

Supplementary Material: Organ at Risk Segmentation for Head and Neck Cancer using Stratified Learning and Neural Architecture Search

1. Performance of OAR segmentation

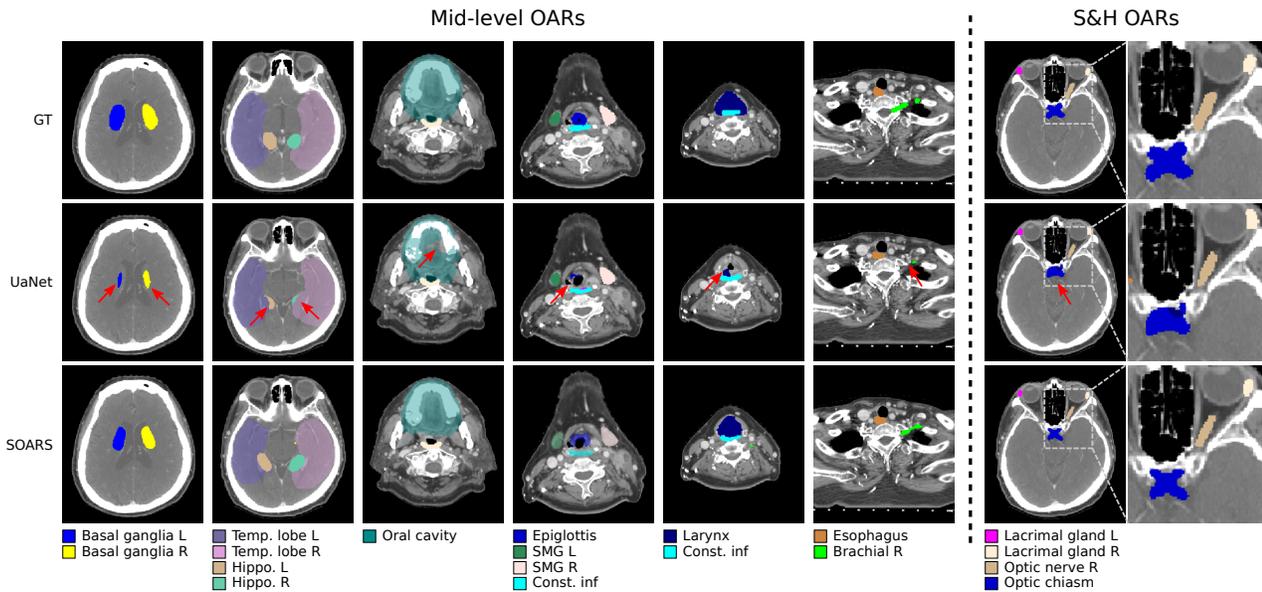


Figure 1. Qualitative illustration of the mid-level (left-hand side) and S&H (right-hand side) OAR segmentation using UaNet and the proposed SOARS. The seven columns are seven representative axial slices in the RTCT image. The 1st column shows the OAR labels from a radiation oncologist, while the 2nd and 3rd columns are the predicted segmentation results by the UaNet and the proposed SOARS, respectively. For better comparison, we use red arrows to indicate the improvements. For visualization purpose, the dashed rectangles are enlarged for highlighting improvements on S&H OAR segmentation.

In Tab. 1, we report the category-by-category Dice score (DSC) of the proposed SOARS against UNet [2], P-HNN [1], and UaNet [3]. In Tab. 2, we report the category-by-category Hausdorff distance (HD) of the proposed SOARS against UNet, P-HNN, and UaNet. For both metrics, SOARS achieved 30 out of 42 OARs best performance. SOARS performed slightly worse than UaNet on temporal lobe and temporomandibular joint segmentations in terms of DSC. Yet, the DSC differences are relatively small. We demonstrate some qualitative comparison results against UaNet in Fig. 1, where the improvements are indicated using red arrows.

