1. Supplementary Material

1.1. DS1 and DS2 Experiments with Efficient Channel (ECA) and Triplet Attention (TA) Mechanisms

Our experiments on Efficient Channel and Triplet attention mechanisms showed that evaluation metrics were either lower or the same as those for the SCA attention mechanism we reported in the paper; see Table S1.

1.2. Rotation Sensitivity of Soft Skeletonization on DS2

We tested different affine rotation transformations on DS2 for both clDice and GATS. The range of the rotation sets were coded as follows for *RangeOfSet(Start, Stop, Step)*: *Set 1*: {0, 90, 15}, *Set 2*: {0, 360, 90}, and *Set 3*: {0, 325, 65}. Table S2 shows the mean and standard deviation on test sets of DS2 for 10 trials. MPR uses N random samples the input volume. As can be seen in Table S2, the performance of a model trained on clDice with Set 2 or Set 3 rotations was degraded, indicating that skeletons produced via soft-skeletonization may be prone to rotation sensitivity due to over-thinning. Rotation sensitivity was not observed for either DS1 or Janelia to as much of an appreciable degree as on DS2 (i.e., 10-trial averages were within their respective standard deviations), likely because DS2 has greater voxel resolution so is more likely to more produce false positives and false negatives due to more instances of small/fine curvilinear structures.

1.3. Skeletal Warping: Discrete Cosine Transform Variant of Warping Loss

Our experiments with the Warping loss showed an increase in training time for DS1 to $\sim$34 hours for 10 trials with Warping loss versus $\sim$4 hours for clDice, $\sim$8 hours for GATS (N=10) or $\sim$6 hours for GATS (N=4). To expedite training time we formulated a novel variant of the Warping loss that applies a discrete cosine transform to the soft-skeletonized inputs of the prediction and the target image, and finds critical points of the skeletal topology as described previously; see Homotopic Warping under Related Literature. The evaluation metrics on DS1 of the Skeletal Warping loss are 0.817 $\pm$ 0.01, 0.795 $\pm$ 0.018, and 0.783 $\pm$ for Dice, clDice and ρ-Dice, respectively. These evaluation metrics are similar to the GATS SCA (N=10) loss, but Skeletal Warping loss is slower than GATS SCA by 20 hours. More experiments are required for testing on different datasets, and a different method is needed for improvements in training time.

Table S 1. **Efficient Channel Attention (ECA) and Triplet Attention (TA) on 3D Residual UNets. Mean and standard deviation evaluated on the test set of Dataset 1 (DS1) and Dataset 2 (DS2) across 10 trials. For clDice: $\alpha = 0.65, k = 3$, GASK and GATS both use MPR, for which N=10, GASK uses soft-skeletonization, GATS uses topological smoothing (TS). For TS: k=5 for DS1, k=3 for DS2.**

Attention	Dataset	Model	Dice ↑	clDice ↑	ρ-Dice ↑	Adj. Rand ↑	β_0 Error ↓	β_1 Error ↓
ECA	DS1	clDice	0.7951 ± 0.0070	0.7665 ± 0.0153	0.7521 ± 0.0184	0.5887 ± 0.0139	0.3572 ± 0.0176	0.0803 ± 0.0092
		GASK	0.7824 ± 0.0138	0.7287 ± 0.0457	0.7296 ± 0.0429	0.5632 ± 0.0277	0.3284 ± 0.0218	0.0698 ± 0.0147
		GATS	0.8031 ± 0.0084	0.7471 ± 0.0194	0.7713 ± 0.0205	0.6047 ± 0.0167	0.3223 ± 0.0250	0.0682 ± 0.0074
	DS2	clDice	0.7491 ± 0.0067	0.6046 ± 0.0155	0.6808 ± 0.0164	0.4976 ± 0.0133	0.0638 ± 0.0164	0.1199 ± 0.0115
		GASK	0.5924 ± 0.0139	0.6500 ± 0.0201	0.7449 ± 0.0137	0.1905 ± 0.0258	0.0602 ± 0.0259	0.4541 ± 0.0135
		GATS	0.7331 ± 0.0114	0.5696 ± 0.0247	0.6424 ± 0.0264	0.4657 ± 0.0227	0.0508 ± 0.0092	0.1040 ± 0.0119
TA	DS1	clDice	0.7970 ± 0.0076	0.7650 ± 0.0189	0.7537 ± 0.0197	0.5924 ± 0.0152	0.3611 ± 0.0207	0.0817 ± 0.0112
		GASK	0.7892 ± 0.0124	0.7399 ± 0.0332	0.7348 ± 0.0344	0.5769 ± 0.0248	0.3085 ± 0.0334	0.0618 ± 0.0115
		GATS	0.8058 ± 0.0050	0.7515 ± 0.0067	0.7701 ± 0.0133	0.6100 ± 0.0100	0.3076 ± 0.0424	0.0627 ± 0.0064
	DS2	clDice	0.7423 ± 0.0115	0.5912 ± 0.0241	0.6673 ± 0.0268	0.4840 ± 0.0229	0.0674 ± 0.0127	0.1268 ± 0.0110
		GASK	0.6059 ± 0.0273	0.6486 ± 0.0261	0.7437 ± 0.0152	0.2165 ± 0.0520	0.0606 ± 0.0237	0.4450 ± 0.0280
		GATS	0.7325 ± 0.0089	0.5707 ± 0.0203	0.6443 ± 0.0205	0.4645 ± 0.0177	0.0455 ± 0.0070	0.1028 ± 0.0080

Table S 2. **DS2 Affine Rotation Transformation Sensitivity**

Model	Dice ↑	clDice ↑	ρ-Dice ↑	Adj. Rand ↑	β_0 Error ↓	β_1 Error ↓
clDice (Set 1)	0.7880 ± 0.0140	0.6890 ± 0.0270	0.7750 ± 0.0300	0.5760 ± 0.0270	0.1270 ± 0.0310	0.1510 ± 0.0170
clDice (Set 2)	0.2360 ± 0.0554	0.0315 ± 0.0159	0.0901 ± 0.0318	−0.0053 ± 0.0017	0.8076 ± 0.1591	0.5619 ± 0.0594
clDice (Set 3)	0.1010 ± 0.1592	0.0718 ± 0.0555	0.1147 ± 0.0612	0.0002 ± 0.0077	0.2317 ± 0.3270	0.6015 ± 0.0928
GATS (N=10, Set 1)	0.7321 ± 0.0065	0.5699 ± 0.0149	0.6418 ± 0.0163	0.4637 ± 0.0129	0.0461 ± 0.0076	0.1023 ± 0.0094
GATS (N=10, Set 2)	0.8029 ± 0.0068	0.7146 ± 0.0156	0.8060 ± 0.0156	0.6052 ± 0.0135	0.0472 ± 0.0100	0.0956 ± 0.0135
GATS (N=10, Set 3)	0.8029 ± 0.0062	0.7149 ± 0.0122	0.7997 ± 0.0159	0.6052 ± 0.0124	0.0565 ± 0.0192	0.1081 ± 0.0220