

Mini-Expert: Bottleneck Augmentation Reduces Boundary Outliers for Robust and Interpretable Multi-Organ Auto-Segmentation

Yunkui Pang¹; Xu Chen, PhD²; Pew-Thian Yap, PhD¹; Jun Lian, PhD¹

¹ University of North Carolina at chapel Hill, ² Huaqiao University

ABSTRACT

Purpose: Accurate multi-organ auto-segmentation supports efficient clinical workflows, but nnU-Net-based models such as TotalSegmentator can produce residual boundary outliers requiring manual correction, especially for small structures. We propose a Mini-Expert bottleneck augmentation to improve robustness and provide routing-derived interpretability for clinical validation.

Methods: We integrated a lightweight Mini-Expert module at the TotalSegmentator bottleneck without altering the overall topology. Clinically disruptive errors often appear as boundary outliers (large HD95) or isolated false positives. The module acts as a learned “expert consultation” step, adaptively allocating extra capacity to uncertain or high-risk regions to correct localized mistakes before they become large surface deviations. It also provides attention maps for qualitative interpretability. We trained baseline and Mini-Expert models on 1,539 CT images and evaluated 117 anatomical structures across 6,214 organ instances using Dice similarity coefficient (DSC) and 95th percentile Hausdorff distance (HD95). A focused subset of 24 clinically important organs was analyzed using paired Wilcoxon signed-rank tests.

Results: Mini-Expert preserved high overlap while improving robustness, increasing mean DSC from 0.9128 to 0.9151 (p=0.002) and reducing mean HD95 from 5.94 to 4.48 mm (p=0.0001). The primary effect was suppression of extreme HD95 outliers, consistent with fewer clinically disruptive boundary failures. In the 24-organ clinical subset, Mini-Expert significantly reduced HD95 for the heart, upper lung lobe, urinary bladder, femurs, duodenum, and lumbar vertebra.

Conclusion: Mini-Expert bottleneck routing improves segmentation robustness by reducing HD95 outliers while preserving Dice. This plug-in design improves contour stability for clinically important organs and may reduce manual contour editing time while increasing clinical confidence in auto-segmentation.

CONTACT

Jun Lian, PhD, DABR

University of North Carolina at Chapel Hill

✉ jun_lian@med.unc.edu

☎ (984) 974-0400

INTRODUCTION

- Reliable **multi-organ auto-segmentation** is essential for clinical workflows.
- Mini-Expert reduces **boundary outliers** and **false positives**, making contours more reliable and easier to edit.

CONCLUSIONS

- Mini-Expert maintained high DSC (0.9128 → 0.9151) while **reducing HD95 from 5.94 to 4.48 mm**, indicating improved boundary robustness.
- Mini-Expert produces **fewer extreme HD95 outliers**, suggesting fewer clinically disruptive boundary failures, especially in smaller organs.
- Mini-Expert improved HD95 for key structures including the **heart, upper lung lobe, urinary bladder, femurs, duodenum, and lumbar vertebra**.

ACKNOWLEDGEMENTS

- This study is in part supported by NIH R01CA206100, R01EB035160 and UNC Lineberger Developmental Award 29242

METHOD

- Mini-Expert adds a lightweight refinement step at the U-Net bottleneck, improving robustness without changing the encoder-decoder design.
- With sparse routing, the model gives extra attention to high-risk regions such as small organs, thin boundaries, and areas prone to HD95 outliers.
- The routing weights can be projected back to the image as attention heatmaps, showing where the model focused during segmentation.
- Trained baseline nnU-Net V2 and Mini-Expert on 1,539 CT images, then evaluated 117 structures across 6,214 organ instances from 89 CT scans using DSC and HD95.

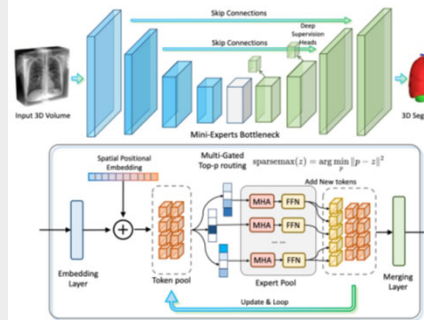


Figure 1. Overview of the nnU-Net v2 TotalSegmentator backbone augmented with a bottleneck Mini-Expert module. Bottleneck features are tokenized and routed to multiple transformer experts in the “Expert pool”. Tokens output by experts are pooled, pruned to a fixed token budget, and fused back via cross-attention.

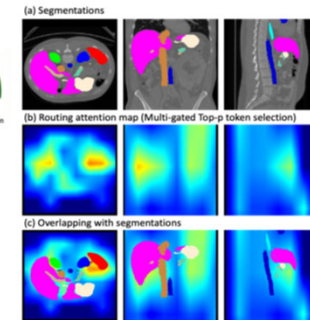


Figure 2. (a) Axial, coronal, and sagittal views of segmentations from Mini-Expert. (b) Routing-derived attention maps at Mini-Expert. (c) Overlay of segmentation and routing attention. The mechanism concentrates computation on anatomically relevant regions for enhanced interpretability.