Organ	UNet	P-HNN	UaNet	SOARS
Basal Ganglia Lt	64.0±12.4	63.5±16.6	63.6±13.7	63.8±13.7
Basal Ganglia Rt	64.7±13.9	63.5±14.2	67.4±15.0	63.6±11.6
Brachial Lt	59.8±13.7	48.8±11.8	49.9±10.3	66.8±17.1
Brachial Rt	58.8±13.7	49.4±7.0	53.5±8.0	65.5±14.2
Brainstem	81.7±5.4	80.1±6.8	80.6±6.3	81.0±5.7
Cerebellum	83.2±2.7	88.8±2.8	90.1±2.8	90.2±2.3
Cochlea Lt	64.0±17.6	67.2±10.4	66.5±12.6	72.3±12.2
Cochlea Rt	64.2±10.0	67.2±10.4	68.2±12.6	69.5±12.4
Const. inf	63.4±17.1	61.8±14.9	73.6±10.6	65.0±18.3
Const. mid	64.9±15.4	63.1±14.5	66.1±11.3	66.9±15.1
Const. sup	64.0±10.2	64.1±10.0	62.3±11.3	67.4±9.2
Epiglottis	65.5±8.6	65.5±11.0	65.4±13.1	67.3±8.2
Esophagus	66.3±23.2	61.6±12.0	69.1±12.9	67.0±14.0
Eye Lt	83.4±7.4	86.4±3.4	85.7±7.4	86.4±3.3
Eye Rt	82.7±6.3	85.9±3.3	86.7±4.3	86.6±4.0
Hippocampus Lt	62.4±12.5	46.2±17.3	50.0±17.3	67.4±16.0
Hippocampus Rt	62.2±14.3	45.2±12.1	52.2±17.6	67.9±18.9
Hypothalamus	63.6±17.3	39.2±16.8	28.7±22.9	72.6±17.1
Innerear Lt	62.4±12.1	58.4±10.6	68.8±10.9	78.8±8.1
Innerear Rt	63.2±16.8	60.1±10.3	73.0±12.2	76.9±9.1
Lacrimalgland Lt	59.2±10.5	54.7±11.5	64.1±16.0	70.7±8.0
Lacrimalgland Rt	58.7±10.5	54.7±11.5	52.1±14.3	70.6±11.0
Larynx core	57.9±17.1	53.9±17.1	56.9±20.1	69.7±20.8
Mandible Lt	87.4±2.9	90.2±2.0	88.2±12.1	91.7±1.8
Mandible Rt	89.1±2.3	90.8±1.8	88.0±6.0	91.1±2.5
Optic Chiasm	49.9±15.4	50.9±13.6	60.4±22.1	72.9±9.2
Optic Nerve Lt	61.7±11.1	67.6±11.0	69.9±9.3	74.3±7.8
Optic Nerve Rt	62.0±12.2	67.6±10.2	69.9±11.0	72.3±8.7
Oralcavity	64.0±5.1	76.3±5.1	77.8±10.2	82.6±5.3
Parotid Lt	64.7±5.8	78.2±5.1	82.8±6.2	84.5±4.2
Parotid Rt	64.7±6.1	78.8±6.5	82.3±6.6	84.1±5.0
Pineal Gland	46.4±29.3	60.2±16.5	63.6±26.4	70.4±14.7
Pituitary	60.4±11.0	65.2±11.0	57.0±14.8	61.5±18.4
Spinalcord	83.5±6.2	83.7±3.6	82.7±7.4	84.6±2.4
SMG Lt	64.2±16.8	71.3±8.8	77.3±9.1	76.9±9.8
SMG Rt	63.2±16.8	69.5±11.7	75.2±9.4	76.1±9.0
Temporal Lobe Lt	66.7±3.6	80.9±3.7	82.6±6.4	81.0±5.2
Temporal Lobe Rt	65.1±5.1	73.6±17.4	82.4±5.7	80.5±4.0
Thyroid Lt	64.9±18.9	76.7±7.7	81.2±6.1	81.6±5.0
Thyroid Rt	64.4±17.7	77.0±6.0	80.5±10.5	82.2±5.1
TMjoint Lt	79.2±6.5	77.2±6.5	79.3±12.8	77.6±7.0
TMjoint Rt	76.5±8.8	75.2±9.3	77.4±9.6	76.2±7.1
Average	66.6	67.6	70.4	75.1

