Multi-Task Learning based on Separable Formulation of Depth Estimation and its Uncertainty

Akari Asai, Daiki Ikami, Kiyoharu Aizawa
The University of Tokyo
{asai, ikami, aizawa}@hal.t.u-tokyo.ac.jp

Abstract

We present an optimization framework for uncertainty estimation in a regression problem. To obtain predictive uncertainty inherent in the observation, we formulate regression with uncertainty estimation as a multi-task learning problem and a new uncertainty loss function, inspired by variational representations of robust estimation. Contrary to existing approaches, our approach allows balancing between the predictive task loss and uncertainty estimation loss. We evaluate the efficacy of our approach on NYU Depth Dataset V2 and show that our proposed method consistently yields better performance than the previous approaches, for both depth and uncertainty estimation.

1. Introduction

Despite the rapid progress made in deep visual learning, the difficulty in interpreting models limits their adoption in practical applications, especially in safety-critical domains (e.g., self-driving, health care). The aleatoric uncertainty is inherent in the observation and cannot be reduced with more data [4].

In this paper, we present a novel optimization framework for the aleatoric uncertainty in a regression problem, by formulating it as multi-task learning. Our method learns to predict the target output y and uncertainty σ by optimizing a separated loss: $L(y, \sigma) = L(y) + \lambda L(\sigma)$, as shown in Figure 1. On the contrary, previous approaches in aleatoric uncertainty estimation use a Bayesian deep learning framework [4, 5], in which the formulation of the loss function cannot be factored into a regression loss $L(y)$ and uncertainty estimation loss $L(\sigma)$ (we call this approach as "Joint Formulation"). This results in a drawback: the trade-off between regression performance and uncertainty estimation performance cannot be adjusted due to this non-separability. Previous studies in multi-task learning show that relative weighting between multiple task objectives is crucial [6].

This work is partially supported by JST CREST JPMJCR19F4.

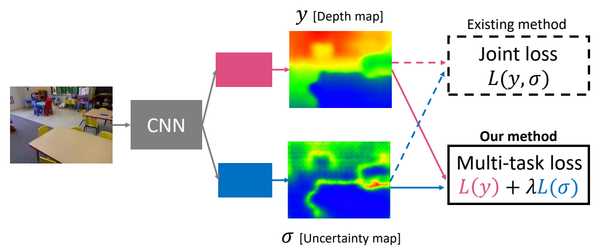

Figure 1: An overview for aleatoric uncertainty estimation in deep depth estimation. The y and σ are learned to predict depth values and its uncertainty, respectively. Whereas existing methods use a joint loss function $L(y, \sigma)$ that cannot be separated into $L(y)$ and $L(\sigma)$, we propose a separate multi-task loss $L(y) + \lambda L(\sigma)$.

Our method addresses the shortcoming of the existing approaches. First, our separable formulation of the loss function can adjust the performance of regression and uncertainty estimation by controlling λ. Second, this formulation also allows us to choose an arbitrary loss function for uncertainty. We introduce a new loss function for uncertainty estimation inspired by the variational representation of robust estimation.

To demonstrate the effectiveness of our proposed framework, we conduct experiments on RGB-based depth prediction dataset, namely NYU Depth Dataset V2. We show that our method substantially outperforms Joint Formulation, for both depth and uncertainty estimation performance, enjoying the benefits of controlling the balance between two losses and choosing an appropriate loss for uncertainty.

2. Proposed Method

In this work, we focus on a pixel-wise regression problem. Let $D = \{(\mathbf{x}_i, \mathbf{y}_i^*)\}_{i=1}^n$ be a training dataset of images and target values. Our goal is to train a network that predicts a target value y and its uncertainty u (or confidence

$w = 1/u$) for every pixel.

