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

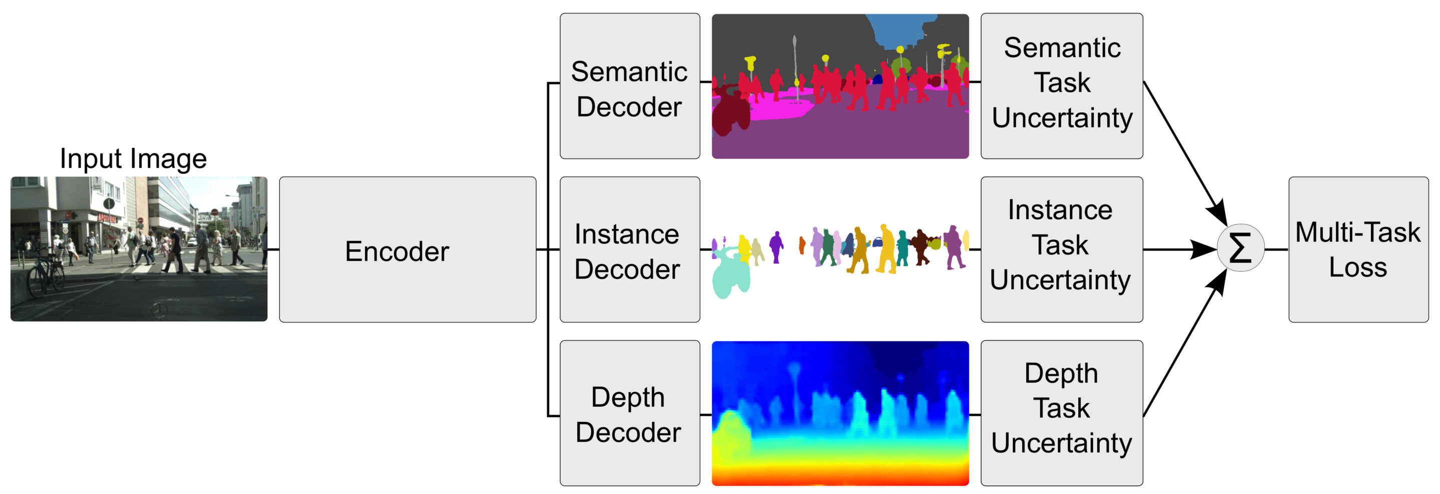


Fig. 1: Model architecture (Kendall et al., 2017)

- 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

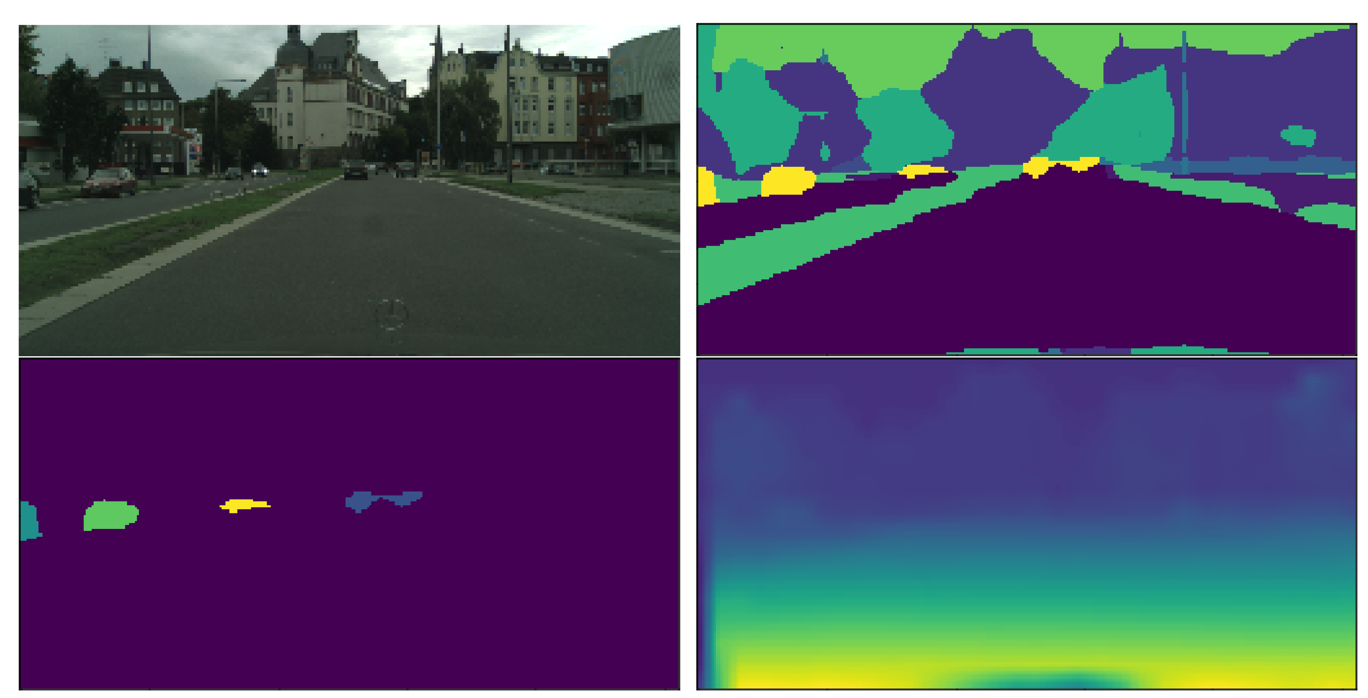
- Overall joint loss function:

