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

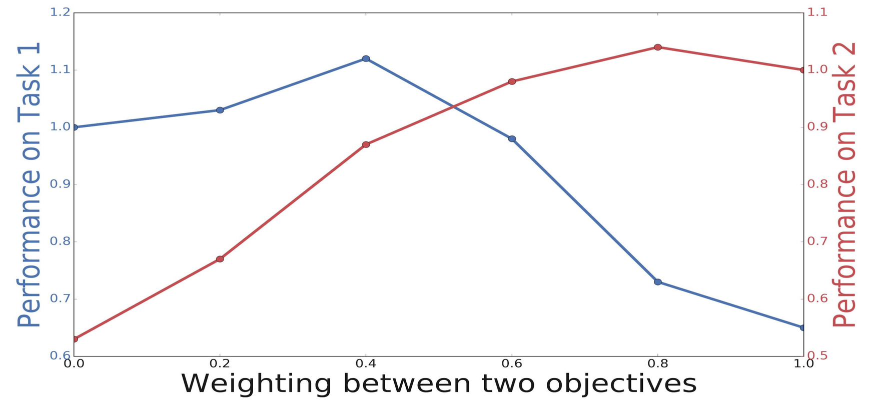
Akari Asai, Daiki Ikami, Kiyoharu Aizawa (The University of Tokyo)

CVPR LONG BEACH CALIFORNIA June 16-20, 2019

Summary

Background

- Quantification of uncertainty is crucial for safety-critical domains.
- Previous approach for aleatoric uncertainty cannot separate the loss due to the target task and the one from uncertainty estimation [1].
- Thus, we cannot adjust the weighting between two losses.



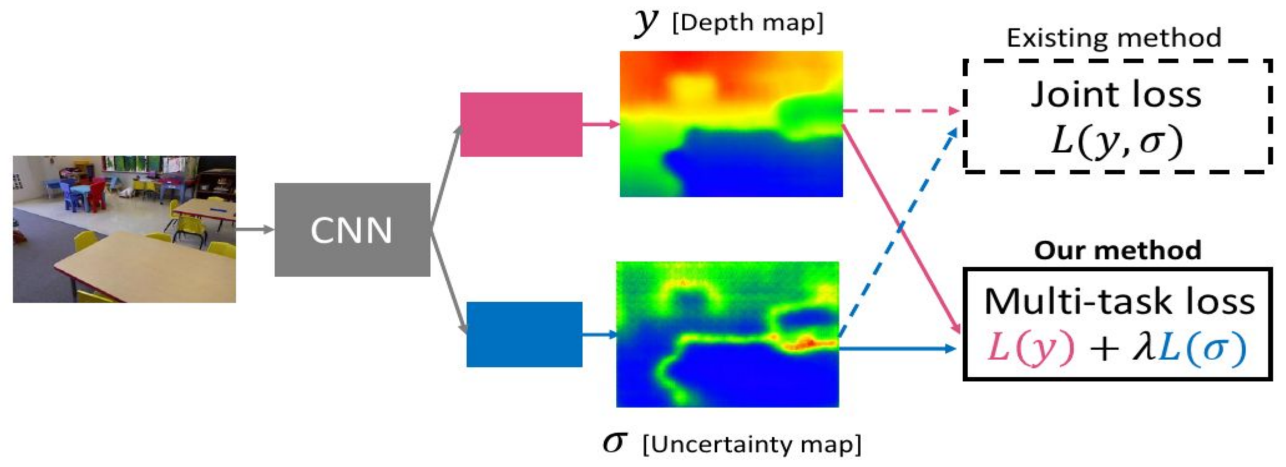
It is widely known that properly balancing multiple objectives is crucial in multi-task learning.

Proposed Method and results

- We propose a new optimization framework which makes (i) two losses separable, and (ii) lets one balance two losses by adjusting a weighting parameter.
- We observed performance improvement on both depth estimation and uncertainty estimation on NYU Depth Dataset V2.

