2026 JULY 19–22 VANCOUVER, BC
JOINT AAPM | COMP MEETING

Overcoming Computational Bottlenecks In Large-Scale Medical Image Segmentation Using Optimized U-Net

Yunkui Pang[1]; Yirong Xu[1]; Pew-Thian Yap, PhD[1]; Jun Lian, PhD[1]
[1]University of North Carolina at Chapel Hill

ABSTRACT

Purpose: Training nnU-Net models on large medical imaging datasets is computationally intensive and can slow model development. We present an optimized nnU-Net training workflow designed to reduce training time while preserving multi-organ segmentation performance.
Methods: A total of 1,539 CT scans were used to train nnU-Net v2 models for segmentation of 117 anatomical structures. The optimized workflow incorporated mixed-precision training, dynamic batch-size scaling, and convergence-guided training. Network architecture, input resolution, and training data were unchanged. Both models were trained on a single NVIDIA L40S GPU and evaluated on 89 held-out CT scans using Dice similarity coefficient (DSC) and 95th-percentile Hausdorff distance (HD95).
Results: Baseline training required 97.4 GPU-hours, whereas the optimized workflow required 35.7 GPU-hours, corresponding to a 2.7-fold speedup and a 63.3% reduction in training time. Segmentation performance remained comparable across evaluated structures, with no systematic deterioration in DSC or HD95.
Conclusion: The optimized nnU-Net workflow substantially reduced training time while maintaining comparable multi-organ segmentation performance. This approach may enable faster model iteration and more efficient development of automated segmentation tools for radiation oncology.

CONTACT

Jun Lian, PhD, DABR
University of North Carolina at Chapel Hill
jun_lian@med.unc.edu
(984) 974-0400

INTRODUCTION

Background: Deep learning-based auto-segmentation supports treatment planning by enabling rapid delineation of multiple anatomical structures on CT images [1]. nnU-Net is a widely adopted self-configuring segmentation framework [2], but training on large clinical datasets can require substantial computational time. This limits model iteration and increases the cost of evaluating updated datasets, preprocessing strategies, and contouring protocols.
Objective: To determine whether workflow-level optimization can accelerate nnU-Net v2 training for large-scale multi-organ CT segmentation while maintaining comparable segmentation performance.

METHODS

Dataset and task
- 1,539 CT scans for training
- 117 anatomical structures
- 89 held-out CT scans for testing

Evaluation
- Single NVIDIA L40S GPU
- Total GPU training time
- DSC and HD95 on held-out test set
- Paired Wilcoxon signed-rank test

Baseline: Standard workflow
- Standard nnU-Net v2 workflow
- Default preprocessing, sampling, and training schedule

Optimized workflow
- BF16 mixed precision training
- Dynamic batch size scaling
- Efficiency-driven schedule with early convergence monitoring

RESULTS

Training efficiency: The optimized workflow reduced total GPU training time **from 97.4 to 35.7 hours.** Corresponds to 2.7× faster training, 63.3% less GPU time.
Segmentation performance: Across 15 sampled clinically important organs, segmentation quality was maintained between workflows, with no statistically significant degradation in either DSC or HD95. Mean DSC values were highly comparable between the baseline and efficient workflows for the prostate (0.794 vs. 0.797), brain (0.925 vs. 0.923), and right middle lung lobe (0.951 vs. 0.955). HD95 values were also comparable across all structures, further supporting preserved boundary and surface accuracy.
Overall pattern: Performance differences between workflows were small and structure-dependent, with volumetric overlap and surface agreement remaining consistent across evaluated structures. Qualitative review of representative cases also showed visually similar multi-organ contours between the baseline and optimized models.
Practical impact: The optimized workflow saved approximately **62 GPU-hours per training run**. This reduction could substantially increase the number of model-development experiments that can be completed within a fixed computational budget.

