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

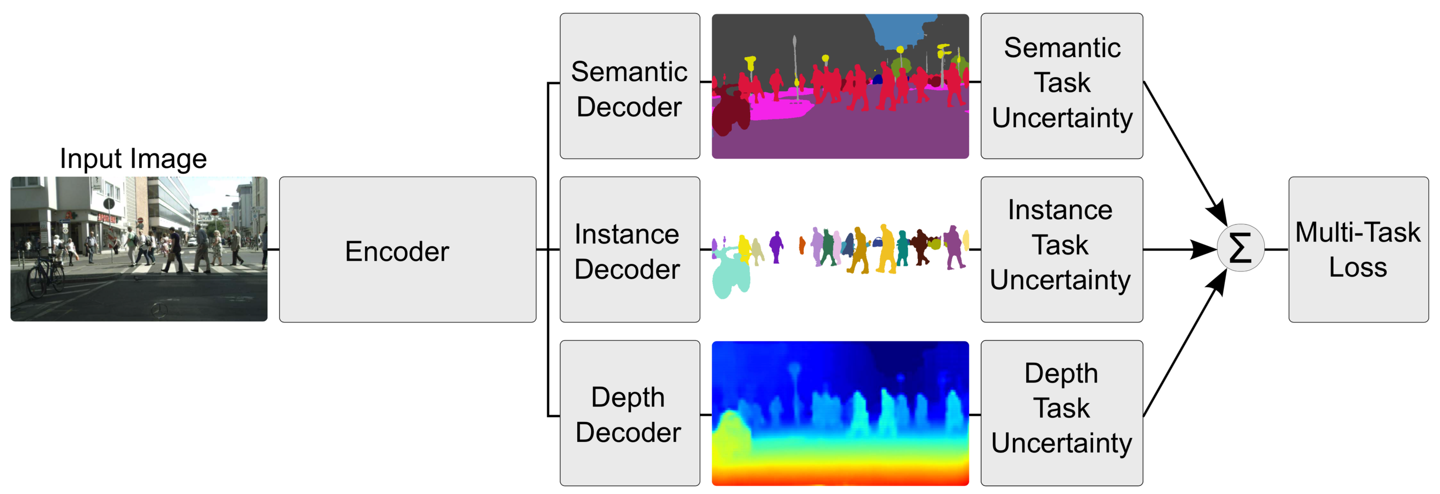


Fig. 1: Model architecture (Kendall et al., 2017)

- 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
- 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 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

Selected reproduction objective

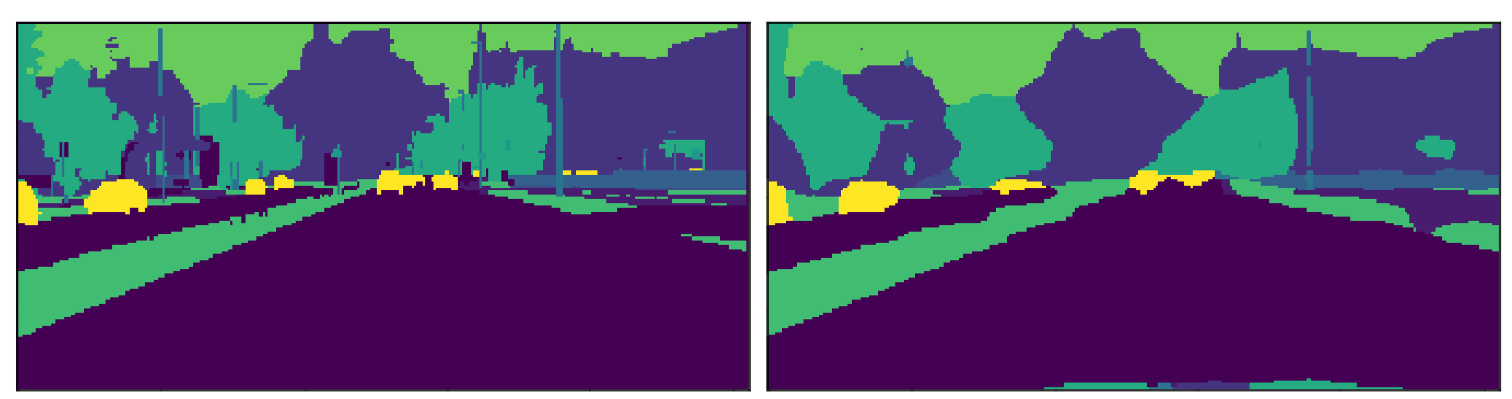
- Comparison of multi-task learning with single-task learning and learned loss weights with fixed loss weights
- Subsampled dataset, Tiny Cityscapes, requires less computation time

