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

[1,3]Zdravko Marinov* [1]Simon Reiß* [2,5,6]David Kersting† [4,7]Jens Kleesiek† [1]Rainer Stiefelhagen*

[1]Institute for Anthropomatics and Robotics (IAR), Karlsruhe Institute of Technology, [2]German Cancer Consortium (DKTK)

[3]HIDSS4Health - Helmholtz Information and Data Science School for Health

[4]Institute for AI in Medicine (IKIM), University Hospital Essen, [5]Department of Nuclear Medicine, University Hospital Essen

[6]West German Cancer Center, [7]Cancer Research Center Cologne Essen (CCCE), University Medicine Essen

*firstname.lastname@kit.edu

†firstname.lastname@uk-essen.de

1. Supplementary

This supplementary document is intended to provide additional experiments, discussions, and figures that were not included in the main manuscript due to space limitations.

2. BrainTumor Pipeline

Figure 2 illustrates our multi-task Mirror U-Net pipeline for brain tumor segmentation. The CT and PET tasks are replaced with FLAIR and T1Gd tasks respectively. We believe that it makes more sense to assign segmentation tasks for each branch since unlike CT, both FLAIR and T1Gd have a strong signal for tumor lesion boundaries. However, for consistency, we also include **(v2)-rec** following the same default tasks as **(v2)** in our AutoPET [4] experiments.

3. λ Hyperparameters

We set the λ_{rec}, λ_{seg}, λ_{cls} parameters following previous work [9] that utilizes multimodal **fusion**, where $\lambda_{\text{seg}} = 1.0$, $\lambda_{\text{rec}} = 0.1$, $\lambda_{\text{cls}} = 0.1$ are used with a higher λ_{seg} since segmentation is the primary task. However, we explore further reducing the influence of λ_{rec} and λ_{cls} in a grid search over $\{0.1, 0.01, 0.001, 0.0001, 0.00001\}$. We find that there is a sweet spot for both parameters (see Figure 1), however, the difference in performance is not substantial and Mirror U-Net is robust to λ-parameter changes in $[0.1 - 0.00001]$.

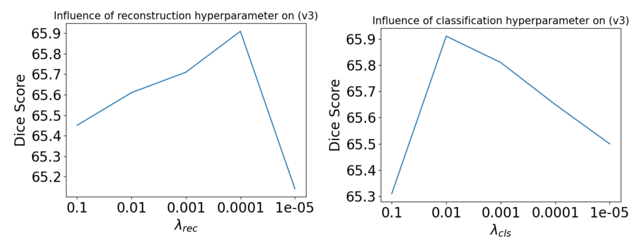

Figure 1. Dice Scores of **(v3)** when varying λ_{rec} and λ_{cls}.

4. Discussion

Statement of Novelty. Mirror U-Net is a novel combination of multimodal **fission** and multi-task learning. Unlike previous methods, which either combine multimodal **fusion** with multi-task learning [9, 1, 6, 7] or employ fission-only models [8], Mirror U-Net conditions the fission of features with explicitly defined tasks. We compare Mirror U-Net to a plethora of such approaches [6, 7, 1, 9, 8] to show that this combination of fission and multi-task learning is essential.

Clinical Relevance. The segmentation of paired PET/CT volumes can aid the diagnosis by estimating the metabolic tumor volume (MTV) [10] as well as the affected anatomical regions [5], predicting an exact tumor location and size. The separation of anatomical knowledge in CT and physiological knowledge in PET is often seen in clinical practice [3] where PET is used to identify highly active lesions and CT is used to localize which anatomy is affected. Mirror U-Net's state-of-the-art performance on AutoPET [4] as well as the possibility to assign explicit tasks to each modality, which correspond to the tasks for which the modalities are used in clinical practice, brings us a step closer to deploying segmentation models in practice.

Limitations. To assign the optimal tasks to each modality branch and the shared bottleneck decoder, one has to consider which tasks make the most sense for each modality, e.g., reconstruction for CT as CT has fine structures but little information about the lesions. For MSD BrainTumor [2] using segmentation tasks for different parts of the tumor led to the best results. However, we did achieve good per-

