Supplementary Material
Modality-Incremental Learning with Disjoint Relevance Mapping Networks for Image-based Semantic Segmentation

Niharika Hegde [1,2] * Shishir Muralidhara [2] * René Schuster [1,2] Didier Stricker [1,2]
[1] RPTU – University of Kaiserslautern-Landau
[2] DFKI – German Research Center for Artificial Intelligence
firstname.lastname@dfki.de

A. Introduction

In this supplementary material to our paper *Modality-Incremental Learning with Disjoint Relevance Mapping Networks for Image-based Semantic Segmentation*, we show the impact of forgetting on previously learned modalities, test the robustness of Disjoint Relevance Mapping Networks (DRMNs) against variation of the pruning parameter μ, and list the exact utilization of network connections for the experiment on InfraParis [12].

B. Task-wise Evaluation

To quantify the amount of forgetting due to the incremental learning of different modalities, Tab. 1 provides the mIoU for each modality through the learning sequence. *I.e.*, each known modality is evaluated after each task. This way, the mutual negative influence of the modalities can be measured. With regularization-based approaches such as EWC [27] and ILT [31], the model learns optimally during the initial step as expected. However, when learning other modalities incrementally, EWC prevents overwriting important parameters from the previous modalities, hindering its learning on the new modality. On the other hand, ILT which uses distillation, exhibits better performance on the initial task compared to EWC. However, the performance on new modalities is significantly worse due to the diverse nature of the modalities. In RMN [25] and the proposed DRMN, even for the initial modality the results are slightly lower compared to the single-task models. This is due to the use of relevance maps, which preserve network capacity for future tasks by not utilizing the entire network capacity at each step. This approach effectively preserves information and completely mitigates catastrophic forgetting, ensuring that performance on previously learned modalities remains consistent over the sequence of tasks. Additionally, with DRMN, isolating parameters and learning task-specific weights enhances the learning of new modalities, as evident in improved performance on both Gray and RGB tasks.

* These authors have contributed equally to this work.

C. Relevance Map Pruning

To recall, the hyperparameter μ defines the threshold at which network weights (connections) are considered relevant. Any connection below this threshold will be pruned after every epoch above 50. The values in the relevance map of the pruned connections will be permanently set to zero for this task, removing the influence of that connection entirely. The unused connections might be used in a later task, though. The network's weights for relevant connections will be frozen. The default value for μ is 0.6. In Tab. 2, we show the results for a threshold of 0.5 and 0.7. The variation of μ in both directions indicates a high robustness of DRMN in this regard. Same holds for the original RMN. Interestingly, we point out that varying the pruning parameter has no significant impact on the sparsity (utilization) of neural connections.

D. Task Utilization on InfraParis

One of our claims in the main paper is that despite the strict separation of task-specific connections, the network's capacity is not exceeded faster than with regular RMNs. To back this claim further, we have also computed the network utlization for the four tasks of InfraParis [12]. The result is shown in Tab. 3. For a description of ORMN and PRMN, we refer to Sec. 5.5 of the main paper. It is striking that even with just about 6 % of the network's overall connections, the final task can be learned even better than with RMN, which uses about half of all weights. Another remarkable observation is that each task approximately consumes half of the remaining connections in DRMN.

Table 1. Results on Freiburg Thermal [39] for the original RMN [25] and our proposed DRMN after each step of training the sequence (***IR*** $\rightarrow$ ***Gray*** $\rightarrow$ ***RGB***).

Method	M_0 (IR)	M_1 (Gray)		M_2 (RGB)		
	IR	IR	Gray	IR	Gray	RGB
Fine Tuning	59.56	07.44	74.10	07.12	73.17	75.24
EWC [27]	59.75	06.89	58.04	10.56	59.47	61.69
ILT [31]	59.56	20.39	21.08	20.68	22.57	23.53
RMN [25]	55.30	55.18	68.85	55.10	68.90	69.46
DRMN (Ours)	55.30	55.16	70.61	54.97	70.56	71.19

Table 2. Results and task-wise network utilization on Freiburg Thermal [39] for the original RMN [25] and our proposed DRMN with varying pruning parameters.

Method	**Prune** μ	**Results (mIoU)**				**Task Utilization (%)**			**Overall (%)**	
		IR	Gray	RGB	**Avg**	IR	Gray	RGB	Shared Weights	Network Utilization
RMN [25]	0.5	55.02	69.06	68.96	64.35	49.91	49.85	49.83	49.80	87.39
	0.6	55.10	68.90	69.46	64.49	47.79	47.78	47.80	46.68	85.77
	0.7	54.25	68.33	68.66	63.75	49.10	49.03	49.08	48.62	86.78
DRMN (Ours)	0.5	55.23	70.64	71.21	65.69	49.91	24.94	12.50	0.00	87.35
	0.6	54.97	70.56	71.19	65.57	47.79	24.95	13.03	0.00	85.78
	0.7	54.34	70.69	71.05	65.36	49.10	24.93	12.73	0.00	86.76

Table 3. Network utilization on InfraParis [12] for the original RMN [25] and our proposed DRMN on the task sequence (***IR*** $\rightarrow$ ***RGB*** $\rightarrow$ ***Depth*** $\rightarrow$ ***Gray***).

Method	**Task Utilization (%)**				**Overall (%)**	
	IR	RGB	Depth	Gray	Shared Weights	Network Utilization
RMN [25]	49.52	49.54	49.49	49.54	68.02	93.50
ORMN	49.52	49.46	49.40	49.41	67.91	93.47
PRMN	49.52	27.16	15.84	10.16	5.97	93.48
DRMN (Ours)	49.52	24.98	12.61	6.37	0.00	93.49

References

[12] Gianni Franchi, Marwane Hariat, Xuanlong Yu, Nacim Belkhir, Antoine Manzanera, and David Filliat. InfraParis: A multi-modal and multi-task autonomous driving dataset. In *Winter Conference on Applications of Computer Vision (WACV)*, 2024. 1, 2

[25] Prakhar Kaushik, Alex Gain, Adam Kortylewski, and Alan Yuille. Understanding catastrophic forgetting and remembering in continual learning with optimal relevance mapping. In *arXiv preprint arXiv:2102.11343*, 2021. 1, 2

[27] James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, *et al*. Overcoming catastrophic forgetting in neural networks. In *Proceedings of the national academy of sciences*, 2017. 1, 2

[31] Umberto Michieli, and Pietro Zanuttigh. Knowledge distillation for incremental learning in semantic segmentation. In *Computer Vision and Image Understanding*, 2021. 1, 2

[39] Johan Vertens, Jannik Zürn, Wolfram Burgard. HeatNet: Bridging the Day-Night Domain Gap in Semantic Segmentation with Thermal Images. In *arXiv preprint arXiv:2003.04645*, 2020. 2