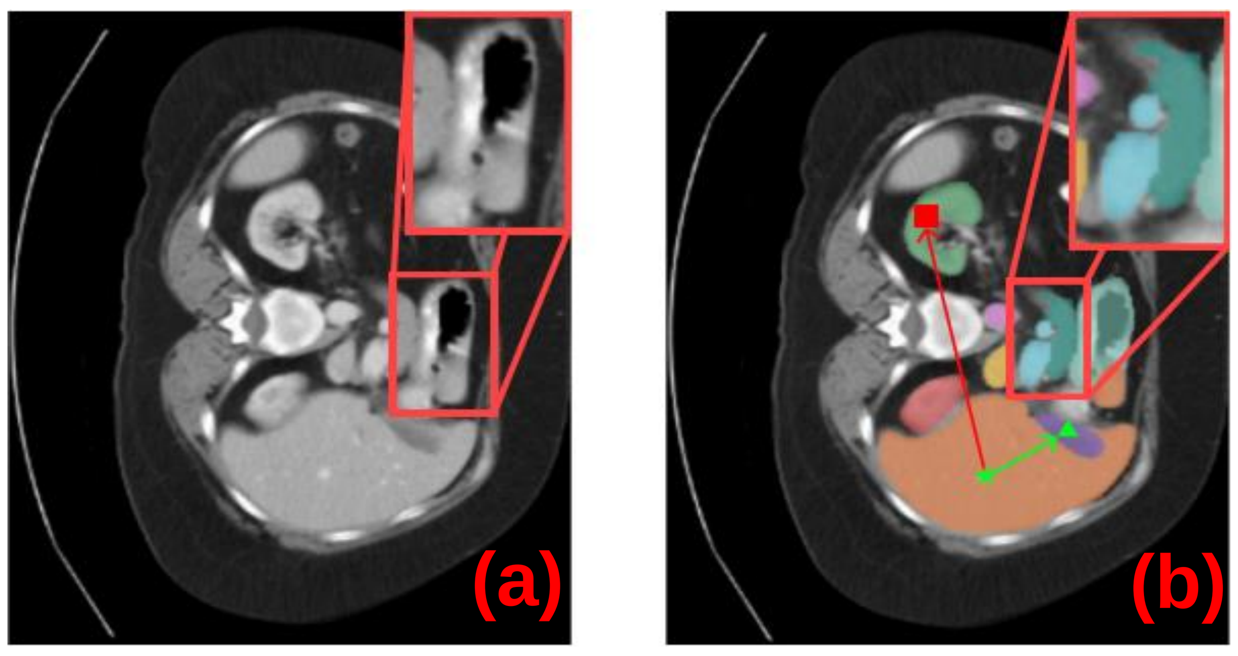
Advancing Medical Image Segmentation: Morphology-Driven Learning with Diffusion Transformer

Sungmin Kang[1] Jaeha Song[2] Jihie Kim[1]

[1]Department of Computer Science and Artificial Intelligence, Dongguk University, Seoul, Korea
[2]Department of Computer Science and Engineering, Dongguk University, Seoul, Korea
rkdtjdals97@dgu.ac.kr archiiive99@gmail.com jihie.kim@dgu.edu