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

- Model consisted of DeepLabv3 encoder architecture, including ResNet-101 layers with dilated convolutions and Atrous Spatial Pooling Pyramid module to increase contextual awareness with a simple decoder architecture with two convolutional layers for each task

Reproduction results

Task Weights Seg.	Inst.	Depth	Segment. IoU [%]	Instance Mean Error [px]	Inverse Depth Mean Error [px]
1	0	0	33.2%	-	-
0	1	0	-	**5.3**	-
0	0	1	-	-	0.45
0.333	0.333	0.333	32.3%	5.7	0.46
0.89	0.01	0.1	32.7 %	8.4	0.53
Learned	Learned	-	33.9%	7.2	-
Learned	-	Learned	33.2%	-	0.43
-	Learned	Learned	-	6.3	0.43
Learned	Learned	Learned	**34.0%**	5.8	**0.42**



- Learning rate search

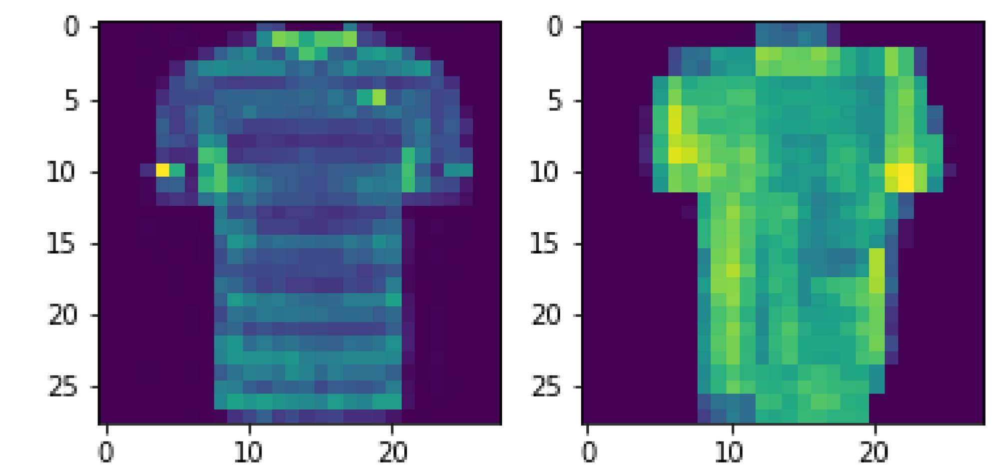
Learning rate	Segmentation IoU
0.01	**33.3%**
0.001	33.0%
0.0001	31.2%
0.00001	25.2%

Success on other multi-task problems

- Applying the principled approach to Fashion MNIST classification and reconstruction yields a learned weights multi-task model which outperforms individually trained models and fixed weights multi-task model

Task weights Classif.	Reconstr.	Classification accuracy [%]	Reconstruction error
1	0	89.7%	-
0	1	-	0.087
0.5	0.5	89.3%	0.085
Learned	Learned	**90.2%**	**0.075**

- Reconstruction output sample from learned weights multi-task model



- Optimal weight search