Uncertainty Estimation as Multi-task Learning

Formulate the estimation of aleatoric uncertainty in regression as a multi-task learning problem.



Our separable formulation allows us to tune the weighting between two losses.

$$L(\theta) = \frac{1}{|D|}\sum_{i=1}^{n}\sum_{j=1}^{m}\mathcal{L}_j(\mathbf{x}_i, \theta), \quad \mathcal{L}_j(\mathbf{x}, \theta) = \mathcal{L}^t(y_j(\mathbf{x};\theta)) + \boxed{\lambda}\mathcal{L}^u(w_j(\mathbf{x};\theta))$$

A New Uncertainty Loss with Multiplicative Symmetry

Problem

- Uncertainty should be larger for data with larger errors.
- Existing loss such as L2 underestimates the loss of infinite uncertainty.

Inspired by variational representation of robust estimation

Variational representation of robust loss minimizes $f(r) = \min_{w} wg(r) + h(w)$.
w is smaller when r is large. $\boldsymbol{w}^*$ would be obtained by first-order derivative.

Formulate new uncertainty loss

Given an arbitrary loss function $g(r)$, we formulate uncertainty loss as follows:

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

The loss gets larger when $w \to 0$. it also has a multiplicative symmetry w.r.t. w^*.

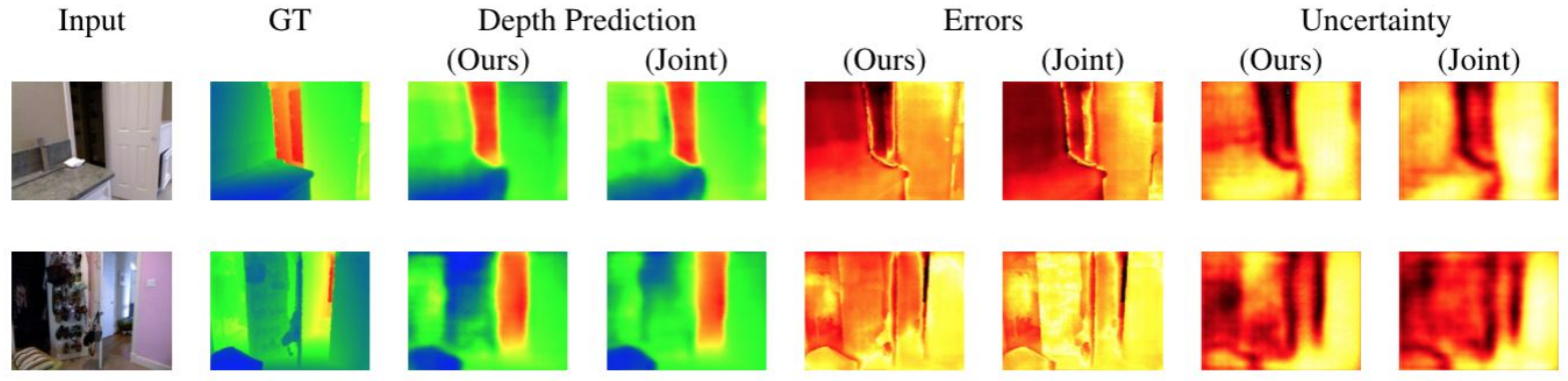
Training

To keep two objectives separate, we detach the gradient of $g(r)$.

