Supplementary for SD-MTCNN: Self-Distilled Multi-Task CNN

Ankit Jha[1]
ankitjha16@gmail.com
Awanish Kumar[1]
awanishk389@gmail.com
Biplab Banerjee[1]
getbiplab@gmail.com
Vinay Namboodiri[2]
vinaypn@iitk.ac.in

[1] Indian Institute of Technology, Bombay
[2] Indian Institute of Technology, Kanpur

In the supplementary document, we report the followings,

- Analysis in terms of the evalution of different performance metrics.
- We showcase some of the qualitative results for the CityScapes dataset.
- Detailed baseline analysis for Mini-Taskonomy, NYUv2 for two tasks, and Cityscapes, respectively.

1 Training performance analysis on NYUv2 dataset

We showcase the validation error at the Teacher end (SD-MTCNN) for segmentation (mIoU), depth perception (relative error), and surface-normal estimation (median error) of the NYUv2 dataset for 100 iterations (Figure 1) and compare the same with the performance of the vanilla Segnet (Baseline-1 main paper).

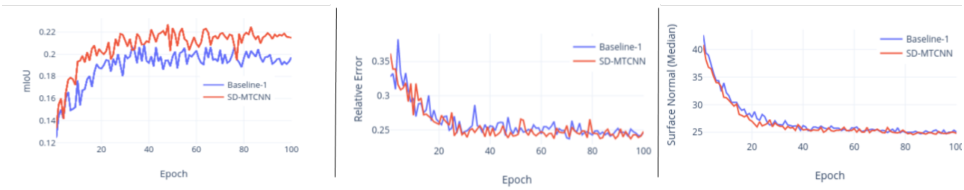

Figure 1: Performance graph between SD-MTCNN (Eq. 5 main paper) and SD-Vanilla Segnet (Baseline-1 main paper) for three tasks on the NYUv2 dataset. From left to right: mIoU (semantic segmentation), relative error (depth estimation), and median error (surface normal).

2 Baseline analysis for Mini-Taskonomy, NYUv2 (two tasks), and CityScapes

We mention the performance of different baseline approaches as detailed in Section 4.2 (main paper) for Mini-Taskonomy, NYUv2 (2 tasks), and CityScapes in Table 1 and 2, respectively. Figure 2 showcases the performance comparison of SD-MTCNN with the traditional distillation strategy for a case from the CityScapes dataset qualitatively.

Method	Segmentation ↑		Depth error ↓		SN ↑	Key ↓	Edge ↓
	IoU	mIoU	Abs.	Rel.	CS	Abs.	Abs.
Vanilla Segnet†	89.04	50.38	0.0336	0.3607	0.8805	0.0153	0.0103
SD-Vanilla Seg.^{T} $(\star)$‡	89.30	53.53	0.0301	0.3628	0.8988	0.0149	0.0108
SD-Vanilla Seg.^{S}	88.39	48.93	0.0432	0.4150	0.8543	0.0343	0.0122
SD-Vanilla Seg.^{T} $(\star\star)$‡	89.17	49.88	0.0352	0.3755	0.8896	0.0192	0.0108
SD-Vanilla Seg.^{S}	80.38	10.27	0.0744	2.8007	0.8482	0.0469	0.0764
Trad. K.D.∓	88.70	48.82	0.0293	0.3987	0.8818	0.0153	0.0105
SD-MTCNN^{T} $(\star)$#	89.03	54.87	0.0325	0.4619	0.8933	0.0151	0.0102
SD-MTCNN^{S}	88.90	54.21	0.0418	0.5523	0.8659	0.0331	0.0120
SD-MTCNN^{T} $(\star\star)$#	88.97	55.19	0.0331	0.5683	0.8852	0.0173	0.0106
SD-MTCNN^{S}	88.76	54.07	0.0423	1.0034	0.8664	0.0335	0.0131
SD-MTCNN^{T} $(full)\star$	**89.36**	**55.36**	**0.0290**	**0.3572**	**0.9058**	**0.0138**	**0.0097**
SD-MTCNN^{S}	89.02	54.78	0.0370	0.4010	0.8608	0.0327	0.0118
SD-MTCNN^{T} $(\star\star) full$	**89.21**	**55.26**	**0.0327**	0.4318	**0.8907**	**0.0147**	**0.0101**
SD-MTCNN^{S}	88.93	53.68	0.0419	0.5221	0.8616	0.0340	0.0129
Ablation on U-Net							
U-Net	89.11	52.76	0.0328	0.5294	0.8901	**0.0121**	0.0105
U-Net^{T} $(\star) full$	**89.61**	**55.96**	**0.0277**	**0.3519**	**0.9080**	0.0125	**0.0099**
U-Net^{S}	89.37	55.17	0.0322	0.5213	0.8730	0.0211	0.0121
U-Net^{T} $(\star\star) full$	**89.17**	**55.02**	**0.0272**	**0.3883**	**0.8942**	0.0133	**0.0104**
U-Net^{S}	88.54	54.39	0.0367	0.5575	0.8691	0.0237	0.0136

Table 1: 5-task validation results on the Mini-Taskomomy dataset for semantic segmentation, depth estimation, surface normal prediction, key-point estimation, and edge prediction on Segnet based models and ablation analysis on U-Net architecture. ^{T} Teacher, ^{S} Student. † Baseline-1, ‡ Baseline-2, ∓ Baseline-3, # Baseline-4, $(\star)$ by Eq. 5, $(\star\star)$ by Eq. 6. We compare SD-MTCNN$(\star)$ against all the $(\star)$ baselines and similar for $(\star\star)$.