Table 1. Dice score comparison on the H&N 42 OAR dataset (unit: %): Lt is short for left and Rt is short for right. Const. is short for constrictor muscle, SMG is short for submandibular gland, and TMjoint is short for temporomandibular joint. The proposed SOARS achieved the best performance in 30 (in bold) out of 42 OARs.

Organ	UNet	P-HNN	UaNet	SOARS
Basal Ganglia Lt	10.0±2.8	9.8±3.2	10.5±4.0	9.3±3.2
Basal Ganglia Rt	9.3±3.8	10.2±3.3	10.5±3.8	11.1±3.4
Brachial Lt	14.9±6.2	15.1±9.6	14.2±11.7	17.3±10.9
Brachial Rt	17.9±8.2	11.4±5.0	16.2±9.6	14.0±7.3
Brainstem	8.4±2.9	8.8±2.9	10.3±3.8	8.1±2.2
Cerebellum	8.9±3.8	9.4±4.7	14.1±9.8	7.7±3.1
Cochlea Lt	3.6±9.0	1.8±0.5	2.3±0.8	1.6±0.4
Cochlea Rt	2.1±0.8	2.0±1.0	2.4±0.9	1.9±0.6
Const. inf	5.7±2.6	8.5±3.9	7.5±4.9	5.4±2.4
Const. mid	7.4±2.8	8.7±3.1	14.7±10.1	7.4±3.3
Const. sup	7.4±3.0	8.0±3.6	12.7±8.2	7.0±3.6
Epiglottis	6.7±2.3	6.9±3.6	9.9±8.5	6.9±2.5
Esophagus	25.1±26.4	21.9±13.7	24.0±15.0	21.1±15.8
Eye Lt	2.8±0.8	3.0±1.8	4.0±5.4	3.3±1.1
Eye Rt	3.1±0.9	3.4±0.9	3.1±0.7	3.0±1.0
Hippocampus Lt	11.0±6.7	16.9±8.6	15.9±8.9	12.2±7.7
Hippocampus Rt	10.7±6.1	12.7±5.8	13.3±6.6	12.5±8.2
Hypothalamus	16.9±8.6	9.3±4.3	10.3±3.7	2.5±1.3
Innerear Lt	12.7±5.8	11.9±33.7	4.0±1.4	2.6±0.7
Innerear Rt	9.3±4.3	4.1±1.3	4.7±2.8	2.9±0.8
Lacrimal Gland Lt	4.3±1.0	4.3±1.3	4.6±1.6	2.9±1.1
Lacrimal Gland Rt	4.1±1.2	5.5±1.5	5.1±2.2	2.9±0.9
Larynx core	12.4±7.3	10.4±7.3	9.2±7.2	9.0±7.1
Mandible Lt	7.9±2.9	6.7±2.8	10.3±24.4	5.3±2.3
Mandible Rt	7.0±2.6	5.6±2.3	12.2±15.8	5.5±1.6
Optic Chiasm	8.0±3.9	8.4±5.3	11.4±7.8	5.3±4.2
Optic Nerve Lt	4.2±3.6	4.6±3.5	5.2±3.1	3.4±1.9
Optic Nerve Rt	4.1±2.3	3.9±1.7	4.9±4.2	3.3±1.4
Oralcavity	16.4±5.0	18.4±5.0	7.6±10.3	13.8±6.2
Parotid Lt	9.0±3.4	10.0±2.8	8.0±5.8	7.0±2.5
Parotid Rt	8.9±7.8	8.3±2.0	9.7±4.2	6.8±1.6
Pineal Gland	3.4±1.8	2.5±1.1	4.0±1.9	1.7±0.6
Pituitary	3.9±1.4	4.4±1.6	4.4±1.3	4.2±2.2
Spinalcord	34.9±13.9	10.2±18.1	17.3±27.2	5.7±2.2
SMG Lt	7.3±4.0	18.6±30.3	6.1±5.4	6.5±3.1
SMG Rt	7.3±4.0	11.1±8.3	7.0±4.9	6.1±2.3
Temporal Lobe Lt	14.3±21.4	16.0±6.8	16.5±6.7	14.6±6.9
Temporal Lobe Rt	12.8±3.6	38.6±85.2	15.0±5.0	13.5±5.9
Thyroid Lt	9.0±2.9	6.9±3.2	7.4±4.8	5.1±2.5
Thyroid Rt	8.7±10.4	7.9±3.3	7.1±4.0	5.5±2.3
TMjoint Lt	3.5±1.2	3.9±1.4	4.4±2.4	3.6±1.7
TMjoint Rt	3.6±1.7	4.6±1.1	4.3±2.9	3.5±1.3
Anchor OARs	9.3	9.4	9.2	7.0