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

Experimental Settings

Dataset: NYU Depth Dataset V2 / **Network:** Based on Sparse-to-Dense [2].
Depth Metrics: **RMSE, REL, MAE, δ_1** / **Uncertainty Metrics:** spearman's correlation coefficient (**CC**) btw errors and uncertainty values, area under the curve (**AUC**) of sparficification plot, **RMSE $_{p30}$**



Experiment

Depth and uncertainty estimation performance

- Significantly improves depth estimation performance from [1][3].
- Yields better uncertainty estimation performance than [1]

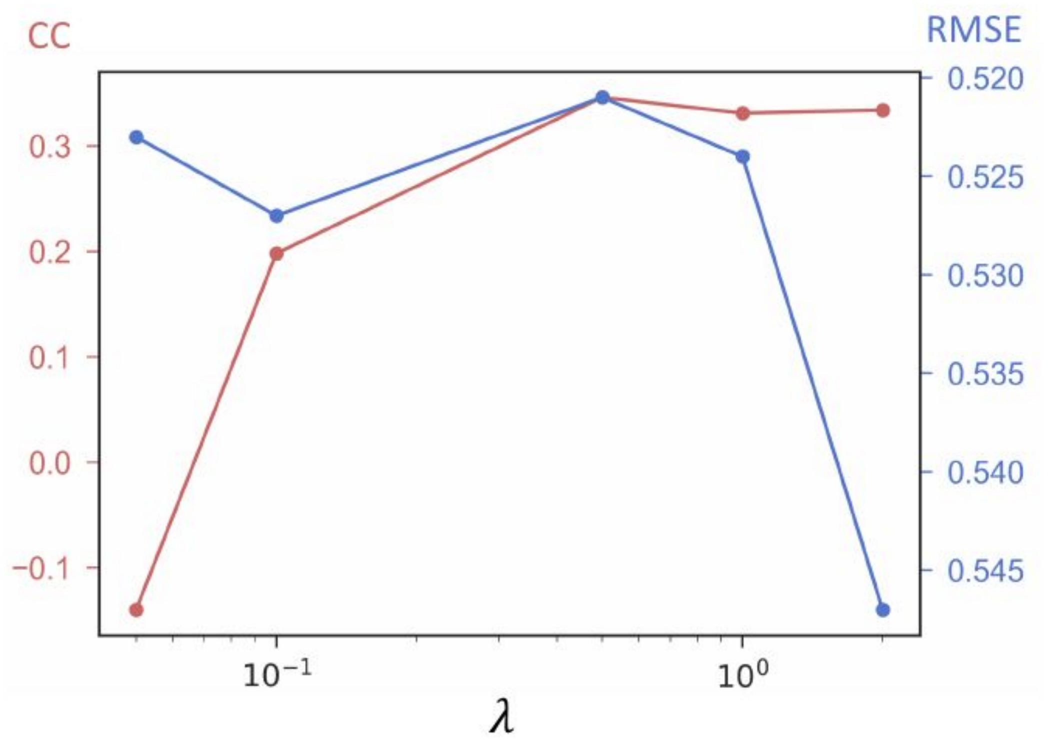
Depth Estimation

Loss	RMSE	REL	MAE	δ
L1 [3]	0.530	0.148	0.387	0.804
Joint [1]	0.531	0.149	0.388	0.803
Ours	**0.521**	**0.146**	**0.380**	**0.814**

Uncertainty Estimation

	CC	AUC	RMS p30
Joint [1]	0.334	**0.180**	0.463
Ours	**0.346**	0.185	**0.453**

Trade-off of uncertainty and depth estimation performance



- When λ is a large value:
  - **depth** ↓
  - uncertainty →
- When λ is a small value:
  - **depth** ↑
  - **uncertainty** ↓
- At a optimal value (0.5), both reach the best.

References

[1] Alex Kendall and Yarin Gal. 2017. What Uncertainties Do We Need In Bayesian Deep Learning for Computer Vision. In NeurIPS.

[2] Michael J. Black and Anand Rangaraja. 1996.On the unification of line processes, outlier rejection, and robust statistics with applications in early vision. IJCV

[3] Fangchang Ma and Sertac Karaman. 2018. Sparse-to-Dense: Depth Prediction from Sparse Depth Samples and a Single Image. In ICRA.

email: asai@hal.t.u-tokyo.ac.jp, akari@@cs.washington.edu
twitter: @AkariAsai