Reproduction results

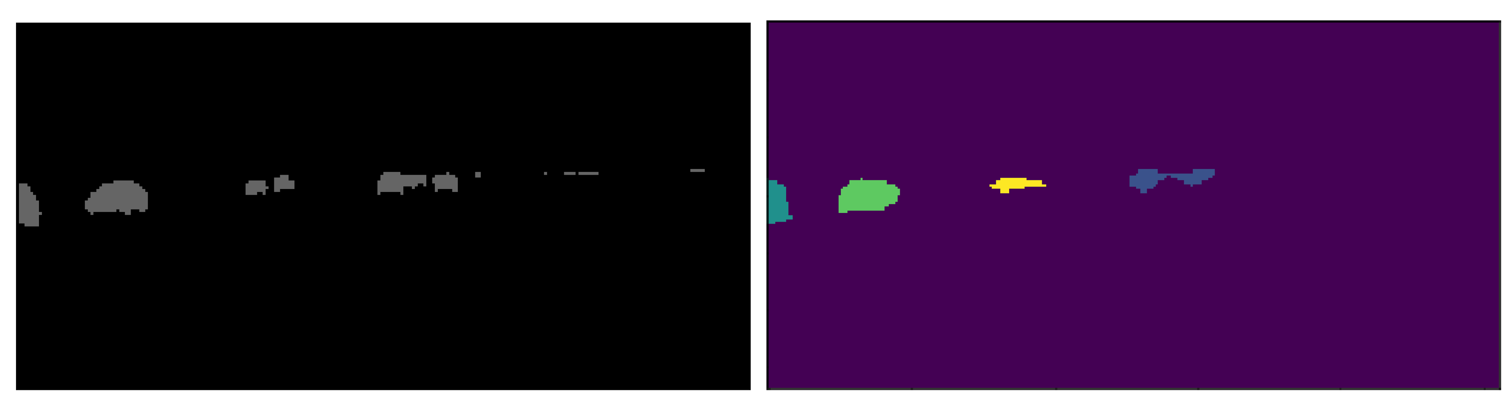
- Quantitative improvement for semantic segmentation and depth regression for multi-task model with learned weights
- Improved baseline for inverse mean depth error from 0.522 to 0.425

Task Weights			Segment.	Instance Mean	Inverse Depth
Seg.	Inst.	Depth	IoU [%]	Error $[px]$	Mean Error $[px]$
1	0	0	33.2%	-	-
0	1	0	-	**5.27**	-
0	0	1	-	-	0.446
0.333	0.333	0.333	32.3%	5.72	0.459
0.89	0.01	0.1	32.7 %	8.39	0.528
Learned	Learned	-	33.9%	7.17	-
Learned	-	Learned	33.2%	-	0.426
-	Learned	Learned	-	6.26	0.430
Learned	Learned	Learned	**34.0%**	5.82	**0.425**

