Memory efficient localized segmentation by object detection with multi-channel input

Chan Woong Lee [1, 2], Sang Joon Park [2], Jaehee Chun[3]

[1] Medical Physics and Biomedical Engineering Lab(MPBEL), Yonsei University College of Medicine, Seoul, South Korea

[2] Department of Radiation Oncology, Yonsei Cancer Center, Heavy Ion Therapy Research Institute, Yonsei University College of Medicine, Seoul, South Korea

[3] Oncosoft Inc, Seoul, South Korea

Introduction

- Accurate medical image segmentation is critical in optimizing and evaluating radiation treatment plans in clinical settings
- Traditionally, this process has been carried out manually, a practice that is subject to intra- and inter-observer variability, demanding in terms of workload, and time-consuming.
- Despite the emergence of AI-based auto-segmentation solutions leveraging advancements in computer vision and artificial intelligence technologies, many existing methods process all areas of an image even when the Organ at Risk (OAR) or Gross Tumor Volume (GTV) occupies a small fraction of the total image space.
- Our method leverages an object detection network to first localize areas of interest in entire images before conducting segmentation, focusing only on identified pivotal regions, consequently reducing the necessary computational resources and segmentation time.

Methods

Overall workflow

- The proposed method is delineated in two successive stages. (Figure 1)
- Stage 1 (Localization): Initially, a bounding box is identified to encompass the OAR utilizing an object detection network.
- Stage 2 (Segmentation): Following the localization stage, the cropped CT image undergoes segmentation through a dedicated Segmentation network.

