
Spatial Uncertainty Sampling for End to End Control

Alexander Amini
CSAIL, MIT
Cambridge, MA 02139
amini@mit.edu

Ava Soleimany
Biophysics, Harvard
Cambridge, MA 02139
soleimany@g.harvard.edu

Sertac Karaman
LIDS, MIT
Cambridge, MA 02139
sertac@mit.edu

Daniela Rus
CSAIL, MIT
Cambridge, MA 02139
rus@csail.mit.edu

Abstract

End-to-end trained neural networks (NNs) are a compelling approach to autonomous vehicle control because of their ability to learn complex tasks without manual engineering of rule-based decisions. However, challenging road conditions, ambiguous navigation situations, and safety considerations require reliable uncertainty estimation for the eventual adoption of full-scale autonomous vehicles. Bayesian deep learning approaches provide a way to estimate uncertainty by approximating the posterior distribution of weights given a set of training data. Dropout training in deep NNs approximates Bayesian inference in a deep Gaussian process and can thus be used to estimate model uncertainty. In this paper, we propose a Bayesian NN for end-to-end control that estimates uncertainty by exploiting feature map correlation during training. This approach achieves improved model fits, as well as tighter uncertainty estimates, than traditional element-wise dropout. We evaluate our algorithms on a challenging dataset collected over many different road types, times of day, and weather conditions, and demonstrate how uncertainties can be used in conjunction with a human controller in a parallel autonomous setting.

Introduction

End-to-end deep learning methods have yielded promising results in autonomous lane following [1] and full end-to-end driving [2,3]. Steering control outputs (\mathbf{Y}) can be directly learned from raw sensor data (\mathbf{X}) by finding a mapping \mathbf{f} parameterized by weights \mathbf{W} which minimizes some objective error. While powerful, these deep NNs are trained end-to-end as black boxes, and as such, lack a definitive measure of associated confidence with the output. Integration of these models into deployable or shared control frameworks necessitates a method for determining predictive uncertainty.

Bayesian neural networks aim to learn a posterior distribution over the weights, $P(\mathbf{W}|\mathbf{X}, \mathbf{Y})$, and thus provide a grounded mathematical framework for determining model uncertainty. In practice, it is computationally infeasible to directly learn this posterior, $P(\mathbf{W}|\mathbf{X})$ from vast, and often noisy, observational data. Thus, variational inference (VI) methods have been used to obtain an approximation for the posterior to estimate model certainty. However, VI methods can also suffer from prohibitive computational cost [4–7]. Dropout sampling has emerged as an accurate, computationally efficient way to estimate model uncertainty. Applying dropout before every weight layer in a NN approximates a deep Gaussian process, since a dropout NN model minimizes the Kullback-Liebler divergence between the approximated posterior and the true posterior, $P(\mathbf{W}|\mathbf{X}, \mathbf{Y})$ [8,9]. Thus, dropout sampling provides an estimation of the predictive variance. Dropout, initially developed to avoid overfitting

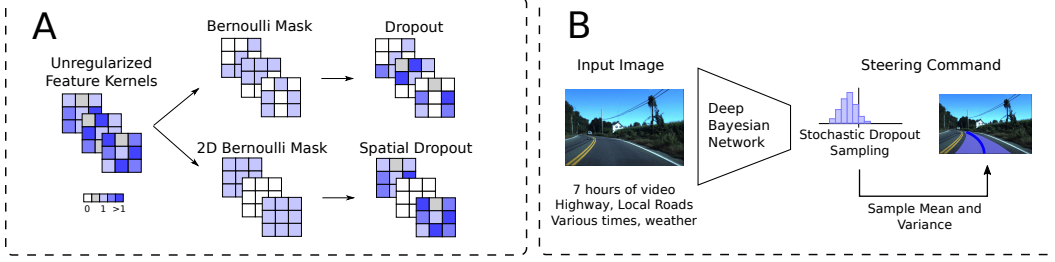


Figure 1: **A. Dropout Comparison.** Element wise dropout (top) compared to 2D spatial dropout with Bernoulli random variables. **B. End-to-end uncertainty estimation for autonomous vehicle control.** Pipeline for training a deep Bayesian network to output stochastic samples of control, which in turn can be used to predict a single steering command and associated uncertainty.

during training, places independent identically distributed Bernoulli random variables over every neuron to either “drop” or “keep” with some probability [10]. This approach is problematic when individual learned feature maps exhibit strong correlation, such as in the preliminary network layers for computer vision applications. Spatial dropout overcomes this limitation and achieves uncertainty estimates while also promoting independence between convolutional feature maps [11, 12].