2.1. Uncertainty Estimation as Multi-task Learning

Here, we assume that we have a deep learning model which predicts both a target value y and its confidence value w. We formulate regression with uncertainty estimation as multi-task learning problem. Therefore, our loss function $L(\theta)$ has two terms, $\mathcal{L}^t(y)$ and $\mathcal{L}^u(w)$, which are loss functions of regression and uncertainty estimation as follows:

$$L(\theta) = \frac{1}{|D|}\sum_{i=1}^{n}\sum_{j=1}^{m}\mathcal{L}_j(\mathbf{x}_i, \theta), \tag{1}$$

$$\mathcal{L}_j(\mathbf{x}, \theta) = \mathcal{L}^t(y_j(\mathbf{x}; \theta)) + \lambda\mathcal{L}^u(w_j(\mathbf{x}; \theta)), \tag{2}$$

where λ is a weighting parameter between the two losses. For the choice of $\mathcal{L}^t(y)$, we can use any loss function of the regression problem, such as the L_2 loss and L_1 loss between a network output y and a ground truth y^*. Contrary to the existing approach, this separate formulation enables us to choose the uncertainty loss, as well.

2.2. A New Uncertainty Loss

Uncertainty is the expected estimation error and is larger for data with larger error. The loss of uncertainty $\mathcal{L}^u(w)$ should be constructed so that w predicts the inverse estimation error. For example, we can consider the L_2 distance: $\mathcal{L}^u(w) = \|1/|y^* - y| - w\|_2^2$. However, this loss functions is counter-intuitive – two uncertainty outputs of *the target output value with infinite uncertainty (w=0)* and *the target output value with half value of correct uncertainty (w=2/r)* have the same loss, which seems to under-estimate the loss of infinite uncertainty.

To overcome this problem, we introduce the loss function inspired by variational representations of robust estimation. The optimization problem of robust regression is formulated as

$$\min_{\theta}\sum_{i} f(r_i(\theta)), \tag{3}$$

where $r_i(\theta)$ is the ith residual, θ is a model parameter, and $f(r)$ is a robust loss function, which is sometimes nonconvex. To solve Eq. (3), a variational representation of robust loss [1, 12] is employed, which transforms $f(r)$ into

$$f(r) = \min_{w} wg(r) + h(w). \tag{4}$$

The optimal w can be obtained by the first-order derivative of Eq. (4), and it reflects data reliability: w is small when r is large. Similar to Eq. (4), we formulate $\mathcal{L}^u(w)$ as $\mathcal{L}^u(w) = wg(r) + h(w)$, where $g(r)$ is an arbitrary loss function. We use $g(r) = |y^* - y|$ in this paper.

In this work, we introduce a new uncertainty loss function, and we employ $h(w) = \frac{1}{w}$:

$$\mathcal{L}^u_{inv}(w) = wg(r) + \frac{1}{w}, \tag{5}$$

The derivative of Eq. (5) on w is

$$\frac{\partial L^u_{inv}}{\partial w_i} = g(r) - \frac{1}{w^2}. \tag{6}$$

The optimal value is $w^* = 1/\sqrt{g(r)}$. $L^u_{inv}(w)$ also has a large value at $w \to +0$. $\mathcal{L}^u_{inv}$ has multiplicative symmetry with respect to w_i^*: $\mathcal{L}^u_{inv}(\alpha w^*) = \mathcal{L}^u_{inv}(w^*/\alpha)$ holds. This symmetry is an intuitively proper characteristic for uncertainty estimation.

Our loss function Eq. (1) is differentiable, and we can train a network via gradient descent. The derivative of $\mathcal{L}_j$ defined by Eq. (2) on θ is computed by

$$\frac{\partial\mathcal{L}_j}{\partial\theta} = \frac{\partial\mathcal{L}^t_j}{\partial\theta} + \lambda\left(\frac{\partial w_j}{\partial\theta}g(r) + w_j\frac{\partial g(r)}{\partial r}\frac{\partial r}{\partial\theta} + \frac{\partial h(w_j)}{\partial w_j}\frac{\partial w_j}{\partial\theta}\right).$$

However, the term $\frac{\partial g(r)}{\partial r}$ hinders us to separate two objectives, and thus in this work, we detach the gradient of $g(r)$ and treat the depth regression residual of g(r) as constant (we do not detach the gradient of $\mathcal{L}^t_j$). The derivative of $\mathcal{L}^u(w)$ on θ is computed by

$$\frac{\partial\mathcal{L}^u_j}{\partial\theta} = \frac{\partial w_j}{\partial\theta}g(r) + \frac{\partial h(w_j)}{\partial w_j}\frac{\partial w_j}{\partial\theta}. \tag{7}$$

3. Experiments

We conduct experiments on a widely used RGB-based depth prediction dataset, NYU Depth Dataset V2 [10]. We split the data with the official split, where 249 scenes are used for training and the remaining 215 are used for testing, and use the network based on Sparse-to-Dense [8], following the experimental settings in [8].