Background

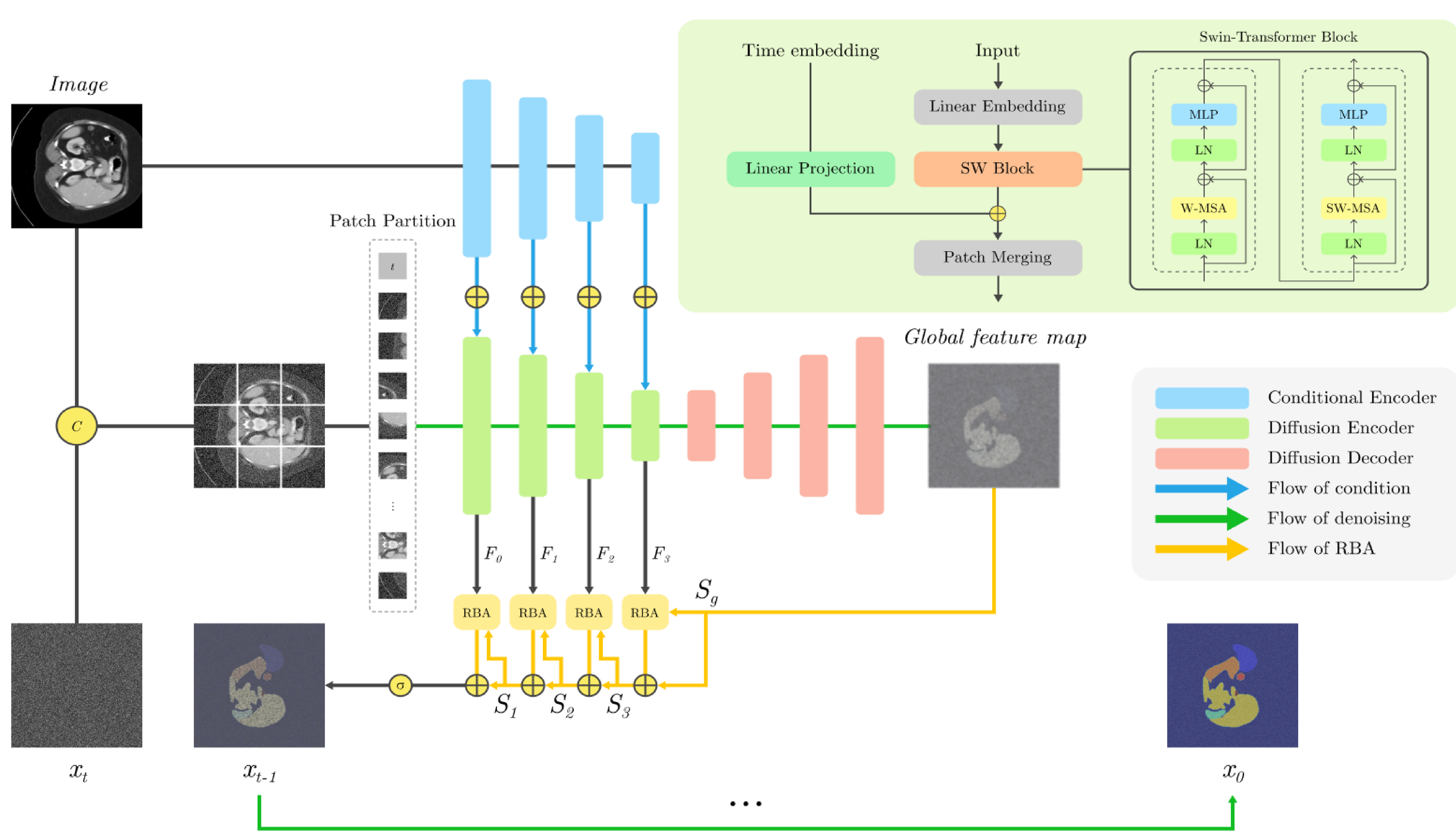
- ***The unique properties of medical imaging(a)*** make clear segmentation difficult, and the high cost and time-consuming task of labeling leads to a ***coarse-grained representation of ground truth(b).***



Contributions

- We presents a new ***diffusion transformer segmentation (DTS)*** model which performs better than previous framework.
- We introduce a novel approach to address the medical image segmentation by integrating ***morphology-driven learning*** into the image processing, such as **(1)** ***k-neighbor label smoothing,*** **(2)** ***reverse boundary attention,*** **(3)** ***self-supervised learning.***
- Our model demonstrates the generality in segmentation tasks in medical modalities such as CT, MRI, and lesion images and further suggests that this approach may be adaptable to other domains.

Architecture



DTS: Diffusion Transformer Segmentation

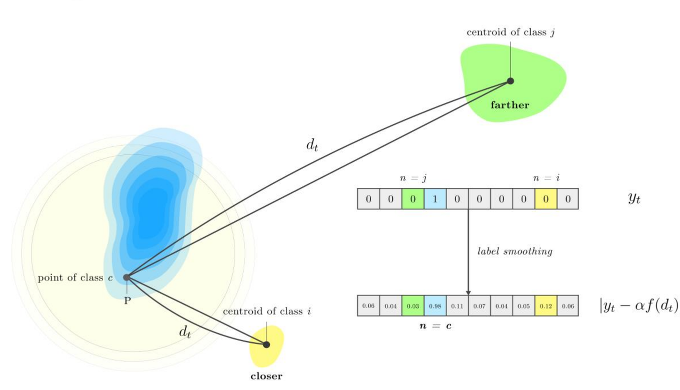
- We suggest the possibility of **replacing the diffusion encoder with a Swin transformer,** which offer scalability and computational efficiency when processing various images due to its hierarchical structure.

$$\varepsilon_\theta(x_t, I, t) = DTS((x_t, I), t, \tau_\theta(I))$$

Morphology-driven learning

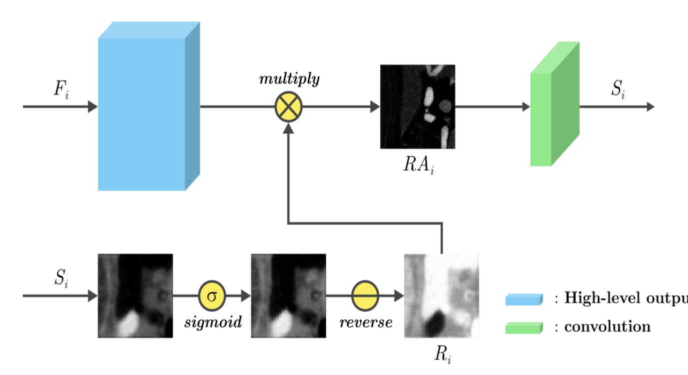
k-Neighbor Label smoothing by organ distance.