Here, we address the need for reliable uncertainty estimation in end to end learning systems for autonomous vehicles. The contributions of this paper can be summarized as follows: (1) Demonstration of accurate uncertainty estimation utilizing 2D spatial dropout; (2) A deep Bayesian network for end-to-end autonomous vehicle control; and (3) A shared control framework which incorporates uncertainty estimates, thus enabling shared human-machine interaction.

Related Work

The idea of using Bayesian integration over neural network parameters [13, 14] and of averaging over the prior Gaussian weight distributions to estimate uncertainty of the output [15–17] have been widely explored. Recently, due to increase in training data sizes, more computationally tractable approaches have emerged. Specifically, it was shown that dropout [10] can be used to estimate uncertainty by sampling the approximate underlying Gaussian process [8, 18]. The effect of different dropout distributions (Bernoulli, Gaussian, and Spike-Slab) [19], convolution specific dropout techniques [11], as well as the impact of model versus data uncertainty [20] has also been explored.

This notion of stochastically sampling from Bayesian NNs to achieve uncertainty estimates has been applied to disease detection [21], reinforcement policy learning [8], as well as scene segmentation [22] and depth estimation [20] for autonomous vehicles. Bayesian uncertainty has not previously been applied to the more challenging autonomous actuation aspect of end-to-end control [2, 3], and it has not yet been shown how uncertainty can be incorporated into a parallel autonomy system [23] where human and robot (i.e., vehicle steering system) share control. Our approach addresses these limitations by using spatial dropout [11] to estimate uncertainty while also promoting independence between feature maps during training of our end-to-end driving network.

Methodology

We implement two dropout based approaches for estimating model uncertainty in a Bayesian NN for end-to-end autonomous vehicle control. It can be shown that applying dropout before every weight layer is equivalent to approximating a probabilistic deep Gaussian process [8]. From the Dropout-based approximation of the posterior $q(\mathbf{W})$, we obtain a predictive distribution $q(\mathbf{Y}|\mathbf{X}) = \int P(\mathbf{Y}|\mathbf{X}, \mathbf{W})q(\mathbf{W})d\mathbf{W}$. Given T stochastic forward passes through the network using dropout, $\{\mathbf{W}_t\}_{t=1}^T$, we define the predictive mean as: $\mathbb{E}[\mathbf{Y}|\mathbf{X}] = \frac{1}{T} \sum_{t=0}^T \mathbf{f}(\mathbf{X}, \mathbf{W}_t)$ and define the model uncertainty (e.g., predictive variance) as $\text{Var}[\mathbf{Y}|\mathbf{X}] = \frac{1}{T} \sum_{t=0}^T \mathbf{f}(\mathbf{X}, \mathbf{W}_t)^2 - \mathbb{E}[\mathbf{Y}|\mathbf{X}]^2$.

In this paper, we use two methods for the aforementioned stochastic sampling: (1) element-wise Bernoulli dropout [10] and (2) spatial Bernoulli dropout [12], illustrated in Figure 1A. In the second approach, uncertainty estimates result from sampling a Bernoulli random variable \mathbf{Z} to drop entire feature maps: $\mathbf{z}^{(k,l)} \sim \text{Bernoulli}(p)$ corresponds to k -th feature map in the l -th layer, where p

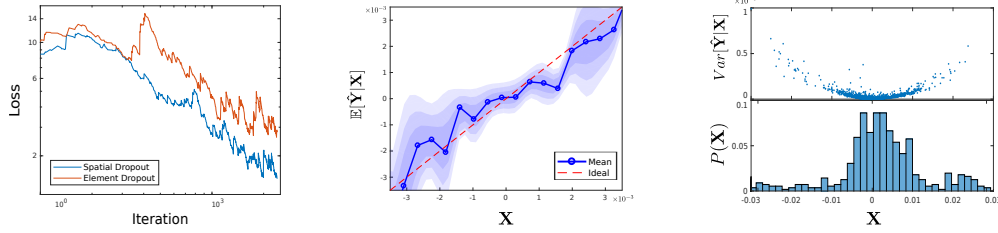


Figure 2: **Results.** Comparison of the loss optimization using spatial and element wise dropout (left). Evaluation of the predictive mean (center) and predictive variance (right top) against the ground truth control, \mathbf{X} . The underlying data distribution $P(\mathbf{X})$ is also illustrated on the bottom right. Note that, as expected, uncertainty (variance) increases as the amount of data decreases.