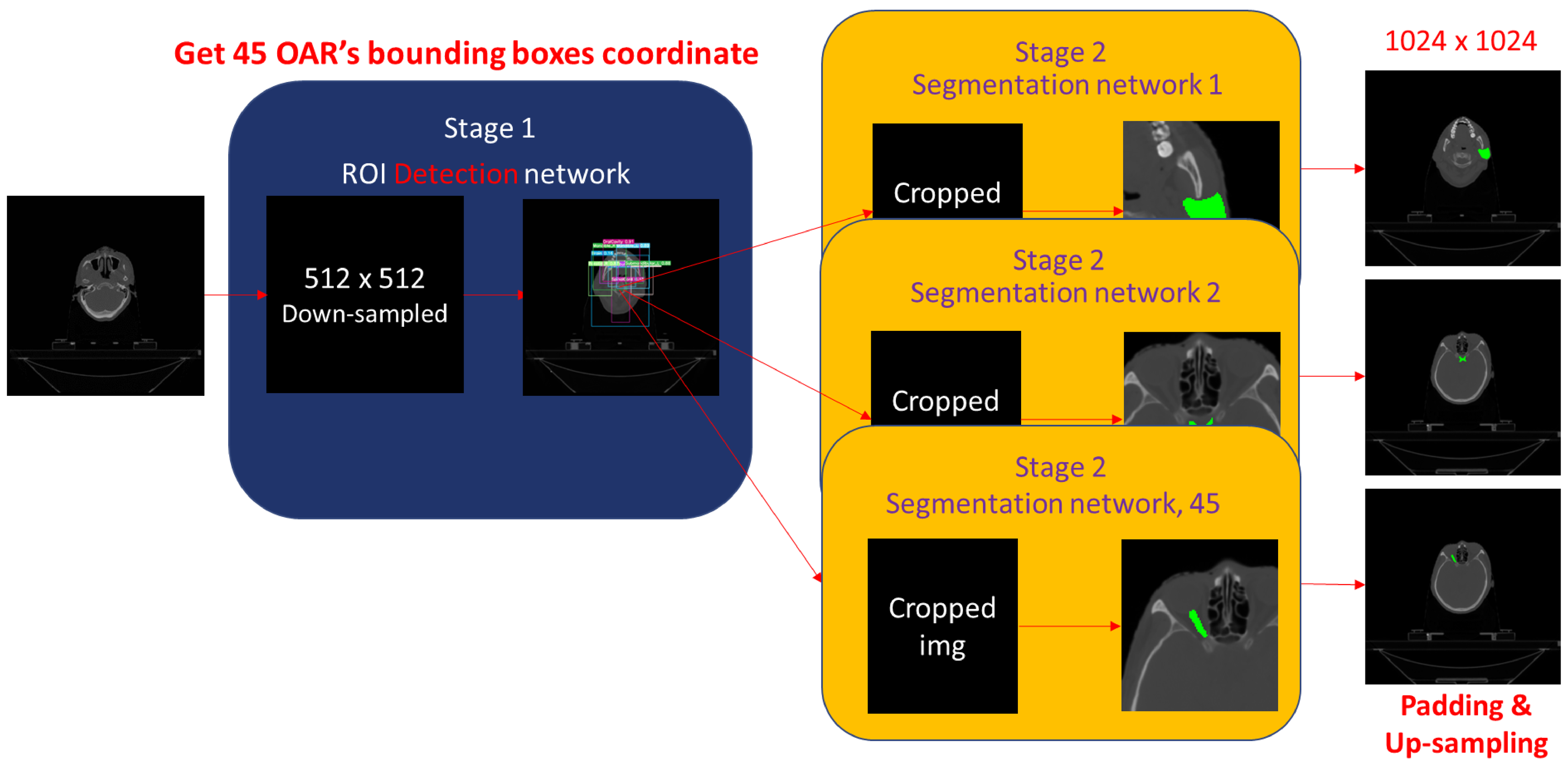


Figure 1. The structure of the proposed method, consisting of the region of interest (ROI) localizing network (blue) and the segmentation network (yellow).

Network

- Localizing stage leverages a 2D-based object detection network powered by the YOLO-v7 model[1].
- In segmentation stage, Dyn-U-Net as implemented in MONAI library was chosen.
- It inherits certain rule-based hyper-parameter settings from nn-U-Net, allowing for an adaptable approach to determining kernel size and stride based on the patch size[2].

Data preprocessing – multi-channel input

- we adjusted the window widths and levels to generate a set of windowed CT images, each emphasizing different anatomical structures.
- These images were then combined to create a 6-channel input, enhancing the network's ability to distinguish detailed features during the training phase.

Ch	Windowing	[Min, Max]	Normalization
1	Total HU (NCECT)	[-1024, 3071]	None
2	Bone	[-1000, 2000]	[0, 1]
3	Soft tissue	[-160, 350]	[0, 1]
4	Brain	[-5, 65]	[0, 1]
5	Stroke	[15, 45]	[0, 1]
6	Total HU (CECT)	[-1024, 3071]	None