- k –nls that leverages the relative positions of organs for ***distance-aware smoothing of the labels of k-neighbors for a given class or organ. (k > 2)***



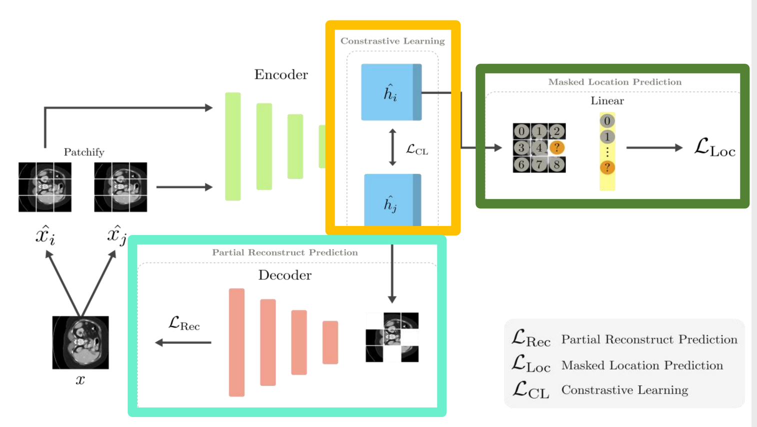
RBA: Reverse-boundary Attention.

- RBA aims to improve the prediction of models by gradually capturing and specifying areas that may have been initially ambiguous.

