Multi-Aperture Fusion of Transformer-Convolutional Network (MFTC-Net) for 3D Medical Image Segmentation and Visualization

Siyavash Shabani, Muhammad Sohaib, Sahar Mohammad, and Bahram Parvin
Biomedical Engineering, University of Nevada, Reno
{sshabani, bparvin}@unr.edu

CVPR SEATTLE, WA JUNE 17-21, 2024

Background:

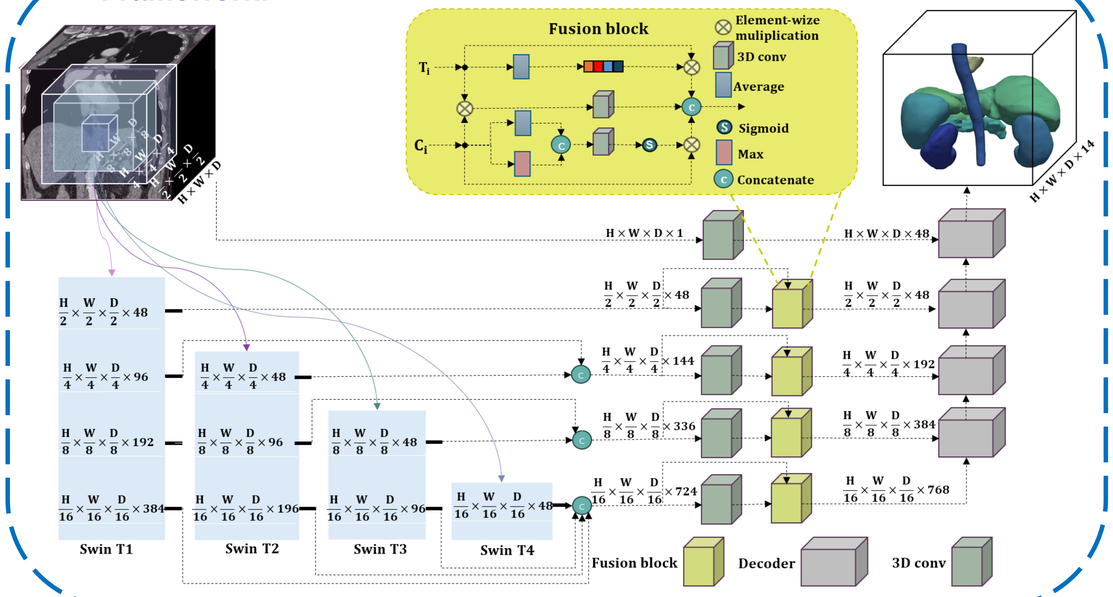
Motivation:

- Designing a framework for 3D medical image segmentation and visualization.
- Finding a proper combination of global and local feature maps from convolutional and vision transformers.
- Incorporating morphometric features of medical images into the loss function.

Contributions:

- Four parallel Swin Transformers modules that couple the pyramid representation with the original resolution.
- A 3D fusion block with a squeeze and excitation block and a convolutional block attention module (CBAM) for combining the outputs of each Transformer and its convolutional block.
- A loss function that integrates the Dice cross-entropy with a custom distance transform for morphometric consistency
- A reduce model complexity with 40M parameters—an improvement over published result.

Framework:



Ablation study:

The ablation study indicates superior performance of multi-aperture Swin Transformer.

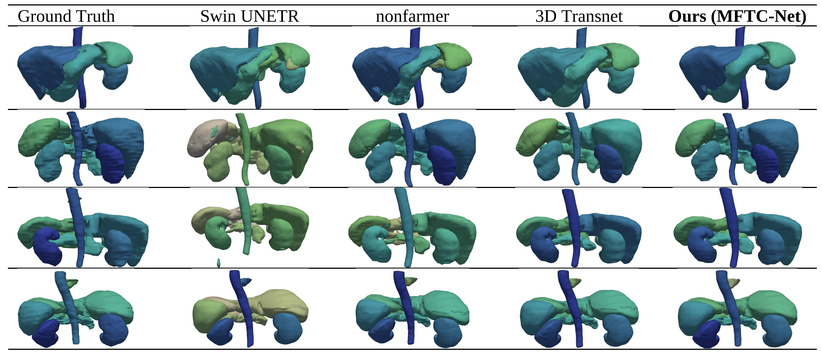
Loss Function	Dice
L_{dice}	89.04
L_{focal}	89.50
$L_{dice} + L_{ce}$	89.17
$L_{dice} + L_{ce} + DistLoss$	**89.73**

The ablation study indicates an improved performance with the inclusion of DistLoss function.

Blocks	Parameters	Dice
T_1	26M	86.12
$T_1 + T_2$	30M	86.76
$T_1 + T_2 + T_3$	34M	87.39
$T_1 + T_2 + T_3 + T_4$	36M	88.75
$T_1 + T_2 + T_3 + T_4 + Fusion$	40M	**89.73**

Results: Comparable performance with SOFA

Methods	Avg Dice↑	Avg HD95 ↓	Par	Spl	Kid(R)	Kid(L)	Gal	Liv	Sto	Aor	Pan
MIST [1]	86.92	11.07	-	92.83	93.28	92.54	74.58	94.94	**87.23**	89.15	72.43
Swin UNETR [2]	83.47	10.55	62M	95.37	86.26	86.99	66.54	95.72	77.01	91.12	68.80
nnFormer [4]	86.57	10.63	150M	90.51	86.25	86.57	70.17	96.84	86.83	92.04	**83.35**
UNETR++ [6]	87.22	07.53	42M	95.77	87.18	87.54	71.25	96.42	86.01	92.52	81.10
MISSFormer [7]	81.96	18.20	-	91.92	82.00	85.21	68.65	94.41	80.81	86.99	65.67
LeVit-UNet [10]	78.53	16.84	-	88.86	80.25	84.61	62.23	93.11	72.76	87.33	59.07
3DTransUNet [11]	88.10	-	81M	93.39	87.47	85.76	**81.15**	**97.34**	85.31	92.97	81.76
MFTC-Net	89.17	07.47	40M	95.98	92.66	93.30	76.41	96.61	85.05	92.62	80.77
MFTC-Net +DistLoss	**89.73**	**07.31**	**40M**	**96.51**	**93.35**	**93.64**	76.87	97.16	85.69	**93.05**	81.20



Discussion:

- Proposed a new framework, labeled as MFTC-Net includes four coupled Swin Transformers, at multiple apertures.
- The model was evaluated on the Synapse multi-organs dataset.
- Our model outperformed published research on Synapse multi-organs datasets, with better mean Dice and mean HD95 scores.

Acknowledgment: The research was supported by a grant from NIH RO1CA279408.