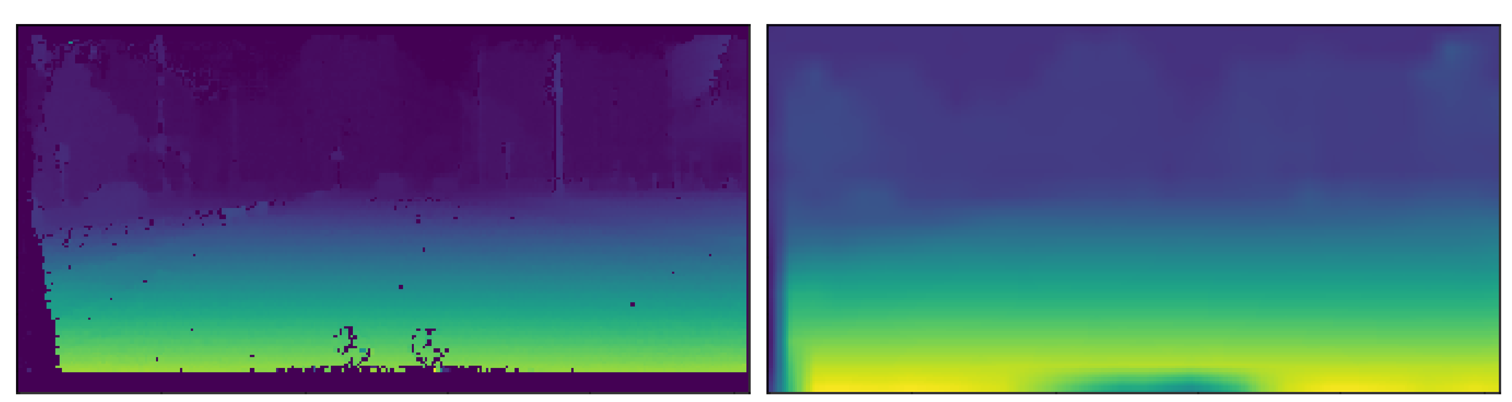
Ground truth Predicted



Semantic segmentation



Instance segmentation

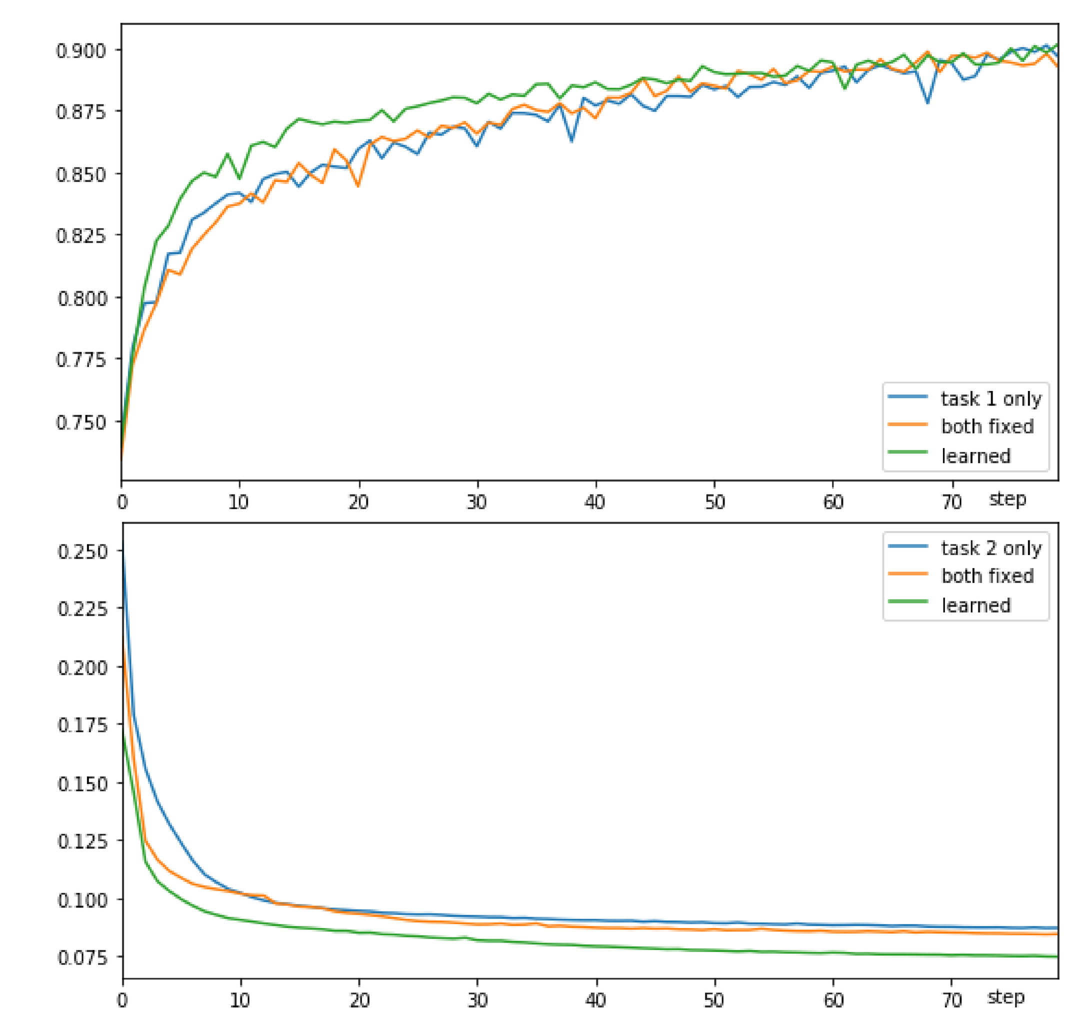


Depth regression

Success on other multi-task problems

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

Task weights		Classification	Reconstruction
Classif.	Reconstr.	accuracy [%]	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 samples from learned weights multi-task model