Appendix

Comparison With MTL Methods

As mentioned in the Empricial Evaluation Section, DiSparse surpasses several dedicated multitask learning approaches despite the high sparsity enforced in our model. In Table 1, we show comparison of DiSparse in both static and dynamic sparse training setting with several MTL approaches including DEN [1], Sluice [10], and Cross-Stitch [8] applied on exactly the same model with the same optimization settings. The superiroty of DiSparse is clearly observed in the table, demonstrating that DiSparse is not only an effective compression approach but also a powerful tool for multitask learning.

Model	T1: Semantic Seg.		T2: SN Prediction					Sparsity	Pre-trained
	mIoU↑	PixelAcc↑	Mean Err↓	Median Err↓	11.25↑	22.5↑	30↑	(%)↑	
Cross-Stitch [8]	25.3	57.4	16.6	13.2	43.7	72.4	83.8	0	✗
Sluice [10]	26.6	59.1	16.6	**13.0**	**44.1**	73.0	83.9	0	✗
DEN [1]	26.3	58.8	17.0	14.3	39.5	72.2	84.7	0	✗
DiSparse(*Static*)	26.5	57.8	**16.4**	13.7	41.2	**74.1**	**85.9**	**90**	✗
DiSparse (*Dynamic*)	**28.2**	**59.2**	16.5	13.5	**42.3**	73.1	84.7	**90**	✗

Table 1. DiSparse semantic segmentation and surface normal prediction results on NYU-v2 [9] compared to other MTL approaches.

Model	T1: Semantic Seg.		T2: SN Prediction					Sparsity	Pre-trained
	mIoU↑	PixelAcc↑	Mean Err↓	Median Err↓	11.25↑	22.5↑	30↑	(%)↑	
DeepLab [2](baseline)	27.69	58.77	16.55	14.17	39.62	73.54	86.33	0	N/A
LTH [5]	23.84	56.35	16.81	13.84	40.91	72.31	**84.28**	30.00	✓
SNIP [7]	26.57	59.85	16.91	13.55	42.01	71.72	82.01	30.00	✗
Random	25.08	55.56	17.60	14.27	40.49	70.12	81.68	30.00	✗
DiSparse (*Ours*)	**28.24**	**60.33**	**16.62**	**13.37**	**42.98**	**72.29**	83.96	30.00	✗
RigL [4]	24.83	57.92	16.78	14.84	37.76	72.18	86.15	30.00	✗
DiSparse (*Ours*)	**28.41**	**59.77**	**16.54**	**13.48**	**43.42**	**73.55**	**86.76**	30.00	✗
IMP [6]	29.23	59.83	16.57	13.38	43.16	72.41	**84.14**	30.00	✓
Random	26.43	58.25	16.89	13.71	41.92	71.72	83.77	30.00	✓
DiSparse (*Ours*)	**29.44**	**59.98**	**16.56**	**13.35**	**43.21**	**72.25**	84.06	30.00	✓

Table 2. DiSparse semantic segmentation and surface normal prediction results on NYU-v2 [9] compared to static sparse training, dynamic sparse training, and pre-trained model pruning methods.

Model	T1: Semantic Seg.			T2: Depth Prediction					Sparsity	Pre-trained
	mIoU ↑	PixelAcc ↑	Error ↓	Abs. Error ↓	Rel. Error ↓	$\delta 1.25$ ↑	$\delta 1.25^2$ ↑	$\delta 1.25^3$ ↑	(%) ↑	
DeepLab [2](baseline)	42.58	74.84	0.49	0.016	0.33	74.22	88.90	94.47	0	N/A
LTH [5]	40.21	72.59	0.51	0.017	0.36	72.54	87.39	93.69	30.00	✓
SNIP [7]	41.03	**74.65**	0.51	0.018	0.36	74.80	**89.53**	94.53	30.00	✗
Random	38.17	72.77	0.52	0.019	0.38	72.83	83.73	92.33	30.00	✗
DiSparse (*Ours*)	**42.34**	74.55	**0.49**	**0.016**	**0.33**	**74.91**	89.22	**94.62**	30.00	✗
RigL [4]	40.68	74.40	0.51	0.018	0.36	72.47	**87.39**	93.53	30.00	✗
DiSparse (*Ours*)	**42.53**	**74.82**	**0.49**	**0.016**	**0.33**	**74.62**	85.96	**93.73**	30.00	✗
IMP [6]	42.39	72.73	0.51	**0.016**	0.36	72.96	87.80	93.77	30.00	✓
Random	40.14	74.41	0.52	0.018	0.39	72.38	87.90	93.85	30.00	✓
DiSparse (*Ours*)	**42.47**	**74.69**	**0.50**	**0.016**	**0.34**	**73.32**	**88.46**	**94.37**	30.00	✓

Table 3. DiSparse semantic segmentation and depth prediction results on Cityscapes [3] compared to static sparse training, dynamic sparse training, and pre-trained model pruning methods.

Results at Lower Sparsity Levels

In the Empricial Evaluation Section, we showed the results of DiSparse and other pruning and sparse training approaches at high sparsity level(90%). Here, in Table 2, 3, we show the results at a sparsity of 30%, demonstrating the superiority of DiSparse at low sparsity level as well. From the table, we can see that DiSparse is better across all configurations and evaluation metrics. Moreover, we observed that DiSparse achieved lossless compression performance, achieving close or even better performance than the baseline unsparsified model.

References

[1] Chanho Ahn, Eunwoo Kim, and Songhwai Oh. Deep elastic networks with model selection for multi-task learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6529–6538, 2019. 1

[2] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 40(4):834–848, 2017. 1

[3] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3213–3223, 2016. 1

[4] Utku Evci, Trevor Gale, Jacob Menick, Pablo Samuel Castro, and Erich Elsen. Rigging the lottery: Making all tickets winners. In *International Conference on Machine Learning*, pages 2943–2952. PMLR, 2020. 1

[5] Jonathan Frankle and Michael Carbin. The lottery ticket hypothesis: Finding sparse, trainable neural networks. In *International Conference on Learning Representations*, 2018. 1

[6] Song Han, Huizi Mao, and William J Dally. Deep compression: Compressing deep neural networks with pruning, trained quantization and huffman coding. *arXiv preprint arXiv:1510.00149*, 2015. 1

[7] Namhoon Lee, Thalaiyasingam Ajanthan, and Philip Torr. Snip: Single-shot network pruning based on connection sensitivity. In *International Conference on Learning Representations*, 2018. 1

[8] Ishan Misra, Abhinav Shrivastava, Abhinav Gupta, and Martial Hebert. Cross-stitch networks for multi-task learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3994–4003, 2016. 1

[9] Pushmeet Kohli Nathan Silberman, Derek Hoiem and Rob Fergus. Indoor segmentation and support inference from rgbd images. In *ECCV*, 2012. 1

[10] Sebastian Ruder, Joachim Bingel, Isabelle Augenstein, and Anders Søgaard. Latent multi-task architecture learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 4822–4829, 2019. 1