ID	Organs	Avg Volume (mm³)	Sample Size	Average DSC		p-value	Average HD95 (mm)		p-value
				Baseline	Efficient		Baseline	Efficient	
1	adrenal gland right	3568.8	65	0.871	0.869	0.363	2.238	2.487	0.894
2	adrenal gland left	4133.3	65	0.878	0.876	0.796	2.725	2.512	0.203
3	thyroid gland	14319.8	22	0.918	0.918	0.241	1.957	1.921	0.144
4	prostate	30825.8	77	0.794	0.797	0.136	4.857	5.517	0.283
5	duodenum	47704.3	39	0.826	0.825	0.285	11.234	8.776	0.194
6	spinal cord	54536.3	55	0.947	0.946	0.072	1.507	1.552	0.679
7	pancreas	59248.3	22	0.872	0.870	0.360	3.951	6.639	0.629
8	kidney left	129936.3	39	0.944	0.944	0.502	3.927	2.999	0.091
9	kidney right	133045.2	39	0.926	0.927	0.225	3.886	10.009	0.135
10	urinary bladder	222992.8	39	0.910	0.910	0.608	16.327	16.136	0.702
11	stomach	283063.7	81	0.936	0.936	0.941	5.872	5.938	0.500
12	spleen	287603.2	54	0.981	0.981	0.224	6.485	5.820	0.533
13	lung middle lobe right	312171.9	15	0.951	0.955	**0.013**	4.019	3.658	0.421
14	heart	463566.9	60	0.900	0.900	0.345	7.393	12.237	0.128
15	brain	790446.9	71	0.925	0.923	0.133	2.781	3.146	0.180

Table 1. Segmentation performance for 15 sampled clinically important organs. Mean organ volume, number of evaluable test cases, Dice similarity coefficient (DSC), and 95th-percentile Hausdorff distance (HD95) are reported for the standard and optimized nnU-Net v2 workflows. Organ-level comparisons were performed using paired Wilcoxon signed-rank tests. Organs are sorted by average volume in mm³.

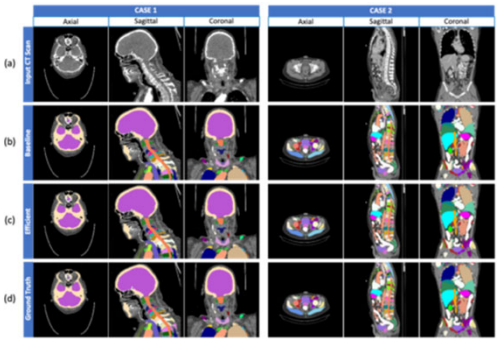

Figure 1. Multi-organ segmentation results from two test CT scans. Rows show (a) input CT images, (b) standard nnU-Net v2 predictions, (c) optimized-workflow predictions, and (d) reference contours. Axial, sagittal, and coronal views are shown for each case. Corresponding colors represent the same anatomical structures across prediction and reference images.

DISCUSSION

These results suggest that substantial nnU-Net v2 training acceleration can be achieved through workflow-level optimization without reducing input resolution or changing the network architecture. This may allow faster model iteration for preprocessing strategies, dataset updates, contouring protocols, and training configurations within the same computational budget. Comparable DSC and HD95 indicate that both volumetric overlap and boundary accuracy were maintained.
One limitation is that validation was performed on one large CT dataset and one GPU architecture. Additional testing across other datasets, disease sites, imaging modalities, and hardware platforms is needed to evaluate generalizability.

CONCLUSIONS

Targeted optimization of the nnU-Net v2 training workflow substantially reduced training time while preserving segmentation performance in a large-scale clinical CT dataset. This approach may lower computational barriers for large-scale auto-segmentation research and enable faster model iteration in radiation oncology.

REFERENCES

[1]. Shi, Feng, et al. "Deep learning empowered volume delineation of whole-body organs-at-risk for accelerated radiotherapy." Nature communications 13.1 (2022): 6566.

[2]. Isensee, Fabian, et al. "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation." Nature methods 18.2 (2021): 203-211.