas. This is not

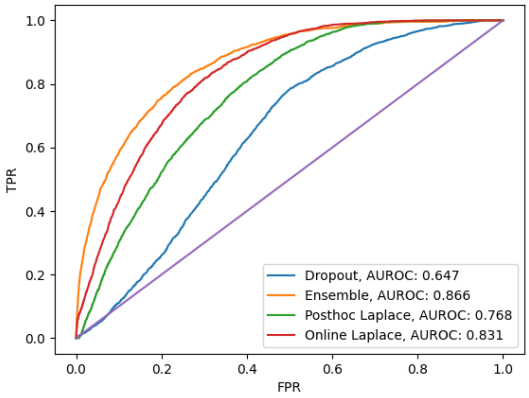

Figure 4: **Detection of OOD, ROC-curves**: Ensembles outperform all other models, Online Laplace. Across runs, this behaviour seemed to appear as a function of accuracy, rather than well-calibrated uncertainty.

Model type	$\delta_1 \uparrow$	$\delta_2 \uparrow$	$\delta_3 \uparrow$	$REL \downarrow$	$RMSE \downarrow$	$SIL \downarrow$	AUROC $\uparrow$
Dropout	0.917	0.982	0.995	0.091	2.606	13.855	0.647
Ensemble	0.955	0.986	0.994	0.064	2.378	12.358	0.866
Posthoc Laplace (Ours)	0.943	0.982	0.992	0.073	2.637	13.904	0.768
Online Laplace (Ours)	0.943	0.981	0.991	0.073	2.701	14.330	0.831

Table 1: **Results on testset.** Posthoc and Online Laplace is slightly outperformed by the 5x train-compute heavier Ensemble method.

seen in MCDropout, nor Ensembles, where uncertainty is much more haphazard. Furthermore, models all attribute higher uncertainty to out-of-distribution boxes, although the difference is not as pronounced for MCDropout. This is also accentuated in figure 3, and quantified in figure 4. In figure 4, OOD-data (pink boxes) are predicted using uncertainty - here both Online and Ensembles perform fairly, with Posthoc and MCDropout trailing. Turning to figure 2, we see that only Online Laplace has monotonically increasing loss in uncertainty. This is somewhat surprising given that it does not predict as strongly as say, Ensembles, see table 1. In general, we saw across runs, that models that predicted well, often tended to identify OOD better, and as such it is no surprise that Ensembles with the additional training compute, outperformed other models. What is striking, is, however, that these uncertainties are much more haphazard, as shown in 1, an observation that held across different images and runs. Note also, that this overperformance of Ensembles is not found elsewhere in literature [11].

6 Summary

We introduce both the Posthoc and Online Laplace depth estimators and show that while both slightly underperform the Ensemble method when concerned with predictive power, they provide meaningful uncertainties, capable of classifying OOD data on par with other methods. In addition, Online Laplace had a particularly attractive relationship between training compute and reliability of predictions, and captured the only positively monotone relationship between uncertainty and loss.

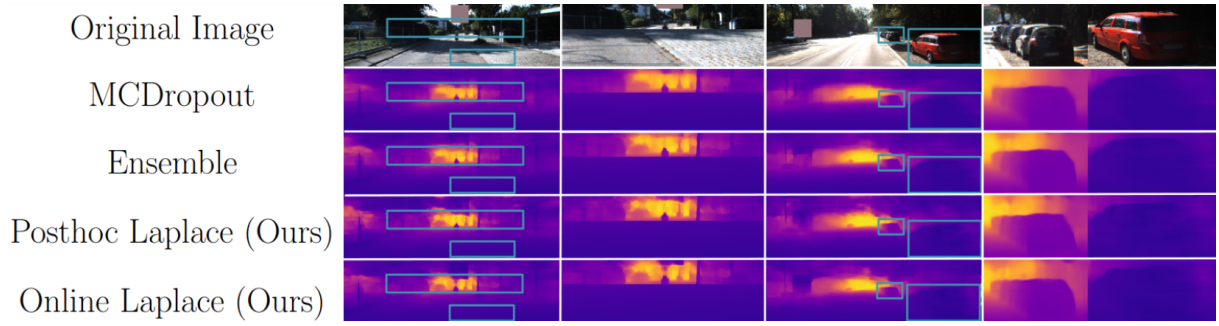

Table 2: **Qualitative assessment of predictions.** Two different images, were chosen at random. Brighter = further away. A pink box is placed randomly on the images, to test models' awareness of OOD-data. All models predict reasonably at mean, and all attribute OOD box to follow background distribution.

References

[1] David Eigen, Christian Puhrsch, and Rob Fergus. "Depth Map Prediction from a Single Image using a Multi-Scale Deep Network". In: *Advances in Neural Information Processing Systems*. Ed. by Z. Ghahramani et al. Vol. 27. Curran Associates, Inc., 2014. URL: https://proceedings.neurips.cc/paper_files/paper/2014/file/7bccfde7714a1ebadf06c5f4cea752c1-Paper.pdf.

[2] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. 2015. arXiv: 1505.04597 [cs.CV].

[3] Yarin Gal and Zoubin Ghahramani. "Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning". In: *Proceedings of The 33rd International Conference on Machine Learning*. Ed. by Maria Florina Balcan and Kilian Q. Weinberger. Vol. 48. Proceedings of Machine Learning Research. New York, New York, USA: PMLR, 20–22 Jun 2016, pp. 1050–1059. URL: https://proceedings.mlr.press/v48/gal16.html.

[4] Ravi Garg et al. *Unsupervised CNN for Single View Depth Estimation: Geometry to the Rescue*. 2016. arXiv: 1603.04992 [cs.CV].

[5] F. Dan Foresee and Martin T. Hagan. *GAUSS-NEWTON APPROXIMATION TO BAYESIAN LEARNING*. 2017.

[6] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. "Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles". In: *Advances in Neural Information Processing Systems*. Ed. by I. Guyon et al. Vol. 30. Curran Associates, Inc., 2017. URL: https://proceedings.neurips.cc/paper_files/paper/2017/file/9ef2ed4b7fd2c810847ffa5fa85bce
Paper.pdf.

[7] Jonas Uhrig et al. "Sparsity Invariant CNNs". In: *International Conference on 3D Vision (3DV)*. 2017.

[8] Erik Daxberger et al. *Laplace Redux – Effortless Bayesian Deep Learning*. 2022. arXiv: 2106.14806 [cs.LG].

[9] Marco Miani and Frederik Warburg. *NNJ*. 2022. URL: https://ilmiofrizzantinoamabile.github.io/nnj/nnj/index.html.

[10] Marco Miani et al. "Laplacian Autoencoders for Learning Stochastic Representations". In: *Advances in Neural Information Processing Systems*. Ed. by Alice H. Oh et al. 2022. URL: https://openreview.net/forum?id=aaar9y7qjfw.

[11] Pola Schwobel et al. "Probabilistic Spatial Transformer Networks". In: *The 38th Conference on Uncertainty in Artificial Intelligence*. 2022. URL: https://openreview.net/forum?id=HFUxb_Uiqec.

[12] Hong Liu et al. *Sophia: A Scalable Stochastic Second-order Optimizer for Language Model Pre-training*. 2023. arXiv: 2305.14342 [cs.LG].

[13] Frederik Rahbæk Warburg et al. "Bayesian Metric Learning for Uncertainty Quantification in Image Retrieval". In: *Thirty-seventh Conference on Neural Information Processing Systems*. 2023. URL: https://openreview.net/forum?id=58XMiu8kot.