Reproduction of ‘Multi-Task Learning using Uncertainty to Weigh Losses’

D. Baerg, O. Key, J. Materzynska, M. Tong

Department of Computer Science, University of Oxford

A new approach to multi-task learning

Paper on ‘Multi-Task Learning using Uncertainty to Weigh Losses’ – Kendall, Gal & Cipolla, 2017.

- Principled approach to multi-task learning, demonstrated on the multi-task problem of semantic segmentation, instance segmentation and depth regression for the Cityscapes dataset

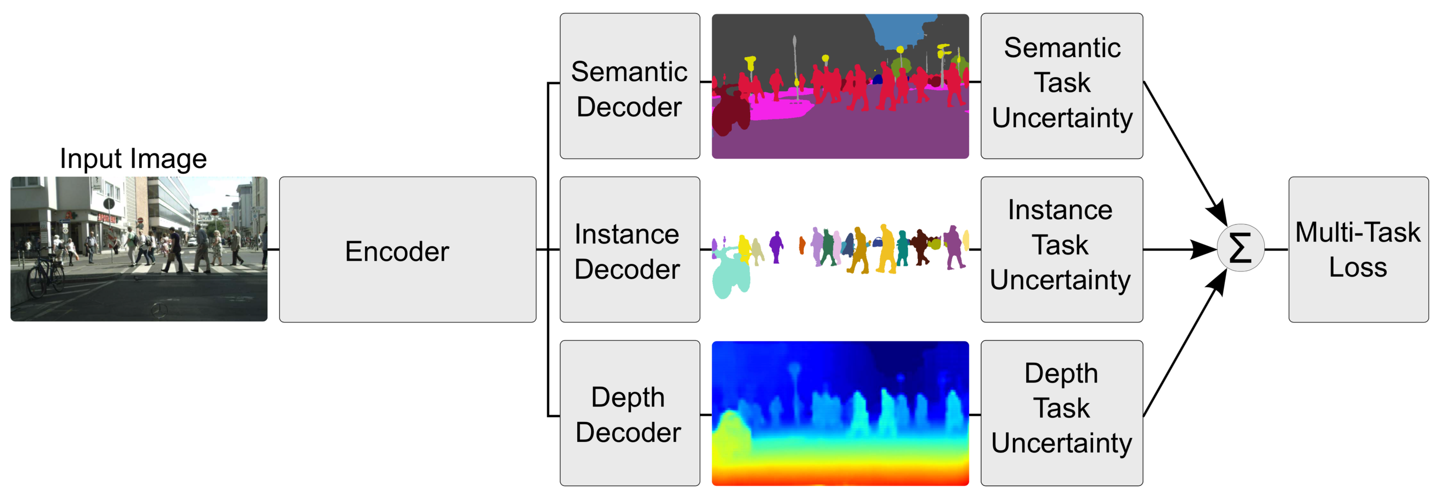

Fig. 1: Model architecture (Kendall et al., 2017)

- 3 key contributions:

1. Aleatoric homoscedastic uncertainty to weigh losses in different tasks
2. Unified architecture for all three tasks with the same encoder for different tasks, with a separate decoder for each task
3. Demonstrating that loss weighting is important for performance, and showing how superior performance can be achieved on multi-task models compared to individually trained single-task models

Selected reproduction objective

- Comparison of multi-task learning with single-task learning and learned loss weights with fixed loss weights
- Subsampled dataset, Tiny Cityscapes, requires less computation time

Loss	Task Weights Seg.	Inst.	Depth	Segmentation IoU [%]	Instance Mean Error [px]	Inverse Depth Mean Error [px]
Segmentation only	1	0	0	59.4%	-	-
Instance only	0	1	0	-	4.61	-
Depth only	0	0	1	-	-	0.640
Unweighted sum of losses	0.333	0.333	0.333	50.1%	3.79	0.592
Approx. optimal weights	0.89	0.01	0.1	62.8%	3.61	0.549
2 task uncertainty weighting	✓	✓		61.0%	**3.42**	-
2 task uncertainty weighting	✓		✓	62.7%	-	0.533
2 task uncertainty weighting		✓	✓	-	3.54	0.539
3 task uncertainty weighting	✓	✓	✓	**63.4%**	3.50	**0.522**

Fig. 2: Results that we aimed to reproduce (Kendall et al., 2017)

Technical details

- Overall joint loss function:

$$\mathcal{L} \approx \frac{1}{\sigma^2_{\text{sem}}}\mathcal{L}_{\text{sem}} + \log \sigma_{\text{sem}} + \frac{1}{2\sigma^2_{\text{ins}}}\mathcal{L}_{\text{ins}} + \log \sigma_{\text{ins}} + \frac{1}{2\sigma^2_{\text{dep}}}\mathcal{L}_{\text{dep}} + \log \sigma_{\text{dep}}$$

- Model architecture
  - DeepLabv3 encoder architecture, including ResNet-101 layers with dilated convolutions and Atrous Spatial Pooling Pyramid module to increase contextual awareness
  - Simple decoder architecture with two convolutional layers for each task
- Sacred with MongoDB to save results

Reproduction results

- Table 1 with paper hyperparams
- Semseg lr search
- Semseg weight decay search
- ASPP dilation size search
- Single batch (+ IoU calibration)

Success on other multi-task problems

- Applying the principled approach to other multi-task learning problems

MNIST and Fashion MNIST

- Table 1 style
- Grid search for optimal weights on two task
- Uncertainty convergence from different initial values