Table 2. Average Hausdorff distance comparison on the H&N 42 OAR dataset (unit: mm): Lt is short for left and Rt is short for right. Const. is short for constrictor muscle, SMG is short for submandibular gland, and TMjoint is short for temporomandibular joint. The proposed SOARS achieved the best performance in 30 (in bold) out of 42 OARs.

2. Performance of S&H OAR detection

In Tab. 3, we report the category-by-category detection accuracy of the regressed center points using the detection-by-segmentation network. Moreover, we binarise both the regressed and ground-truth heat maps by keeping the top 1000 largest intensity voxels, and report their HD. Note, as cochlea is spatially enclosed by inner-ear, we use a single heat map, *i.e.* ear, for both OARs detection. As shown in Tab. 3, we achieve an average HD reduction of 13.7 mm (from 18.9 mm to 6.2 mm) as compared to the detection using only RTCT images. The HD for all OARs are reduced, especially the lacrimal gland, optic chiasm, and pineal gland. These significant HD reductions indicate that the anchor OARs serve as effective references to better detect the S&H OAR locations.

	Dist (mm)		HD (mm)	
	CT Only	CT+Anchor	CT Only	CT+Anchor
Ear Lt	3.9±2.5	3.9±2.6	6.7±3.3	5.7±2.1
Ear Rt	1.9±1.4	1.6±1.0	4.4±1.8	3.4±1.3
Hypothalamus	2.6±1.7	2.3±1.5	4.0±2.0	3.6±1.5
Lacrimal Gland Lt	5.6±5.7	4.6±3.1	28.0±76.8	14.7±20.7
Lacrimal Gland Rt	3.3±1.9	3.0±1.7	47.4±112.0	4.7±1.4
Optic Chiasm	3.9±2.5	3.4±1.9	26.6±71.8	10.6±25.6
Optic Nerve Lt	2.5±1.6	2.6±1.5	4.6±1.8	4.5±1.2
Optic Nerve Rt	3.0±1.2	3.1±1.6	21.9±61.0	4.9±1.6
Pineal Gland	2.5±2.5	1.8±0.7	27.7±72.2	3.9±1.3
Average	3.3	2.9	18.9	6.2

Table 3. The detailed S&H detection results measuring the average distances between regressed and true center points, as well as the Hausdorff distances between the binarised regressed and binarised true heat maps. Lt is short for left and Rt is short for right. The best performance is highlighted in bold.

References

- [1] A P Harrison, Z Xu, K George, and et al. Progressive and multi-path holistically nested neural networks for pathological lung segmentation from ct images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 621–629. Springer, 2017. 1
- [2] O Ronneberger, P Fischer, and T Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241. Springer, 2015. 1
- [3] H Tang, X Chen, Y Liu, Z Lu, J You, M Yang, S Yao, G Zhao, Y Xu, T Chen, et al. Clinically applicable deep learning framework for organs at risk delineation in ct images. *Nature Machine Intelligence*, pages 1–12, 2019. 1