Teaching AI the Anatomy Behind the Scan: Addressing Anatomical Flaws in Medical Image Segmentation with Learnable Prior Supplementary Material

Young Seok Jeon[1,*], Hongfei Yang[1,*], Huazhu Fu[2], Mengling Feng[1,†]
[1]National University of Singapore [2]Agency for Science, Technology and Research (A*STAR), Singapore
youngseokjeon74@gmail.com, {hfyang, mornin}@nus.edu.sg, hzfu@ieee.org

1. Thin-plate-spline Deformation

In this section, we provide a detailed explanation of thin-plate splines (TPS) [1], including its unraveled non-matrix definition and the optimization of TPS coefficients by solving a linear equation subject to several constraints. As per convention in computer vision, we call the three coordinate axes in $\mathbb{R}^3$ the h, w and d-axis.

Given an object P in $\mathbb{R}^3$, we wish to alter its shape by warping the coordinate axes. This can be done by constructing a warping function $\mathcal{D} : \mathbb{R}^3 \to \mathbb{R}^3$, and reconstruct the new shape Y by

$$(h', w', d') = \mathcal{D}(h, w, d), \quad Y(h, w, d) := P(h', w', d').$$

That is, the value of the *target* object Y at the coordinate point (h, w, d), is given by the value of the *source* object P at the warped coordinate point (h', w', d') which is calculated by the warping function $\mathcal{D}$. We call points associated with the target object Y *target points*, and points associated with the source object P *source points*.

The thin-plate-spline deformation is a method to construct the warping function $\mathcal{D}$. Given a sequence of *target control points* $\{\mathbf{p}_i\}_i^N$ and a corresponding *source control points* $\{\mathbf{p}'_i\}_i^N$, the warping $\mathcal{D}$ maps exactly $\mathbf{p}_i \mapsto \mathbf{p}'_i$ with minimal bending energy. Given a general target point $\mathbf{p} = (h, w, d)$, its image under $\mathcal{D}$ is given by

$$\begin{aligned}\mathcal{D}_h(\mathbf{p}) &= a^{(N+1)} + a^{(N+2)}h + a^{(N+3)}w + a^{(N+4)}d \\ &\quad + \sum_{i=1}^{N} a^{(i)} U(|\mathbf{p} - \mathbf{p}_i|), \qquad (1a)\\ \mathcal{D}_w(\mathbf{p}) &= b^{(N+1)} + b^{(N+2)}h + b^{(N+3)}w + b^{(N+4)}d \\ &\quad + \sum_{i=1}^{N} b^{(i)} U(|\mathbf{p} - \mathbf{p}_i|), \qquad (1b)\\ \mathcal{D}_d(\mathbf{p}) &= c^{(N+1)} + c^{(N+2)}h + c^{(N+3)}w + c^{(N+4)}d \\ &\quad + \sum_{i=1}^{N} c^{(i)} U(|\mathbf{p} - \mathbf{p}_i|), \qquad (1c)\end{aligned}$$

where $U(r) = r^2 \log r^2$ is the kernel function, $(a^{(1)}, \cdots, a^{(N+4)})$, $(b^{(1)}, \cdots, b^{(N+4)})$, and $(c^{(1)}, \cdots, c^{(N+4)})$ are TPS coefficients that are determined by mapping the control points. The TPS coefficients can be obtained by solving a linear equation.

Here, we use the h-coordinate coefficients as an example, and the calculation of the w and d coordinate coefficients are done in a similar manner. The function (1a) has $N + 4$ coefficients to be computed, which can be calculated by a closed-form solution.

Let $\mathbf{v} = (h'_1, \cdots, h'_N | 0, 0, 0, 0)^T$, where h'_i is the h-coordinate of the i-th source control point. Also, define