As in previous studies in depth estimation [3, 7, 8], we evaluate the depth estimation performance with root mean squared error (RMSE), mean absolute relative error (REL), mean absolute error (MAE) and δ_i, which represents the percentage of the pixels where the relative error is within a threshold. To accurately evaluate the uncertainty estimation performance, we adopt sparsification plots, which have been commonly used in confidence measurement fields of optical flow estimation [2, 9, 11]. In sparsification plots, the pixels are removed by order of uncertainty, that is, pixels with higher uncertainty would be removed first. We use the following three metrics: area under the curve (AUC) of the sparsification plots, RMSE p_{30} (the RMSE averaged over the pixels after removing ones with 70% highest uncertainties), and spearman's rank correlation coefficient (CC) between the examined uncertainty values and the corresponding end-point errors.

4. Results

Table 1 shows a comprehensive comparison of our proposed method and the previous approach of uncertainty

Table 1: **Depth and uncertainty estimation results on NYU Depth Dataset V2.**

	Depth Estimation				Uncertainty Estimation		
loss	RMSE (↓)	REL (↓)	MAE (↓)	δ_1 (↑)	CC (↑)	AUC (↓)	RMSE p_{30} (↓)
L_1 (w/o uncertainty)	0.530	0.148	0.387	0.804	–	–	–
Joint Formulation	0.531	0.149	0.388	0.803	0.334	**0.180**	0.463
Ours ($\lambda = 0.5$)	**0.521**	**0.146**	**0.380**	**0.814**	**0.346**	0.185	**0.453**

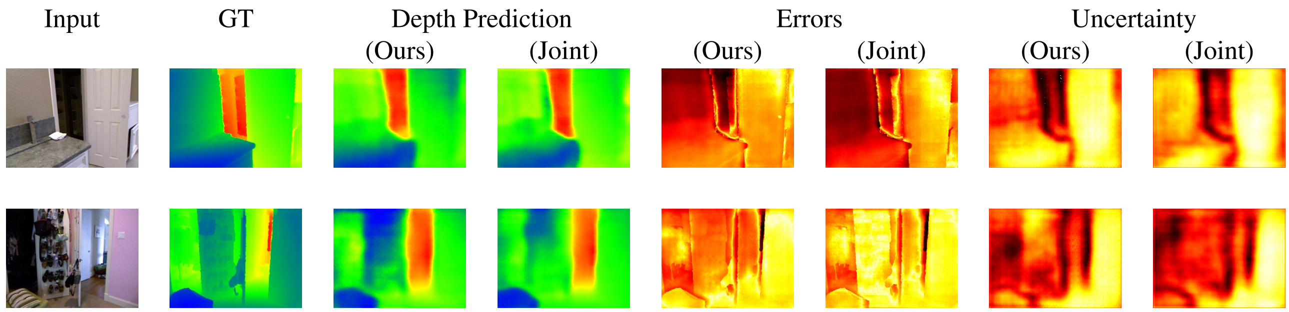

Figure 2: **Examples of the predicted results on NYU Depth Dataset V2.** From left to right: input RGB images (Input), the ground-truth (GT) depth maps for the inputs, predicted depth maps (Depth Prediction), the absolute errors (Errors), and expected uncertainty maps (Uncertainty) from our method and Joint Formulation, respectively. Regarding error maps and uncertainty maps, the red regions have larger values, while the yellow regions contain smaller values. On depth maps, the blue or green regions contain small depth values.

estimation. Our proposed method ("Ours", the bottom row) achieves better depth estimation performance from the model without uncertainty estimation (the top row) and the one with uncertainty estimation given by Joint Formulation (the middle row). Our method also achieves better uncertainty estimation performance than Joint Formulation on all metrics other than AUC. Figure 2 shows a qualitative comparison of the proposed method and the previous method. The two examples show that the previous approach is more likely to produce large uncertainties even for the regions with smaller errors.