is the probability that all units in the feature map remain active [11]. Equivalently, this can be thought of as a special case of element-wise dropout, further supporting the notion that the sample variance from T stochastic passes of spatial dropout provides a valid way to estimate the uncertainty. Specifically, we consider each neuron in a feature map to be drawn from a Bernoulli distribution $\tilde{\mathbf{z}}^{(i,k,l)} \sim \text{Bernoulli}(p)$, corresponding to the i -th neuron in the k -th feature map from the l -th layer. Then, spatial dropout precisely corresponds to the instance where $\tilde{\mathbf{z}}^{(i,k,l)} = \tilde{\mathbf{z}}^{(j,k,l)} \forall (i,j)$ pairs across feature maps in convolutional layers. Because we can think of spatial dropout as a special instance of element wise dropout with an additional constraint, it follows that spatial dropout provides a valid way to estimate the model uncertainty. We implemented both approaches and compared the resulting loss optimization and uncertainty estimates.

We tested the efficacy of spatial dropout sampling of deep Bayesian networks applied to the complex task of end-to-end driving (Figure 1B). The vehicle base platform used for this study is a Toyota Prius 2015 V, which we retrofitted with sensors, power, and computing systems for parallel and autonomous driving. We collected an extensive dataset with over 7 hours and 500GB of driving data across a range of road and weather conditions to train our models. All models were trained on the NVIDIA DGX-1 supercomputer and were able to run in real-time on the car’s NVIDIA Drive PX2.

Our network takes as input a single image frame from a front facing video camera and performs feature extraction with 5 convolution layers with 5×5 kernels and alternating $(2 \times 2, 1 \times 1)$ strides. The resulting features are then fed into 4 fully connected layers which perform dimensionality reduction until we get a single output unit at the last layer. Dropout is performed after every layer with $p = \{0.9, 0.5\}$, where p corresponds to the probability of being kept, for convolutional and fully connected layers, respectively. The network was trained end-to-end to optimize MSE of the inverse-turning radius of the vehicle (monotonic in steering wheel command).

Experiments and Results

In this section, we outline the uncertainty experiments performed during this study and introduce a novel control strategy to incorporate uncertainty estimates in a shared autonomous controller. We observed that utilizing spatial dropout during training lead to faster convergence compared to element wise dropout (Figure 2 left). After training the model, we were evaluated the predictive mean of new data as $\mathbb{E}(\mathbf{Y}|\mathbf{X})$ (center). The shades of blue represent half standard deviations from the mean and we observe a greater variance as we are given data farther from the origin. Specifically, this corresponds to having more variance in the predictive mean when given roads with tighter turns, as opposed to on relatively straight roads. Finally, we evaluate the predictive uncertainty, $\text{Var}(\mathbf{Y}|\mathbf{X})$ (right top), again as a function of the data. Note, that we observe a similar trend where uncertainty rises as the data moves farther from the origin (increased turns). As expected, we also observe that less data is available from these situations as well (right bottom) supporting a general trend where model uncertainty increases as the amount of data used during training decreases.

We also found that the mean uncertainty error, $\left(\frac{y - \mathbb{E}(\mathbf{Y}|\mathbf{X})}{\text{Var}(\mathbf{Y}|\mathbf{X})}\right)$, which penalizes estimators with high error and low uncertainty, for spatial dropout (1.74) was less than that obtained from element-wise dropout (2.67). Furthermore, we use these uncertainty estimates to develop a shared controller which fuses the human’s actions with the network’s output according to the uncertainty. In other words, we formulate a parallel autonomy control strategy $u_{PA} = (1 - \sigma) \cdot u_N + \sigma \cdot u_H$ where u_{PA}, u_H, u_N

correspond to parallel autonomy, human, and network control commands for $\sigma = \kappa \text{Var}(\mathbf{Y}|\mathbf{X})$, with κ being a tunable threshold parameter that depends on the drop probability.

Conclusion

In this paper, we present a novel way to utilize spatial dropout to stochastically sample and compute uncertainty estimates of an end-to-end autonomous vehicle control system. We demonstrate how our estimates allow us to understand the limitations of our model when performing inference and thus provide a tunable framework which fuses model and human control. We are currently working to extend this work to not only estimate the posterior over a single control output but instead over a full parametric Gaussian Mixture Model (GMM) across all possible control outputs. In doing so, we will estimate the uncertainty associated to each of the individual GMM parameters and gain an even richer understanding of our model’s confidence.

Acknowledgments

Support for this work was given by the Toyota Research Institute (TRI). However, note that this article solely reflects the opinions and conclusions of its authors and not TRI or any other Toyota entity. We gratefully acknowledge the support of NVIDIA Corporation with the donation of the DGX-1 used for this research.

