LongiNet: Dual-Encoder Longitudinal Lesion Segmentation

Niels Rocholl[1 [0009−0002−3072−4109]], Ewoud Smit[1 [0000−0002−7090−2742]], and Alessa Hering[1 [0000−0002−7602−803X]]

Department of Medical Imaging, Radboudumc, Nijmegen, The Netherlands
niels.rocholl@radboudumc.nl

Abstract. We present LongiNet, a dual-encoder 3D network for longitudinal lesion segmentation. The baseline encoder takes the baseline CT image together with its baseline mask, while the follow-up encoder takes only the follow-up CT image. Encoders share weights from a pretrained nnU-Net (ULS23 baseline) and features are fused via 1x1x1 convolutions before decoding. A mandatory auxiliary baseline-mask reconstruction task is used during training to improve stability. Data are standardized by CT intensity clamping to [-1000, 400] and rescaling to [0,1], with lightweight spatial and intensity augmentations. Training uses Dice+CE loss, SGD with PolyLR and short transfer warmup. Validation uses a deterministic split. No test-time augmentation or ensembling is applied.

Keywords: autoPET challenge · longitudinal segmentation · nnU-Net · MONAI

1 Introduction

We address longitudinal lesion segmentation in CT where baseline and follow-up scans with lesion clickpoints are provided. Our method fuses baseline and follow-up representations to predict follow-up lesions concisely and robustly.

2 Methods

We follow the template and provide concise details; Table 1 summarizes key settings.

2.1 Data

Training data Longitudinal-CT dataset [2].

Validation data Deterministic split (val_split=0.2) created once and stored (index- and ID-based) for reproducible validation.

2.2 Data pre-processing

CT intensities clamped to [-1000, 400] and rescaled to [0,1]. Channels ensured first; masks binarized. Inputs are formed as follows: the baseline encoder receives two channels [baseline image, baseline mask], and the follow-up encoder receives one channel [follow-up image].

2.3 Algorithm/model

Dual-encoder 3D nnU-Net backbone (ULS) with shared weights for BL and FU streams [1]. Features are fused via 1x1x1 convolutions at all skip levels and bottleneck; decoded by nnU-Net decoder. A mandatory auxiliary baseline mask reconstruction branch is used during training to improve stability.

2.4 Data post-processing

Predictions are resampled into full-volume geometry with nearest-neighbor, then component-wise labeled using nearest clickpoint in physical space; final mask is binary.

2.5 Training and test parameters

Loss: Dice+CE (softmax, one-hot). Optimizer: SGD (momentum=0.99, nesterov), weight_decay=3e-5. LR: initial_lr=2.5e-3 with PolyLR (exp=0.9), max_epochs=1000; transfer warmup 3 epochs at 0.1×LR. Batch size=8, mixed precision (fp16), 4 GPUs (DDP). Augmentations: small affine rotations (±10°), Gaussian noise (std=0.02), intensity scale (±20%), shift (±10%), contrast (gamma 0.8–1.2), Gaussian smooth ($\sigma = 0.5 \dots 1.0$). Test: no TTA, threshold 0.5; no ensembling.

2.6 Github repository

Link to Github repository: https://github.com/DIAGNijmegen/oncology-longinet-container

3 Results

We used an 80/20 train/validation split (single run; no cross-validation). Figure 1 and Figure 2 summarize validation performance.

