Online Knowledge Distillation for Multi-task Learning

Supplemental Material

Geethu Miriam Jacob Vishal Agarwal Björn Stenger

Rakuten Institute of Technology, Rakuten Group, Inc.
{geethu.jacob, vishal.agarwal, bjorn.stenger}@rakuten.com

1. Detailed Experiments

Table 1: Ablation study on AFD component. *This is a detailed table for table 5 from main document. We compare models with different configurations of AFD: AFD for last layer only (AFD-last layer), equal weighting (AFD-equal), softmax weighting (AFD-softmax), select/skip policy (AFD-select/skip) and the proposed random initialization (AFD-ours) for knowledge distillation on intermediate features*

	Sem. Segm.		Depth estimation		Surface Normal Prediction					
Method	mIoU↑	pAcc↑	abs↓	rel↓	mean↓	median↓	<11° ↑	<22.5° ↑	<30° ↑	Δ ↑
AFD-last layer	50.41	73.25	0.4164	0.1749	31.91	21.60	33.67	54.56	63.00	4.89
AFD-equal	51.11	73.15	0.4145	0.1740	31.87	21.55	33.58	54.65	63.12	5.28
AFD-softmax	51.18	73.4	0.4163	0.1614	31.78	21.40	33.68	54.51	62.97	5.68
AFD-select/skip	51.29	73.65	0.4133	0.1746	32.42	22.51	32.12	53.14	61.76	4.39
AFD-ours	51.99	73.75	0.4112	0.1701	31.82	21.50	33.58	54.74	63.20	**6.22**

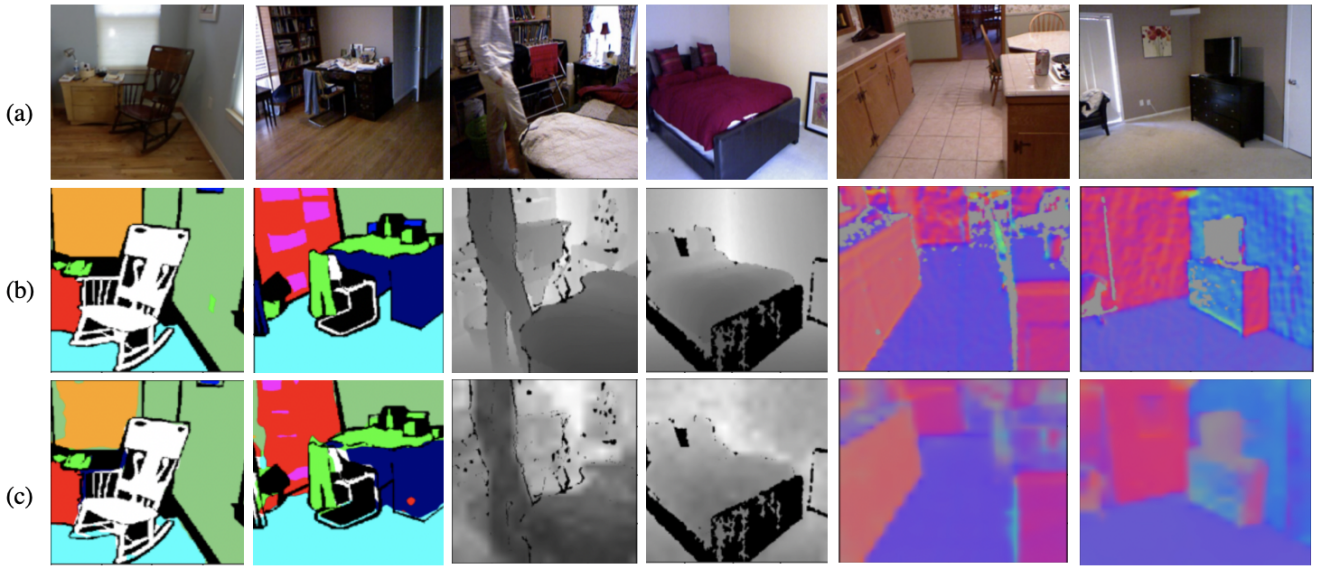

Figure 1: Qualitative results. *Inferences of semantic segmentation, depth estimation ad surface normal estimation on NYUv2 dataset (a) Input image, (b) groundtruth images, (c) predicted images.*

Table 2: Comparison with implemented SOTA on 3-task NYUv2 dataset with DeeplabV3 backbone. *Performance evaluation of state-of-the-art methods and the proposed method, implemented on DeeplabV3 architecture. The last column shows the average performance improvement. This is a detailed table of Table 3 (left) in main document.*

	Sem. Segm.		Depth estimation		Surface Normal Prediction					
Method	mIoU↑	pAcc↑	abs↓	rel↓	<11° ↑	<22.5° ↑	<30° ↑	mean↓	median↓	Δ ↑
Single Task	49.57	72.88	0.5052	0.1962	25.23	48.99	63.23	27.15	22.11	0.15
Baseline (MTL)	48.11	72.38	0.4792	0.1859	24.67	47.99	60.48	28.63	23.60	0
DWA [1]	48.21	72.29	0.4703	0.1817	24.76	48.14	60.69	28.54	23.54	0.81
GradNorm [2]	48.14	72.49	0.4816	0.1842	24.65	48.07	60.52	28.54	23.59	0.14
UW [3]	48.17	72.39	0.4773	0.1844	24.85	48.32	60.75	28.52	23.43	0.42
RLW [4]	48.39	72.43	0.4756	0.1871	24.76	48.17	60.55	28.67	23.57	0.18
Cross Stitch [5]	48.20	72.86	0.4789	0.1834	24.20	47.64	60.23	28.57	23.87	0.11
KD-MTL [6]	48.78	73.07	0.4605	0.1841	25.43	49.02	61.62	28.08	23.04	1.96
Ours	49.06	72.91	0.4880	0.1883	26.86	51.91	64.32	27.04	21.52	**2.45**

Table 3: Comparison with implemented SOTA on 3-task NYUv2 dataset with DeeplabV3-MTAN backbone. *Performance evaluation of state-of-the-art methods and the proposed method, implemented on MTAN-DeeplabV3 architecture. The last column shows the average performance improvement (full table of Table 3 (right) in main document).*

	Sem. Segm.		Depth estimation		Surface Normal Prediction					
Method	mIoU↑	pAcc↑	abs↓	rel↓	mean↓	median↓	<11° ↑	<22.5° ↑	<30° ↑	Δ ↑
Single Task	48.69	72.87	0.6228	0.2344	26.41	21.07	27.65	52.88	65.32	0.15
Baseline (MTL)	46.25	72.01	0.5314	0.2151	28.30	23.72	23.88	47.93	60.69	0
DWA [1]	46.58	72.23	0.5337	0.2079	27.79	23.10	24.47	48.98	61.91	1.39
GradNorm [2]	46.76	72.26	0.5304	0.2072	27.81	23.11	24.57	48.97	61.89	1.64
UW [3]	46.72	72.05	0.5351	0.2136	28.23	23.62	24.20	48.08	60.89	0.36
RLW [4]	46.24	71.64	0.5371	0.2050	28.03	23.57	23.85	47.37	60.21	0.48
KD-MTL [6]	47.35	72.50	0.5148	0.2031	27.66	22.94	25.18	49.30	62.14	3.04
OKD-MTL (ours)	48.30	72.58	0.4957	0.1971	27.36	22.33	25.88	50.50	63.15	**5.18**