*Equal contribution
†Corresponding author

matrices

$$\mathcal{K} = \begin{bmatrix} 0 & U_{12} & \cdots & U_{1N} \\ U_{21} & 0 & \cdots & U_{2N} \\ \cdots & \cdots & \cdots & \cdots \\ U_{N1} & U_{N2} & \cdots & 0 \end{bmatrix}, N \times N;$$

$$\mathcal{P} = \begin{bmatrix} 1 & h_1 & w_1 & d_1 \\ 1 & h_2 & w_2 & d_2 \\ \cdots & \cdots & \cdots & \cdots \\ 1 & h_N & w_N & d_N \end{bmatrix}, N \times 4;$$

$$\mathcal{M} = \begin{bmatrix} \mathcal{K} & \mathcal{P} \\ \mathcal{P}^T & O \end{bmatrix}, (N+4) \times (N+4) \tag{2a}$$

where $U_{i,j} = U(|\mathbf{p}_i - \mathbf{p}_j|)$, h_i, w_i, and d_i are the h-, w-, and d-coordinates of the target control point $\mathbf{p}_i$, and O is a zero matrix of size 4×4. Then the coefficients $\mathbf{a} = (a^{(1)}, \cdots, a^{(N+4)})$ are given by

$$\mathbf{a} = \mathcal{M}^{-1}\mathbf{v}. \tag{3}$$

The additional last four rows of $\mathcal{M}$ guarantee that the coefficients $a^{(i)}$ sum to zero and that their cross-products with the points $\mathbf{p}_i$ are likewise zero. These extra conditions are regularization terms used in TPS formulation.

In our implementation, we keep the target control points fixed, and use neural networks to propose the source control points. By doing so, we only need to calculate $\mathcal{M}^{-1}$ once, and we do not have numerical instability problem.

2. Extended Results

We present the class-wise performance of AIC-Net models with the UNETR-Swin backbone on all tasks in Tabs. 1 to 3. Notably, AIC-Net not only achieves better mean scores but also generally exhibits lower standard deviations, indicating more consistent and reliable performance. We have omitted the segmentation results for kidney_cyst_left and kidney_cyst_right from the Organ task, as both AIC-Net and the baseline models failed to predict these classes.

References

[1] Fred L. Bookstein. Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Transactions on pattern analysis and machine intelligence*, 11(6):567–585, 1989. 1

Table 1. Organ Segmentation Comparison on UNETR-Swin Backbone