RESULTS

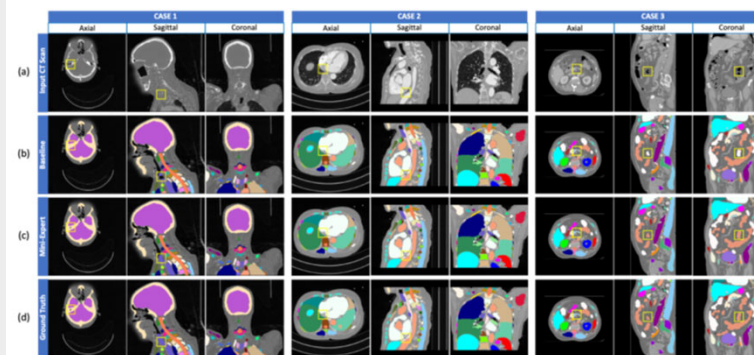


Figure 3. Qualitative results. Regions of interest are highlighted in yellow boxes. Rows: (a) Input CT scan. (b) Baseline (nnU-Net) segmentation results. (c) Mini-Expert segmentation results. (d) Ground Truth.

Organs	Avg Volume (mm ³)	Average DSC		p-value	Average HD95 (mm)	Average DSC		p-value
		Baseline	Min Expert			Baseline	Min Expert	
adrenal gland left	4133.3	0.878	0.876	0.77	2.725	2.458	0.07	
adrenal gland right	3568.8	0.871	0.871	0.19	2.238	2.158	0.52	
aorta	169780.3	0.975	0.978	0.06	1.876	1.708	0.48	
brain	790446.9	0.925	0.926	0.45	2.781	2.741	0.37	
duodenum	47704.3	0.826	0.829	0.92	11.234	5.741	0.03	
femur left	191916.8	0.959	0.959	0.57	12.368	4.521	0.04	
femur right	187771.8	0.957	0.959	0.26	7.952	2.623	0.01	
gallbladder	23244.5	0.809	0.818	0.07	4.255	5.171	0.73	
heart	463566.9	0.9	0.9	0.95	7.393	2.924	0.01	
kidney left	129936.3	0.944	0.938	0.12	3.927	4.023	0.49	
kidney right	133045.2	0.926	0.927	0.37	3.886	3.804	0.47	
liver	1397957.9	0.959	0.962	0.47	9.825	10.037	0.34	
lung lower lobe left	636082	0.975	0.975	0.16	2.66	2.468	0.32	
lung lower lobe right	730855.6	0.974	0.969	0.88	1.938	1.598	0.07	
lung middle lobe right	312171.9	0.951	0.951	0.43	4.019	3.563	0.35	
lung upper lobe left	608003.3	0.967	0.967	0.96	3.412	3.332	0.27	
lung upper lobe right	628974.1	0.873	0.879	0.32	21.754	10.499	0.03	
pancreas	59248.3	0.872	0.885	0.15	3.951	3.136	0.09	
prostate	30825.8	0.794	0.796	0.31	4.857	4.59	0.42	
spinal cord	54536.3	0.947	0.947	0.81	1.507	1.504	0.63	
spleen	387603.2	0.981	0.979	0.27	6.485	5.723	0.12	
stomach	283063.7	0.936	0.938	0.12	5.872	6.796	0.94	
urinary bladder	222992.8	0.91	0.909	0.45	16.327	7.452	0.01	
vertebrae L3	60785.1	0.948	0.963	0.03	5.648	1.253	0.02	

Table 1. Quantitative results of 24 sampled clinically important organs. Average organ volume and segmentation accuracy are reported for the TotalSegmentator baseline and the proposed Mini-Expert model using Dice similarity coefficient (DSC) and 95th percentile Hausdorff distance (HD95, mm). Paired statistical testing is provided (Wilcoxon signed rank).

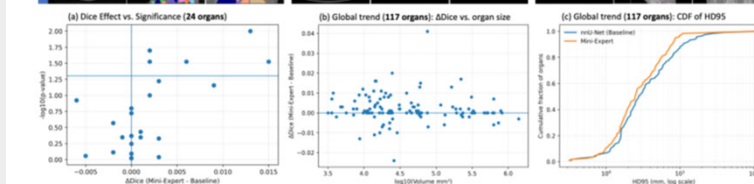


Figure 4. (a) Volcano plot for 24 clinically important organs showing Δ Dice versus $-\log_{10}(p\text{-value})$. (b) Across 117 organs, Δ Dice is plotted against $\log(\text{organ volume})$, indicating larger gains for smaller structures. (c) HD95 CDF (117 organs, log scale) demonstrates a left-shifted distribution for Mini-Expert, consistent with fewer boundary outliers and improved robustness.