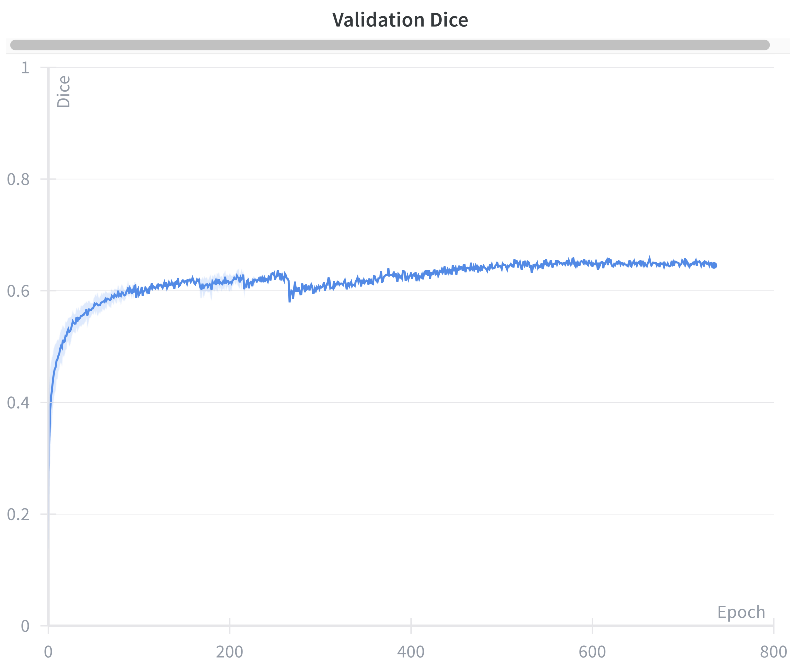

Fig. 1. Validation Dice over training.

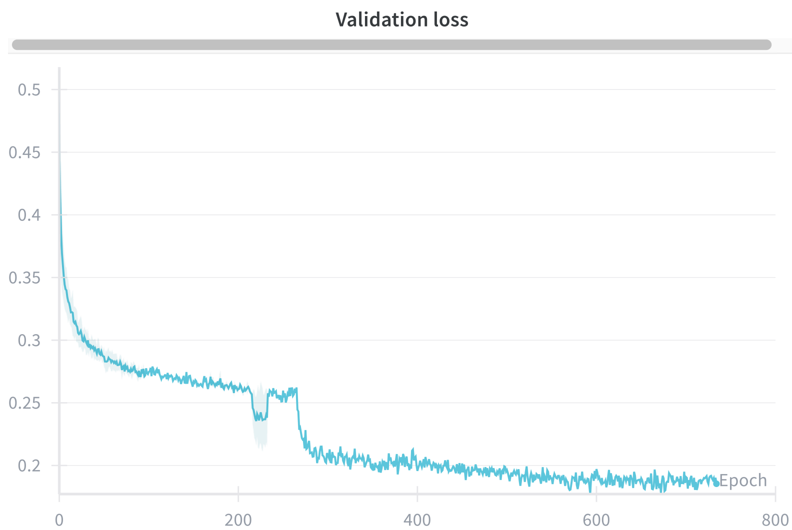

Fig. 2. Validation loss over training.

4 Discussion

During training, we observe improved performance compared to fine-tuning the pure ULS baseline. The dual-encoder fusion together with the auxiliary reconstruction task appears beneficial, but more experiments are required for conclusive results.

5 Conclusion

LongiNet delivers a concise, reproducible longitudinal segmentation pipeline suitable for challenge submission without TTA or ensembling.

Acknowledgments. None.

Disclosure of Interests. The authors have no competing interests to declare.

References

1. M. J. J. de Grauw, E. Th. Scholten, E. J. Smit, M. J. C. M. Rutten, M. Prokop, B. van Ginneken, A. Hering: The ULS23 Challenge: a Baseline Model and Benchmark Dataset for 3D Universal Lesion Segmentation in Computed Tomography (2024). https://arxiv.org/abs/2406.05231
2. Küstner, T., Peisen, F., Gatidis, S., Wagner, A., Megne, O., Othman, A., Sanner, A., Lossau, T., Moltz, J. H., Kohlbrandt, T., & Hering, A. (2025). Longitudinal-CT. University of Tübingen. https://doi.org/10.57754/FDAT.qwsry-7t837

Table 1. Algorithm details

Team name	**algorithm name** (as submitted on grand-challenge)	**data pre-processing**	**data post-processing**	**training data augmentation**
niels rocholl	LongiNet dual-encoder	CT clamp [-1000,400], rescale [0,1]	Resample to full, comp. labeling by nearest clickpoint	Small affine, noise, scale, shift, contrast, Gaussian smooth

test time augmentation	**ensembling**	**standardized framework?**	**network architecture**	**loss**
None	None	MONAI + nnU-Net v2 + Lightning	Dual-encoder UNet (3D)	Dice + CE

training data	**data/model dimensionality and size**	**use of pre-trained models**	**GPU hardware for training**
Longitudinal-CT dataset [2]	3D: 128x128x64 inputs (VOIs 64x128x128 at inference)	Pretrained nnU-Net (ULS23 baseline) [1]	1x Nvidia A100