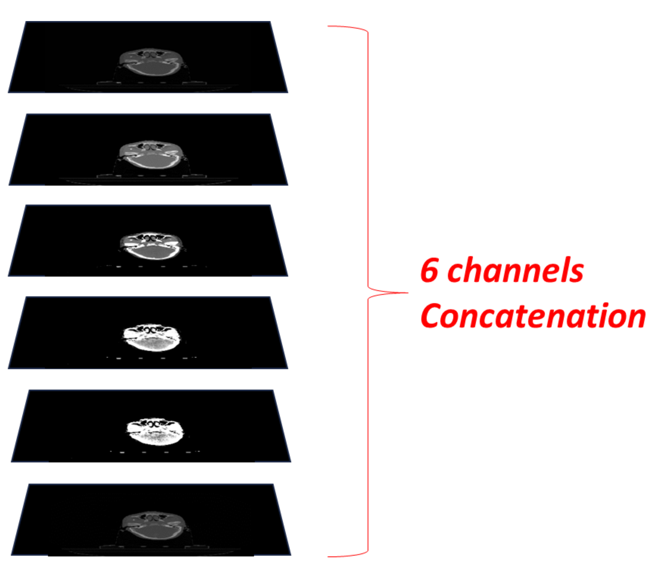
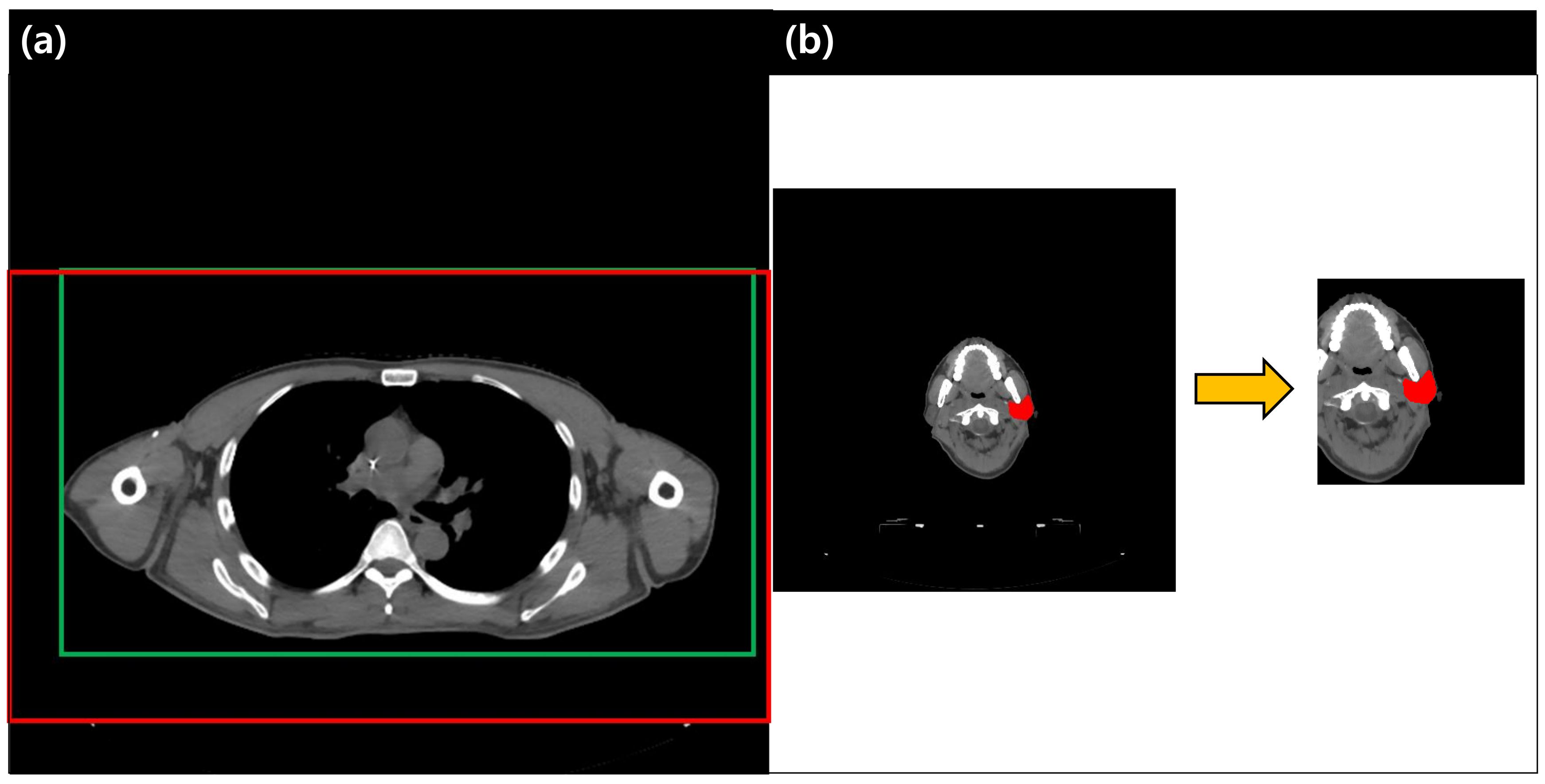


Figure 2. Window values applied to multi-channel input and normalization table for each value.