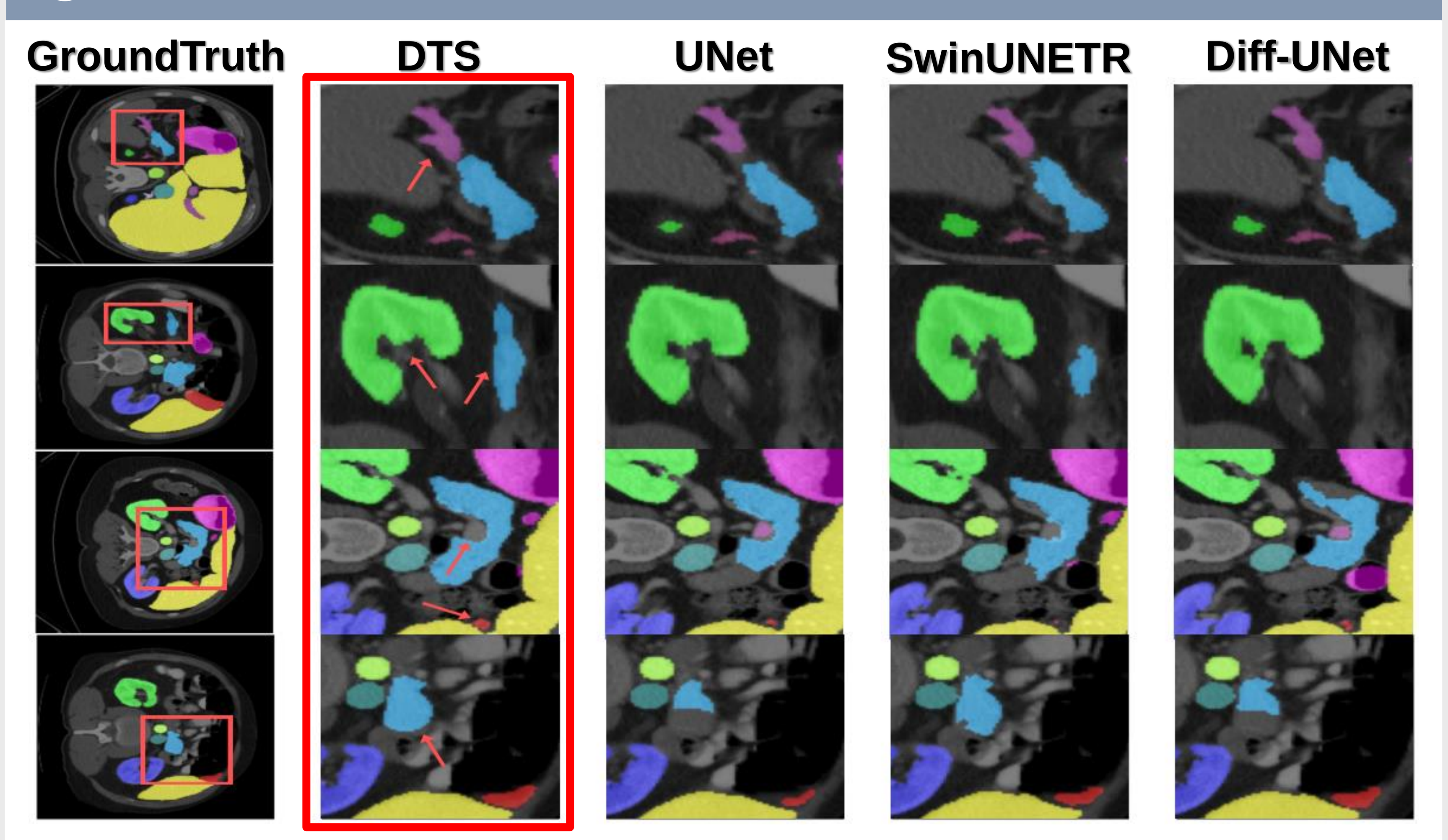


Self-supervised learning (SSL)

- Our SSL framework combine three proxy tasks.
- *Contrastive learning*
- *Masked Location Prediction*
- *Partial Reconstruct Prediction*



Qualitative Results



- The region of interest was highlighted with arrows.
- Our model(DTS) captures fine-grained details and achieves precise boundary representations.

Quantitative Results

Multi-organ segmentation(CT)

Method	Spleen	Kidney	Gall	Esophagus	Liver	Stomach	Aorta	IVC	Veins	Pancreas	AG	Avg.
TransUNet [9]	0.952	0.928	0.662	0.757	0.969	0.889	0.920	0.833	0.791	0.775	0.637	0.828
nnUNet [illegible]	0.947	0.920	0.794	0.812	0.955	0.905	0.908	0.850	0.812	0.829	0.764	0.863
UNETR [18]	0.952	0.928	0.805	0.824	0.963	**0.925**	0.928	0.857	0.828	0.832	0.781	0.874
Swin UNETR [19]	0.956	0.937	0.828	0.827	0.971	0.921	0.928	0.863	0.849	0.858	0.810	0.886
EnsemDiff [illegible]	0.905	0.911	0.732	0.723	0.947	0.838	0.915	0.838	0.704	0.715	0.637	0.805
SegDiff [4]	0.894	0.881	0.703	0.654	0.852	0.702	0.874	0.819	0.715	0.724	0.694	0.774
MedsegDiff [54]	0.969	0.930	0.822	0.817	0.970	0.919	0.912	0.859	0.831	0.813	0.791	0.875
Diff-UNet [56]	**0.973**	**0.942**	0.812	0.815	**0.973**	0.924	0.907	0.868	0.825	0.788	0.779	0.873
Ours*	0.972	**0.942**	**0.903**	**0.847**	0.972	0.924	**0.945**	**0.874**	**0.867**	**0.880**	**0.842**	**0.906**

- Our model(DTS), which demonstrate good segmentation performance for small organs.(e.g. gall bladder, esophagus)

Tumor, Lesion segmentation(MRI, Skin image)

Method	BraTs WT		BraTs TC		BraTs ET		BraTs Average		ISIC Average	
	Dice↑	HD↓	Dice↑	HD↓	Dice↑	HD↓	Dice↑	HD↓	Dice↑	HD↓
TransUNet [9]	78.95	5.87	81.60	5.05	76.15	5.91	78.90	5.87	85.40	3.88
UNETR [18]	89.92	2.49	84.79	4.07	79.51	5.77	84.74	4.08	87.57	3.21
SwinUNETR [19]	**90.04**	**2.41**	85.19	3.94	80.01	5.69	85.09	3.97	89.68	2.57
SegDiff [4]	80.51	5.23	82.32	4.83	73.24	6.84	78.69	5.87	87.30	3.32
MedsegDiff [54]	89.49	2.71	85.12	3.96	79.12	5.81	84.57	4.13	89.89	2.57
Diff-UNet [56]	88.23	2.94	86.94	3.40	79.87	5.79	85.01	4.01	88.64	2.94
Ours*	89.63	2.57	**88.02**	**3.07**	**81.11**	**5.12**	**86.25**	**3.62**	**91.12**	**2.18**

Acknowledgements

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2024-2020-0-01789), and the Artificial Intelligence Convergence Innovation Human Resources Development (IITP-2024-RS-2023-00254592) supervised by the IITP(Institute for Information & Communications Technology Planning & Evaluation).

Contact

Machine Learning LAB
https://www.ml.dongguk.edu/