	NSD ↑		**Haus$_{95}$** ↓		**Dice** ↑	
	AIC-Net	Baseline	AIC-Net	Baseline	AIC-Net	Baseline
adrenal_gland_left	95.0 ± 13.9	96.2 ± 10.2	2.22 ± 2.34	5.29 ± 26.0	82.2 ± 15.2	83.4 ± 15.3
adrenal_gland_right	96.9 ± 9.41	96.8 ± 9.50	1.65 ± 1.47	1.55 ± 1.46	83.2 ± 14.1	83.7 ± 15.2
colon	87.1 ± 15.5	82.0 ± 25.1	10.6 ± 13.7	12.6± 17.6	86.1 ± 11.5	85.1 ± 15.6
duodenum	86.9 ± 16.1	85.0 ± 22.6	5.56 ± 6.47	8.02± 17.4	78.2 ± 18.9	78.6 ± 21.0
esophagus	96.1 ± 13.1	95.7 ± 13.3	2.82 ± 6.06	2.69 ± 6.04	89.5 ± 5.70	89.0 ± 6.59
gallbladder	74.2 ± 38.0	78.1 ± 35.0	7.12 ± 16.8	6.12 ± 17.5	80.2 ± 25.3	81.1 ± 24.0
kidney_left	94.0 ± 14.8	91.5 ± 21.9	5.05 ± 13.2	4.68 ± 13.2	91.2 ± 15.6	91.3 ± 16.8
kidney_right	90.1 ± 25.2	89.7 ± 25.8	3.63 ± 6.58	3.67 ± 7.35	92.2 ± 16.1	91.8 ± 17.7
liver	92.9 ± 17.3	87.9 ± 27.0	6.40 ± 14.7	9.30 ± 30.8	95.0 ± 12.3	95.2 ± 12.3
lung_lower_lobe_left	93.2 ± 11.1	87.4 ± 25.1	2.92 ± 3.12	7.85 ± 25.9	92.9 ± 12.7	92.7 ± 13.3
lung_lower_lobe_right	90.4 ± 20.0	86.5 ± 26.3	2.90 ± 3.38	10.8 ± 37.5	92.3 ± 14.3	92.4 ± 14.6
lung_middle_lobe_right	88.5 ± 15.5	84.5 ± 23.9	4.37 ± 4.96	6.85 ± 18.0	90.6 ± 9.94	90.6 ± 9.70
lung_upper_lobe_left	91.1 ± 17.1	90.0 ± 20.0	3.95 ± 7.28	5.40 ± 14.1	93.5 ± 6.86	93.6 ± 6.03
lung_upper_lobe_right	72.7 ± 38.4	64.5 ± 43.6	5.17 ± 9.24	5.48 ± 9.83	87.3 ± 22.4	87.3 ± 25.1
pancreas	88.1 ± 22.3	88.6 ± 21.5	3.57 ± 3.58	5.29 ± 13.5	82.7 ± 19.4	83.3 ± 19.1
prostate	44.4 ± 45.5	40.3 ± 43.1	3.75 ± 3.36	4.60 ± 6.07	79.1 ± 20.6	79.6 ± 13.4
small_bowel	82.6 ± 22.9	83.5 ± 23.0	12.6 ± 17.7	12.5 ± 25.9	84.8 ± 12.9	85.7 ± 13.0
spleen	95.8 ± 12.9	91.1 ± 23.4	3.15 ± 6.72	6.81 ± 22.4	96.6 ± 2.36	96.6 ± 18.6
stomach	89.6 ± 20.1	82.0 ± 30.8	7.59 ± 14.1	11.8 ± 19.3	90.2 ± 15.8	90.5 ± 15.0
thyroid_gland	92.7 ± 21.6	70.2 ± 44.3	5.74 ± 16.9	4.80 ± 19.6	87.3 ± 8.14	87.1 ± 11.0
trachea	89.9 ± 27.9	82.8 ± 36.2	7.56 ± 34.8	5.24 ± 17.8	92.8 ± 5.40	92.2 ± 6.35
urinary_bladder	83.2 ± 24.3	75.0 ± 32.5	9.52 ± 21.9	17.0 ± 32.5	87.2 ± 15.5	86.4 ± 15.9
mean	80.4 ± 21.1	76.7 ± 25.7	6.18 ± 11.9	7.89 ± 17.9	84.2 ± 15.9	84.1 ± 15.4