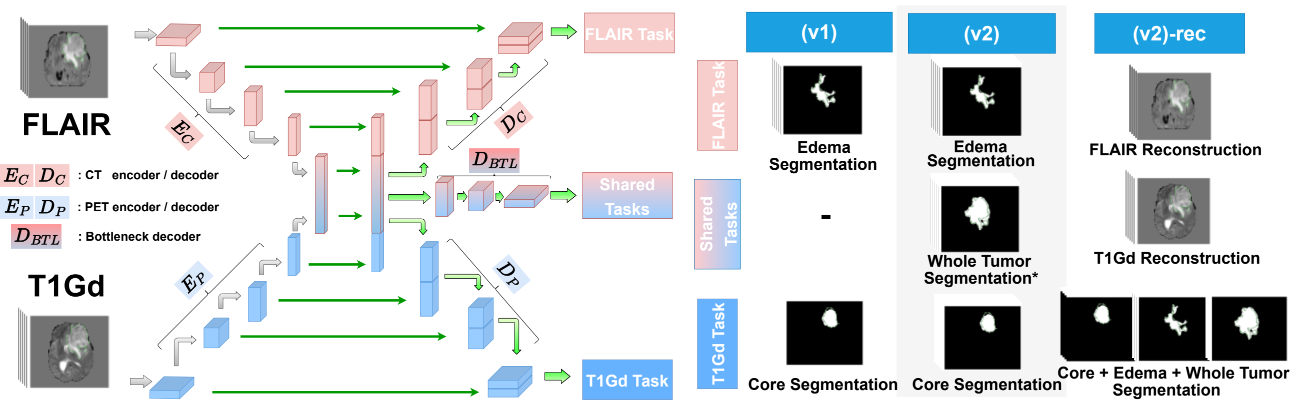

Figure 2. Overview of the tasks used in our MSD BrainTumor [2] experiments. In **(v1)** and **(v2)** we use the union of the edema and core predictions as our final whole tumor prediction mask.*The whole tumor from the bottleneck at **(v2)** is only used as regularization. In **(v2)-rec** we use the default tasks we used in AutoPET [4] for consistency with our AutoPET [4] experiments. We segment all three tumor classes in the T1Gd branch.

formance with the default **(v2)** AutoPET [4] tasks also for **(v2)-rec** on MSD BrainTumor [2].

Future Work. We plan to explore additional task combinations to further improve the effectiveness of Mirror U-Net. For example, we will investigate the potential of adding more anatomical information to the CT branch, such as multi-organ segmentation and incorporating tumor classification (lymphoma, melanoma, lung cancer) from the bottleneck. Additionally, we aim to extend Mirror U-Net to **interactive segmentation**, where human feedback can be incorporated to refine the segmentation results. Specifically, we will explore the use of click-based interactions to correct predictions from each branch individually, leveraging the low parameter sharing between branches and the modality-specific skip connections.

References

[1] Vincent Andrearczyk, Pierre Fontaine, Valentin Oreiller, Joel Castelli, Mario Jreige, John O Prior, and Adrien Depeursinge. Multi-task deep segmentation and radiomics for automatic prognosis in head and neck cancer. In *International Workshop on PRedictive Intelligence In MEdicine*, pages 147–156. Springer, 2021. 1

[2] Michela Antonelli, Annika Reinke, Spyridon Bakas, Keyvan Farahani, Annette Kopp-Schneider, Bennett A Landman, Geert Litjens, Bjoern Menze, Olaf Ronneberger, Ronald M Summers, et al. The medical segmentation decathlon. *Nature communications*, 13(1):4128, 2022. 1, 2

[3] Thomas Beyer, Gerald Antoch, Stefan Müller, Thomas Egelhof, Lutz S Freudenberg, Jörg Debatin, and Andreas Bockisch. Acquisition protocol considerations for combined pet/ct imaging. *Journal of Nuclear Medicine*, 45(1 suppl):25S–35S, 2004. 1

[4] Sergios Gatidis, Tobias Hepp, Marcel Früh, Christian La Fougère, Konstantin Nikolaou, Christina Pfannenberg, Bernhard Schölkopf, Thomas Küstner, Clemens Cyran, and Daniel Rubin. A whole-body fdg-pet/ct dataset with manually annotated tumor lesions. *Scientific Data*, 9(1):1–7, 2022. 1, 2

[5] Russ A Kuker, David Lehmkuhl, Deukwoo Kwon, Weizhao Zhao, Izidore S Lossos, Craig H Moskowitz, Juan Pablo Alderuccio, and Fei Yang. A deep learning-aided automated method for calculating metabolic tumor volume in diffuse large b-cell lymphoma. *Cancers*, 14(21):5221, 2022. 1

[6] Yu Liu, Fuhao Mu, Yu Shi, and Xun Chen. Sf-net: A multi-task model for brain tumor segmentation in multimodal mri via image fusion. *IEEE Signal Processing Letters*, 29:1799–1803, 2022. 1

[7] Mingyuan Meng, Bingxin Gu, Lei Bi, Shaoli Song, David Dagan Feng, and Jinman Kim. Deepmts: Deep multi-task learning for survival prediction in patients with advanced nasopharyngeal carcinoma using pretreatment pet/ct. *IEEE Journal of Biomedical and Health Informatics*, 26(9):4497–4507, 2022. 1

[8] Vanya V Valindria, Nick Pawlowski, Martin Rajchl, Ioannis Lavdas, Eric O Aboagye, Andrea G Rockall, Daniel Rueckert, and Ben Glocker. Multi-modal learning from unpaired images: Application to multi-organ segmentation in ct and mri. In *2018 IEEE winter conference on applications of computer vision (WACV)*, pages 547–556. IEEE, 2018. 1

[9] Leon Weninger, Qianyu Liu, and Dorit Merhof. Multi-task learning for brain tumor segmentation. In *International MICCAI brainlesion workshop*, pages 327–337. Springer, 2019. 1

[10] Weina Xu, Shupeng Yu, Ying Ma, Changping Liu, and Jun Xin. Effect of different segmentation algorithms on metabolic tumor volume measured on 18f-fdg pet/ct of cervical primary squamous cell carcinoma. *Nuclear medicine communications*, 38(3):259, 2017. 1