Paper PDF

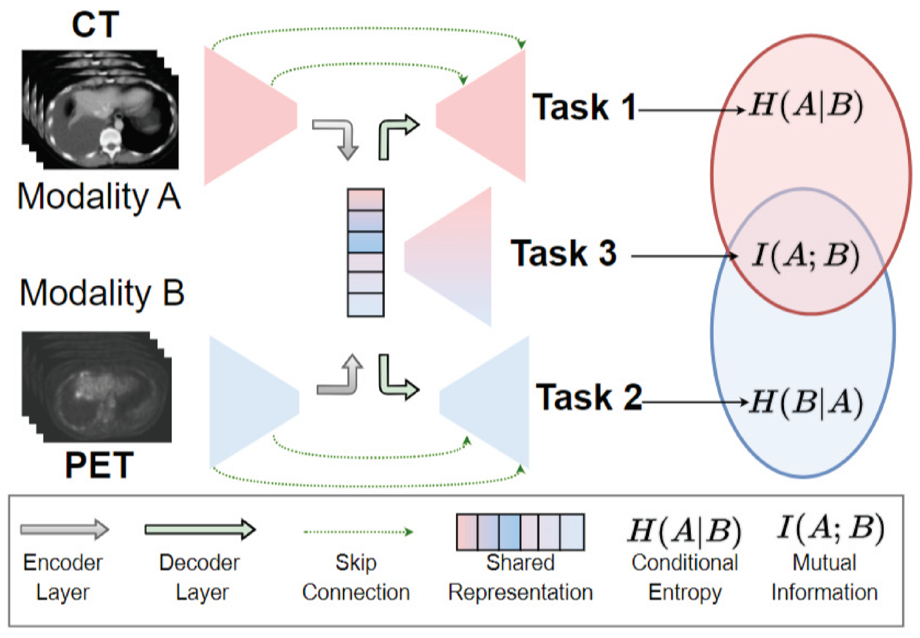
Mirror U-Net: Marrying Multimodal Fission with Multi-task Learning for Semantic Segmentation in Medical Imaging

Zdravko Marinov[*,1,3], Simon Reiß[1], David Kersting[2,5,7], Jens Kleesiek[4,7], Rainer Stiefelhagen[1]

[1] Karlsruhe Institute of Technology, Karlsruhe, Germany, [2] German Cancer Consortium (DKTK), Germany
[3] HIDSS4Health – Helmholtz Information and Data Science School for Health, Karlsruhe/Heidelberg, Germany,
[4] Institute for AI in Medicine, University Hospital Essen, Essen, Germany,
[5] Department of Nuclear Medicine, University Hospital Essen, [6] West German Cancer Center, Germany
[7] Cancer Research Center Cologne Essen (CCCE), University Medicine Essen, Essen, Germany
* **zdravko.marinov@kit.edu**