Table 2. Vertebrae Segmentation Comparison on UNETR-Swin Backbone

	NSD ↑		**Haus$_{95}$** ↓		**Dice** ↑	
	AIC-Net	Baseline	AIC-Net	Baseline	AIC-Net	Baseline
sacrum	92.9 ± 22.1	78.2 ± 38.7	1.45 ± 0.56	16.6 ± 44.0	89.3 ± 21.3	90.0 ± 19.2
vertebrae_C1	93.6 ± 21.0	41.0 ± 48.7	1.48 ± 0.77	12.7 ± 51.5	83.6 ± 20.2	84.1 ± 20.9
vertebrae_C2	97.3 ± 6.00	58.9 ± 48.9	2.04 ± 3.23	1.37 ± 0.71	86.1 ± 14.2	87.6 ± 12.8
vertebrae_C3	97.7 ± 7.06	66.0 ± 47.9	1.29 ± 0.57	1.24 ± 0.49	86.8 ± 17.0	90.3 ± 8.12
vertebrae_C4	91.9 ± 25.6	56.7 ± 49.1	1.36 ± 0.64	9.03 ± 27.8	83.9 ± 23.8	83.4 ± 24.1
vertebrae_C5	93.4 ± 22.2	70.5 ± 43.1	1.38 ± 0.96	1.81 ± 1.49	83.8 ± 21.0	84.9 ± 14.7
vertebrae_C6	92.5 ± 24.3	82.7 ± 35.2	1.15 ± 0.46	17.9 ± 64.8	79.9 ± 25.3	82.4 ± 20.9
vertebrae_C7	96.2 ± 16.4	74.8 ± 43.1	1.42 ± 1.68	3.53 ± 14.5	92.8 ± 2.90	93.8 ± 1.81
vertebrae_L1	93.6 ± 22.1	92.7 ± 24.3	1.74 ± 2.83	2.48 ± 6.98	90.9 ± 20.4	93.3 ± 15.5
vertebrae_L2	96.6 ± 14.0	90.8 ± 27.0	1.73 ± 2.64	2.77 ± 8.28	92.8 ± 13.6	95.1 ± 5.79
vertebrae_L3	98.4 ± 4.97	92.4 ± 24.8	1.92 ± 2.82	1.44 ± 1.77	94.6 ± 6.20	95.7 ± 3.00
vertebrae_L4	97.1 ± 13.5	92.5 ± 24.4	1.57 ± 2.40	8.77 ± 42.5	93.2 ± 13.3	94.9 ± 6.19
vertebrae_L5	99.0 ± 2.98	97.4 ± 13.1	1.49 ± 2.10	1.22 ± 0.72	94.7 ± 3.78	95.4 ± 2.42
vertebrae_S1	93.5 ± 22.8	92.0 ± 25.4	1.35 ± 1.36	1.32 ± 0.93	89.8 ± 18.0	91.3 ± 13.1
vertebrae_T1	99.5 ± 1.45	64.9 ± 47.5	1.30 ± 1.22	12.4 ± 33.6	94.0 ± 23.0	94.6 ± 1.69
vertebrae_T10	96.1 ± 15.0	89.3 ± 28.2	1.78 ± 2.66	2.00 ± 3.06	91.4 ± 17.7	90.6 ± 20.1
vertebrae_T11	95.3 ± 17.9	88.4 ± 30.3	1.70 ± 2.55	1.61 ± 2.26	92.6 ± 14.7	93.1 ± 14.3
vertebrae_T12	96.4 ± 15.6	90.0 ± 28.5	1.92 ± 3.26	2.08 ± 3.78	92.3 ± 17.1	92.1 ± 19.0
vertebrae_T2	98.5 ± 5.17	74.7 ± 42.2	1.68 ± 1.90	8.49 ± 27.3	93.0 ± 6.70	93.5 ± 7.21
vertebrae_T3	97.0 ± 11.0	74.0 ± 42.8	2.05 ± 2.56	14.2 ± 39.0	91.8 ± 12.5	92.3 ± 13.6
vertebrae_T4	96.0 ± 16.1	73.7 ± 42.2	1.67 ± 2.10	19.4 ± 49.4	90.0 ± 16.6	90.1 ± 18.0
vertebrae_T5	95.0 ± 16.4	72.7 ± 42.7	2.27 ± 2.99	16.2 ± 41.7	89.5 ± 15.9	90.3 ± 15.8
vertebrae_T6	94.5 ± 16.3	80.3 ± 37.7	2.03 ± 2.24	1.62 ± 1.68	86.1 ± 20.9	87.7 ± 21.1
vertebrae_T7	87.7 ± 28.2	86.4 ± 30.4	3.57 ± 7.08	3.72 ± 7.09	81.9 ± 28.1	83.8 ± 26.6
vertebrae_T8	92.4 ± 22.8	88.9 ± 28.3	2.71 ± 4.67	5.33 ± 22.6	86.8 ± 23.9	85.9 ± 24.8
vertebrae_T9	95.3 ± 17.6	90.3 ± 26.1	1.71 ± 2.36	2.11 ± 3.01	90.6 ± 20.1	89.9 ± 20.2
mean	95.3 ± 15.7	79.2 ± 35.4	1.76 ± 2.25	6.59 ± 19.3	89.3 ± 16.1	90.2 ± 14.3