Dataset	NYUv2				CityScapes			
Method	Segmentation ↑		Depth error ↓		Segmentation ↑		Depth error ↓	
	IoU	mIoU	Abs.	Rel.	IoU	mIoU	Abs.	Rel.
Vanilla Segnet†	50.64	14.90	0.6244	0.2612	89.73	49.71	0.0161	35.91
SD-Vanilla Seg.^{T} ($\star$)‡	56.52	18.44	0.5793	0.2502	92.37	54.61	0.0137	28.97
SD-Vanilla Seg.^{S}	52.77	15.79	0.6083	0.2576	89.60	49.11	0.0164	37.94
SD-Vanilla Seg.^{T} ($\star\star$)‡	54.37	16.94	0.6051	0.2468	90.44	52.02	0.0148	27.94
SD-Vanilla Seg.^{S}	23.06	6.55	1.7556	0.6943	49.98	14.83	0.0769	214.8
Trad. K.D.∓	50.92	15.76	0.6201	0.2689	89.78	49.87	0.0156	27.15
SD-MTCNN^T ($\star$)#	56.32	21.22	0.6885	0.2560	92.13	55.48	0.0149	**6.98**
SD-MTCNN^S	55.93	21.64	0.7066	0.2683	91.12	52.54	0.0164	14.06
SD-MTCNN^T ($\star\star$)#	56.16	21.89	0.6897	0.2620	91.64	54.03	0.0148	9.82
SD-MTCNN^S	55.84	20.35	0.7166	0.2662	90.96	51.94	0.0164	12.40
SD-MTCNN^T ($\star$) $full$	**57.18**	23.01	**0.5847**	**0.2466**	**92.54**	**56.70**	**0.0131**	27.68
SD-MTCNN^S	55.80	**23.19**	0.6033	0.2588	91.60	54.09	0.0150	33.23
SD-MTCNN^T ($\star\star$) $full$	**56.29**	**22.63**	**0.6042**	0.2544	91.57	**54.63**	**0.0134**	31.11
SD-MTCNN^S	56.09	22.50	0.6103	0.2674	90.99	53.27	0.0151	34.82

Table 2: 2-task validation results on the NYUv2 and CityScapes datasets for 13-classes and 7-classes resp. with semantic segmentation and depth estimation on Segnet based models. ^{T} Teacher, ^{S} Student. † Baseline-1, ‡ Baseline-2, ∓ Baseline-3, # Baseline-4, ($\star$) by Eq. 5, ($\star\star$) by Eq. 6. We compare SD-MTCNN($\star$) against all the ($\star$) baselines and similar for ($\star\star$).

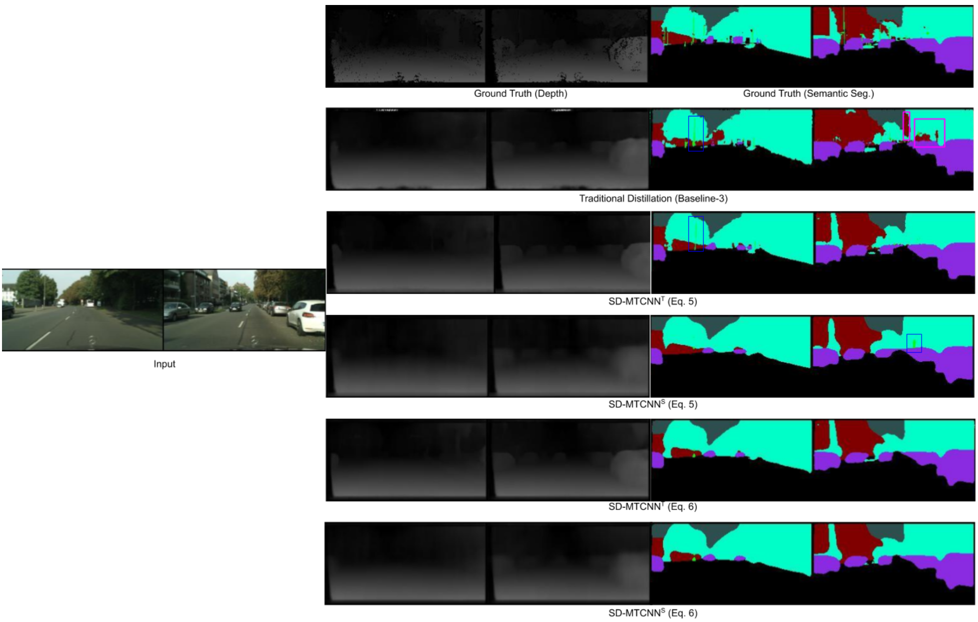

Figure 2: Visualization of segmentation and depth estimation on the CityScapes dataset. From top to bottom: ground truth, and predictions of Baseline-3 (Trad. Dist.), SD-MTCNN^T and SD-MTCNN^S (by Eq. 5 and Eq. 6 our model), respectively. The blue boxes show the correct predictions, whereas red boxes show the wrong predictions.