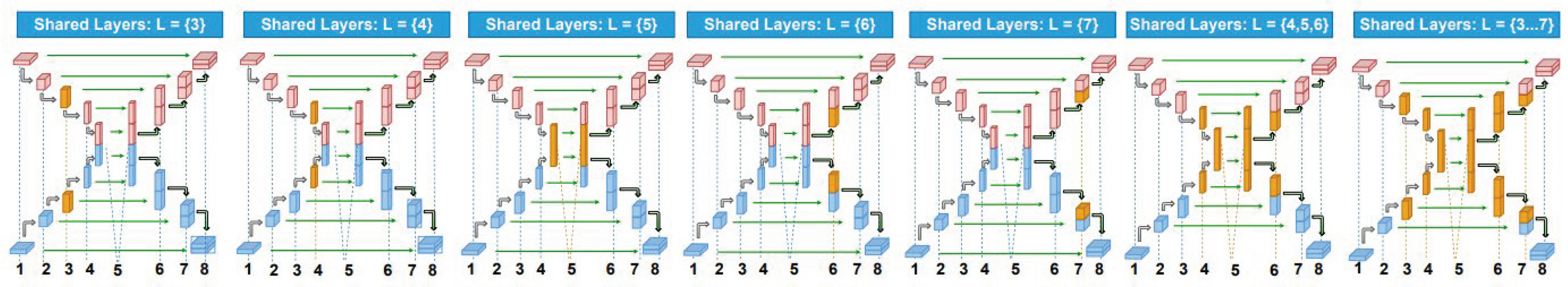
1 Aim

Traditional fusion approaches (early, mid-, and late fusion) do not perform well on multimodal PET/CT data for lesion segmentation since CT provides a very weak signal. We propose to combine **multimodal fission*** with **multi-task** learning to extract useful features from the CT and boost the segmentation on the AutoPET dataset [1].

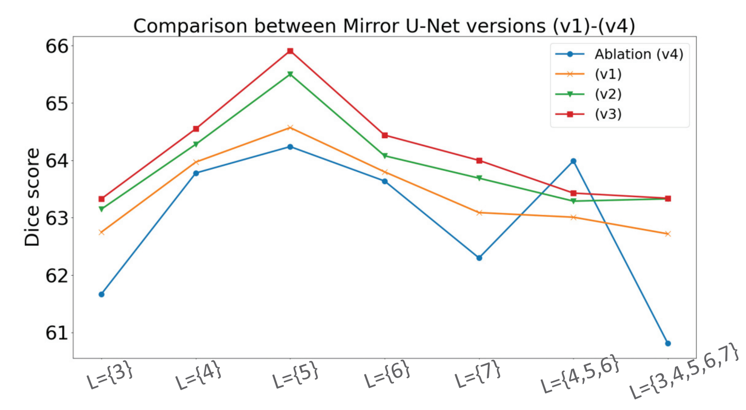


*Multimodal fission: Fusion followed by factorized or partitioned features.

3 Weight Sharing Experiments

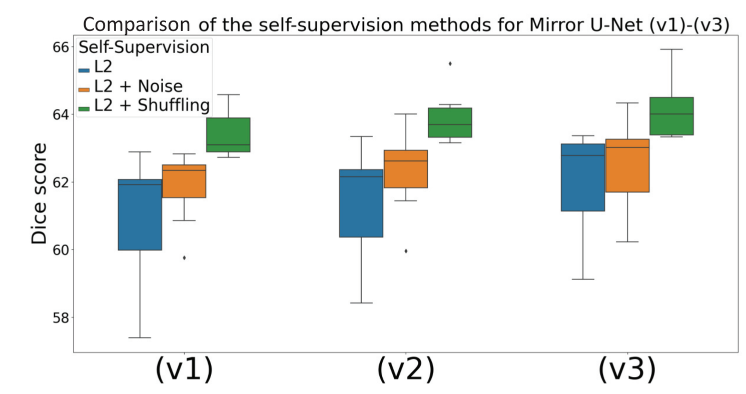


4 Results for (v1) – (v4)



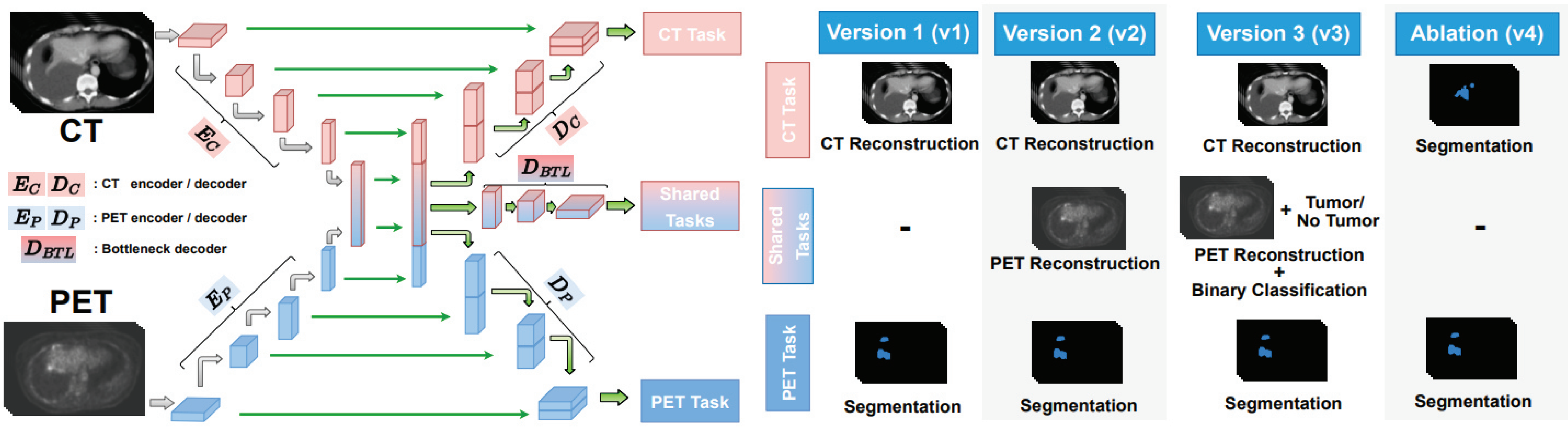
→ **(v3)** is consistently best
→ The bottleneck is the best layer to share