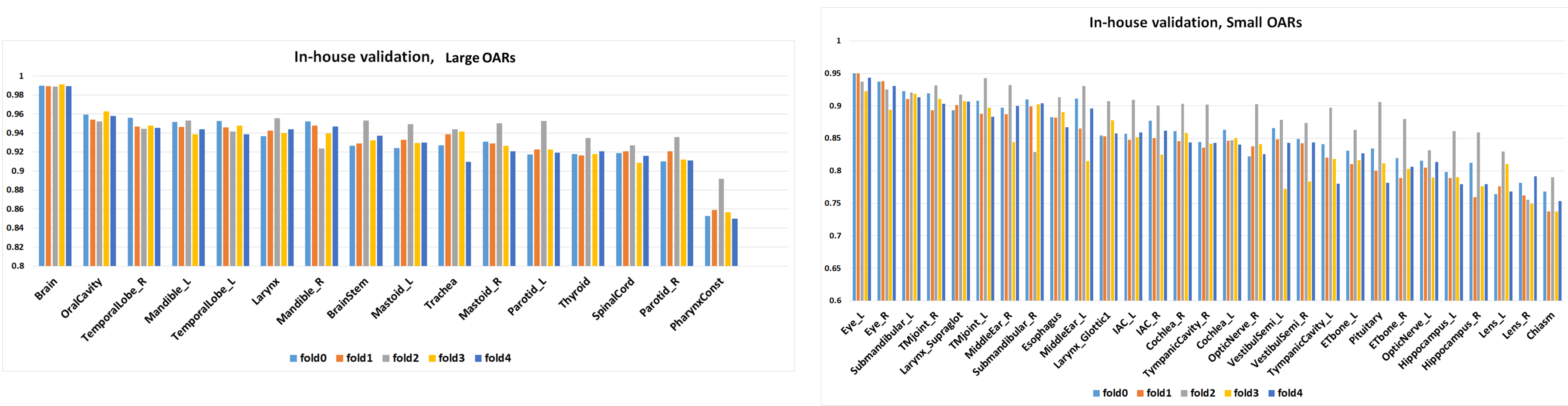
Data preprocessing – cropping method

- 1) In the connected-components body-finding algorithm, we identify the largest connected component within the CT image. This allows for a tighter cropping of the body part compared to standard foreground cropping.
- 2) The localized cropping, feasible due to the presence of a localizer(Object detection network), involved cropping & training based on the Organ at Risk (OAR).



Results

In-house validation result (DSC)



	Fold 0	Fold 1	Fold 2	Fold 3	Fold 4	Average
DSC	0.885±0.056	0.876±0.062	0.906±0.047	0.872±0.063	0.876±0.060	0.883±0.058

Table 1. Dice similarity coefficients(DSC) of each folds. (train: 96/ valid: 24)

Actual validation result, 4th (DSC & Normalized surface distance (NSD))

Rank	DSC	NSD
1	0.8925	0.9110
2	0.8876	0.9052
3	0.8881	0.9046
4 (proposed)	**0.8845**	**0.8985**
5	0.8795	0.8838
6	0.8674	0.8676
7 (baseline)	0.8674	0.8676
8	0.8674	0.8676
9	0.8188	0.8239
10	0.7984	0.8024
11	0.7895	0.7851
12	0.7547	0.7199

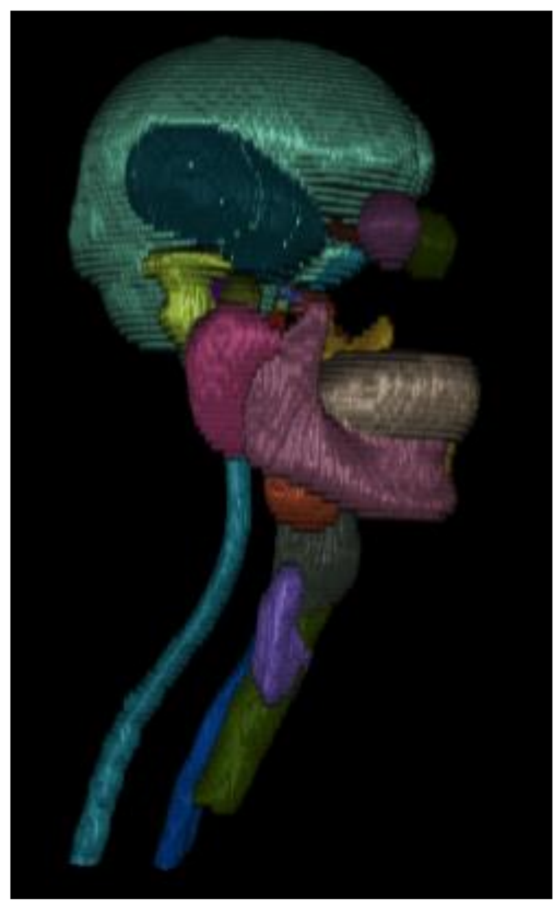


Figure 3. Visualization of inference results from one of the actual validation datasets.

Discussion

Limitation

- Training with 512 x 512 images may lead to aliasing, potentially resulting in a slight loss of DSC.

Conclusion

- There aren't many cases of using object detection in medical image segmentation.
- It is anticipated that efficient training/inference will be possible even on high-capacity/high-resolution datasets

REFERENCES

[1] Wang, Y., Wang, H., Xin, Z. (2022). Efficient detection model of steel strip surface defects based on YOLO-V7. IEEE Access, 10, 133936-133944.

[2] Isensee, F., Jaeger, P. F., Kohl, S. A., Petersen, J., Maier-Hein, K. H. (2021). nnUNet: a self-configuring method for deep learning-based biomedical image segmentation. Nature methods, 18(2), 203-211.