Table 3. Ribs Segmentation Comparison on UNETR-Swin Backbone

	NSD ↑		Haus$_{95}$ ↓		Dice ↑	
	AIC-Net	Baseline	AIC-Net	Baseline	AIC-Net	Baseline
costal_cartilages	94.5± 16.7	84.9 ± 32.2	6.83± 23.5	8.43 ± 23.3	85.7 ± 6.80	86.3 ± 6.30
rib_left_1	99.1 ± 1.30	57.6 ± 48.3	1.28 ± 0.64	32.4 ± 74.6	91.4 ± 3.30	90.1 ± 4.90
rib_left_10	94.9 ± 11.7	80.1 ± 34.8	7.82 ± 23.0	22.1 ± 48.087	89.5 ± 13.0	85.9 ± 19.3
rib_left_11	93.8 ± 16.8	79.2 ± 36.1	4.65 ± 7.65	28.9 ± 55.5	88.1 ± 17.0	85.2 ± 21.8
rib_left_12	92.8 ± 22.3	82.3 ± 32.6	2.92 ± 7.17	38.5 ± 61.2	89.8 ± 9.60	87.4 ± 9.60
rib_left_2	98.4 ± 6.10	67.7 ± 44.9	2.25 ± 5.00	21.1 ± 57.3	90.7 ± 8.70	89.0 ± 12.2
rib_left_3	92.0 ± 22.4	61.9 ± 45.5	3.63 ± 7.27	33.2 ± 59.2	85.7 ± 21.5	82.4 ± 23.6
rib_left_4	90.2 ± 23.0	74.4 ± 38.1	4.67 ± 6.79	19.6 ± 42.1	84.8 ± 22.2	84.0 ± 18.6
rib_left_5	86.9 ± 25.6	79.3 ± 35.8	8.59 ± 27.9	17.9 ± 43.3	83.9 ± 20.4	85.1 ± 19.9
rib_left_6	88.2 ± 24.8	84.8 ± 29.8	10.6 ± 31.8	14.9 ± 41.9	82.0 ± 24.1	84.7 ± 19.0
rib_left_7	91.7 ± 16.9	82.2 ± 32.5	5.57 ± 8.48	16.7 ± 43.4	85.9 ± 18.0	86.5 ± 18.2
rib_left_8	93.9 ± 11.6	82.0 ± 29.9	6.50 ± 14.0	24.0 ± 50.2	88.1 ± 12.9	85.9 ± 14.4
rib_left_9	96.2 ± 8.40	80.4 ± 32.4	3.97 ± 6.19	17.7 ± 39.2	90.5 ± 8.50	87.1 ± 13.5
rib_right_1	96.7 ± 15.7	62.4 ± 47.3	1.21 ± 0.60	38.3 ± 80.8	91.3 ± 3.60	90.5 ± 4.50
rib_right_10	92.7 ± 19.8	78.0 ± 36.3	5.02 ± 8.47	27.4 ± 60.8	87.7 ± 19.3	87.6 ± 15.4
rib_right_11	94.0 ± 18.8	85.3 ± 30.3	3.92 ± 8.25	22.3± 46.2	88.5 ± 18.4	87.4 ± 18.0
rib_right_12	91.0 ± 27.2	83.1 ± 33.7	2.16 ± 7.60	12.0 ± 31.5	89.4 ± 14.3	87.8 ± 12.0
rib_right_2	98.4 ± 7.10	74.2 ± 41.9	1.54 ± 2.98	2.39 ± 4.39	91.3 ± 8.30	89.5 ± 10.7
rib_right_3	96.8 ± 12.9	72.0 ± 42.4	2.65 ± 8.15	8.56 ± 31.0	89.5 ± 13.4	86.7 ± 17.5
rib_right_4	93.8 ± 17.4	75.3 ± 38.9	3.53 ± 7.45	16.44 ± 43.9	89.2 ± 11.8	84.8 ± 19.5
rib_right_5	91.5 ± 20.7	78.9 ± 34.6	6.54 ± 22.7	11.0 ± 31.0	86.6 ± 17.5	85.1 ± 15.8
rib_right_6	93.6 ± 15.4	82.6 ± 30.3	4.37 ± 5.74	11.7 ± 27.7	87.6 ± 16.6	84.8 ± 17.1
rib_right_7	92.6 ± 18.7	83.2 ± 28.6	6.65 ± 14.3	14.3 ± 28.9	87.6 ± 18.2	86.3 ± 13.5
rib_right_8	90.9 ± 22.3	83.6 ± 28.4	4.90 ± 8.63	12.8 ± 22.9	86.4 ± 18.9	86.2 ± 13.9
rib_right_9	91.4 ± 19.9	76.2 ± 36.5	5.44 ± 7.86	22.1 ± 45.8	86.3 ± 20.4	87.0 ± 14.7
sternum	95.8 ± 12.9	85.9 ± 28.1	1.97 ± 3.14	23.2 ± 57.6	89.8 ± 9.50	86.5 ± 14.8
mean	93.5 ± 16.8	77.6 ± 35.8	4.58 ± 10.6	19.9 ± 44.3	88.0 ± 14.5	86.5 ± 15.0