5 Self-supervision Tasks



→ Voxel shuffling is the best task with the least variance

2 Multi-task Settings



6 Comparison to Related Work

Method	Dice ↑	FPV ↓	FNV ↓	Tasks	Multimodal Fission	Multi-task
nnUNet [17]	62.75	2.83	1.59	Seg		
Blackbean [47]	63.15	2.55	1.76	Seg		
SF-Net [27]	61.21	3.44	2.95	Seg + Rec		✓
Andrearczyk et al. [1]	61.45	2.98	1.89	Seg + Class		✓
DeepMTS [31]	61.91	3.22	2.76	Seg + Class		✓
Weninger et al. [44]	61.22	3.98	2.82	Seg + Rec + Class		✓
CT-only Mirror U-Net **(v3)**	12.37	28.24	50.02	Seg + Rec + Class		✓
PET-only Mirror U-Net **(v3)**	56.14	4.81	3.02	Seg + Rec + Class		✓
Mirror U-Net **(v4)**	64.24	2.93	1.99	Seg	✓	
Valindria *et al.* [42]	39.84	7.89	17.00	Seg	✓	
(Ours) Mirror U-Net **(v3)**	**65.91**	**1.55**	**0.76**	Seg + Rec + Class	✓	✓

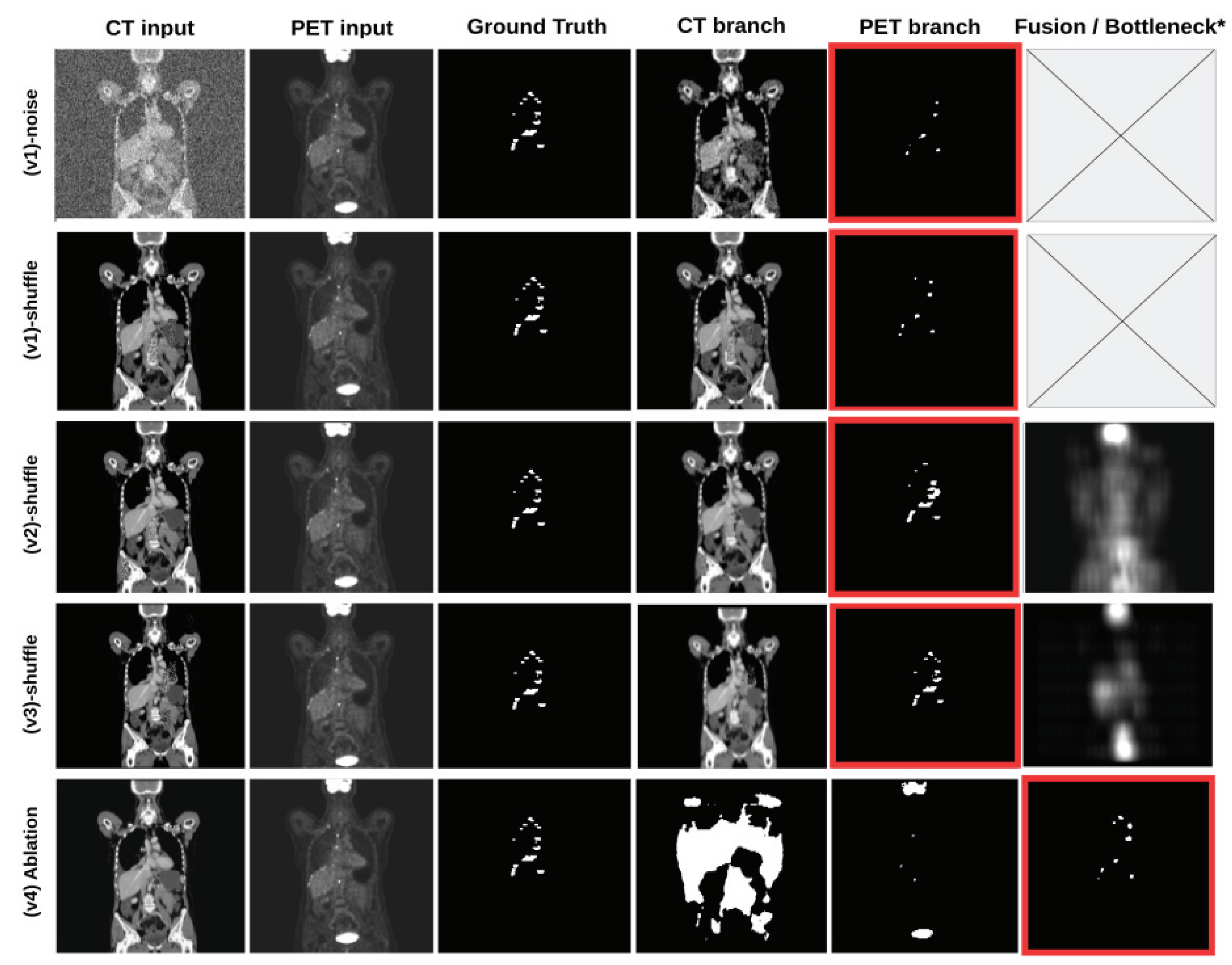
FPV: False Positive Volume, FNV: False Negative Volume

7 Comparison to Traditional Fusion Strategies

Metric	Baselines							Mirror U-Net (Ours)			
	CT	PET	EF	MF	LF-Logit	LF-∪	LF-∩	**(v1)**	**(v2)**	**(v3)**	Ablation **(v4)**
Dice ↑	26.00	**60.99**	54.89	55.53	57.41	59.89	21.60	64.57	65.50	**65.91**	64.24
FPV ↓	15.64	5.38	4.98	4.77	4.88	3.95	**1.67**	2.93	2.83	**1.55**	2.93
FNV ↓	44.15	**2.15**	3.13	3.02	2.88	3.01	99.74	1.66	0.94	**0.76**	1.99

EF: Early Fusion, MF: Mid Fusion, LF: Late Fusion

8 Qualitative Results



9 Conclusion

- Traditional fusion methods do not utilize the information in the CT and lead to overfitting. However, combining multimodal fission with multi-task learning significantly improves the performance on AutoPET [1]
- The modality-specific tasks must be chosen carefully
- The bottleneck proved to be the best location to share features in all experiments

10 References

[1] Gatidis, Sergios, et al. "The autoPET challenge: Towards fully automated lesion segmentation in oncologic PET/CT imaging." (2023)
[2] Valindria, Vanya V., et al. "Multi-modal learning from unpaired images: Application to multi-organ segmentation in CT and MRI." WACV (2018)
[3] Isensee, Fabian, et al. "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation." Nature methods (2021)
[4] Weninger, Leon. "Multi-task learning for brain tumor segmentation." BrainLes MICCAI Workshop (2019)

Acknowledgements: This project is supported by the Helmholtz Association under the joint research school "HIDSS4HealthHelmholtz Information and Data Science School for Health".