We evaluate the performance of both depth estimation and uncertainty estimation using different loss-weighting parameters, λ. As shown in Figure 3, the depth estimation performance measured with RMSE deteriorates when the λ is set to a larger value, while with a smaller λ value, where a model learns to pay more attention for depth estimation, the CC decreases dramatically. Another finding is that, at some optimal weighting, the model performs better than any of other weightings, both for depth and uncertainty estimation.

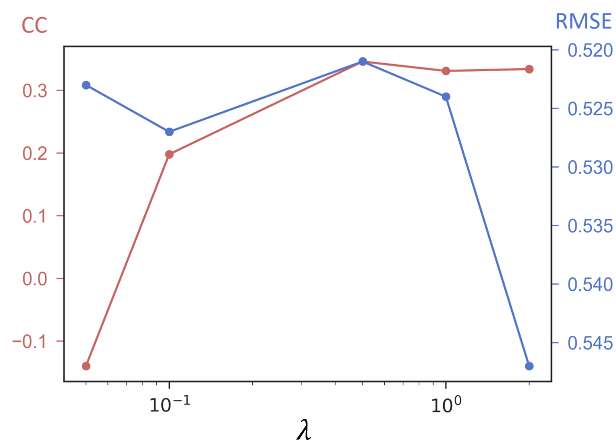

Figure 3: **Results with different weightings on NYU Depth Dataset V2.** RMSE is a measurement for depth estimation performance (the lower, the better), and CC is a measurement for uncertainty estimation performance (the higher, the better). For RMSE, we inverted the y-axis.

5. Conclusion and Future Work

In this paper, we present an optimization framework for aleatoric uncertainty estimation in a regression problem: separating the two loss functions, $\mathcal{L}^t$ and $\mathcal{L}^u$ and deriving a new objective with weight λ. Our experimental results on NYU Depth Dataset V2 showed that our proposed method was superior to previous studies, both in depth and uncertainty estimation. For future work, we intend to conduct experiments on other regression tasks like optical flow estimation as well as on another RGB depth estimation task.

References

[1] M. J. Black and A. Rangarajan. On the unification of line processes, outlier rejection, and robust statistics with applications in early vision. *IJCV*, 19(1):57–91, 1996. 2

[2] A. Bruhn and J. Weickert. A confidence measure for variational optic flow methods. In *Geometric Properties for Incomplete Data*. 2006. 2

[3] D. Eigen, C. Puhrsch, and R. Fergus. Depth map prediction from a single image using a multi-scale deep network. In *NIPS*, 2014. 2

[4] Y. Gal. Uncertainty in deep learning. *University of Cambridge*, 2016. 1

[5] A. Kendall and Y. Gal. What uncertainties do we need in bayesian deep learning for computer vision? In *NIPS*, 2017. 1

[6] A. Kendall, Y. Gal, and R. Cipolla. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. 2018. 1

[7] I. Laina, C. Rupprecht, V. Belagiannis, F. Tombari, and N. Navab. Deeper depth prediction with fully convolutional residual networks. In *3DV*, 2016. 2

[8] F. Ma and S. Karaman. Sparse-to-dense: Depth prediction from sparse depth samples and a single image. In *ICRA*, 2018. 2

[9] O. Mac Aodha, A. Humayun, M. Pollefeys, and G. J. Brostow. Learning a confidence measure for optical flow. *TPAMI*, 35(5):1107–1120, 2013. 2

[10] P. K. Nathan Silberman, Derek Hoiem and R. Fergus. Indoor segmentation and support inference from rgbd images. In *ECCV*, 2012. 2

[11] A. S. Wannenwetsch, M. Keuper, and S. Roth. Probflow: Joint optical flow and uncertainty estimation. In *ICCV*, 2017. 2

[12] Y.-l. Yu, Ö. Aslan, and D. Schuurmans. A polynomial-time form of robust regression. In *NIPS*, 2012. 2