References

- [1] A. Gurgurian, T. Koduri, S. V. Bailur, K. J. Carey, and V. N. Murali, “Deeplanes: End-to-end lane position estimation using deep neural networks,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 38–45, 2016.
- [2] M. Bojarski, D. Del Testa, D. Dworakowski, B. Firner, B. Flepp, P. Goyal, L. D. Jackel, M. Monfort, U. Muller, J. Zhang, *et al.*, “End to end learning for self-driving cars,” *arXiv preprint arXiv:1604.07316*, 2016.
- [3] H. Xu, Y. Gao, F. Yu, and T. Darrell, “End-to-end learning of driving models from large-scale video datasets,” *arXiv preprint arXiv:1612.01079*, 2016.
- [4] J. Paisley, D. Blei, and M. Jordan, “Variational bayesian inference with stochastic search,” *arXiv preprint arXiv:1206.6430*, 2012.
- [5] D. P. Kingma and M. Welling, “Auto-encoding variational bayes,” *arXiv preprint arXiv:1312.6114*, 2013.
- [6] D. J. Rezende, S. Mohamed, and D. Wierstra, “Stochastic backpropagation and approximate inference in deep generative models,” *arXiv preprint arXiv:1401.4082*, 2014.
- [7] M. D. Hoffman, D. M. Blei, C. Wang, and J. Paisley, “Stochastic variational inference,” *The Journal of Machine Learning Research*, vol. 14, no. 1, pp. 1303–1347, 2013.
- [8] Y. Gal and Z. Ghahramani, “Dropout as a bayesian approximation: Representing model uncertainty in deep learning,” in *international conference on machine learning*, pp. 1050–1059, 2016.
- [9] A. Damianou and N. Lawrence, “Deep gaussian processes,” in *Artificial Intelligence and Statistics*, pp. 207–215, 2013.
- [10] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: a simple way to prevent neural networks from overfitting,” *Journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.
- [11] Y. Gal and Z. Ghahramani, “Bayesian convolutional neural networks with bernoulli approximate variational inference,” *arXiv preprint arXiv:1506.02158*, 2015.
- [12] J. Tompson, R. Goroshin, A. Jain, Y. LeCun, and C. Bregler, “Efficient object localization using convolutional networks,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 648–656, 2015.
- [13] J. Denker, D. Schwartz, B. Wittner, S. Solla, R. Howard, L. Jackel, and J. Hopfield, “Large automatic learning, rule extraction, and generalization,” *Complex systems*, vol. 1, no. 5, pp. 877–922, 1987.
- [14] N. Tishby, E. Levin, and S. A. Solla, “Consistent inference of probabilities in layered networks: Predictions and generalization,” in *International Joint Conference on Neural Networks*, vol. 2, pp. 403–409, 1989.
- [15] J. S. Denker and Y. Lecun, “Transforming neural-net output levels to probability distributions,” in *Advances in neural information processing systems*, pp. 853–859, 1991.

- [16] D. J. MacKay, “A practical bayesian framework for backpropagation networks,” *Neural computation*, vol. 4, no. 3, pp. 448–472, 1992.
- [17] R. M. Neal, *Bayesian learning for neural networks*, vol. 118. Springer Science & Business Media, 2012.
- [18] Y. Gal and Z. Ghahramani, “A theoretically grounded application of dropout in recurrent neural networks,” in *Advances in neural information processing systems*, pp. 1019–1027, 2016.
- [19] P. McClure and N. Kriegeskorte, “Representing inferential uncertainty in deep neural networks through sampling,” 2016.
- [20] A. Kendall and Y. Gal, “What uncertainties do we need in bayesian deep learning for computer vision?,” *arXiv preprint arXiv:1703.04977*, 2017.
- [21] C. Lebig, V. Allken, M. S. Ayhan, P. Berens, and S. Wahl, “Leveraging uncertainty information from deep neural networks for disease detection,” *bioRxiv*, p. 084210, 2017.
- [22] A. Kendall, V. Badrinarayanan, and R. Cipolla, “Bayesian segnet: Model uncertainty in deep convolutional encoder-decoder architectures for scene understanding,” *arXiv preprint arXiv:1511.02680*, 2015.
- [23] F. Naser, D. Dorhout, S. Proulx, S. D. Pendleton, H. Andersen, W. Schwarting, L. Paull, J. Alonso-Mora, M. H. Ang, S. Karaman, *et al.*, “A parallel autonomy research platform,” in *Intelligent Vehicles Symposium (IV)*, 2017 IEEE, pp. 933–940, IEEE, 2017.