

Medical Transformer: Gated Axial-Attention for Medical Image Segmentation

Jeya Maria Jose Valanarasu¹, Poojan Oza¹, Ilker Hacihaliloglu², and Vishal M. Patel¹

¹Johns Hopkins University

²Rutgers, The State University of New Jersey

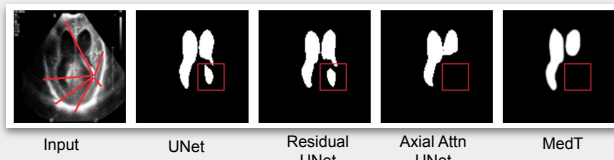


JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING

RUTGERS
THE STATE UNIVERSITY
OF NEW JERSEY

Motivation

- Current Methods for medical image segmentation are based on convolutional neural networks (CNNs).
- CNNs lack understanding of long range dependencies in the image due to inherent inductive biases.



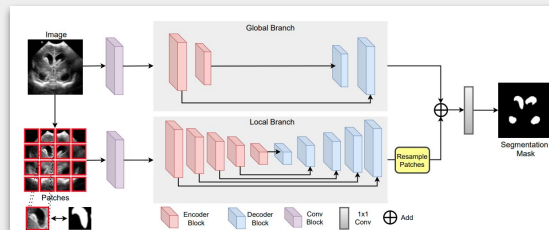
- From the above figure, it can be seen that CNN methods miss-classify some bleeding inside ventricle area as true positive due to failure to understand long-range dependencies where as transformer based methods predict accurate segmentation masks.

Contributions

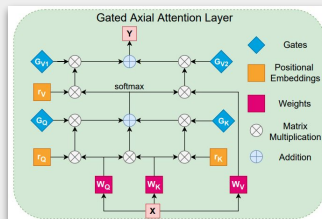
- In this work, we propose a transformer based method for medical image segmentation - **MedT** which focuses on understanding long-range dependencies to improve segmentation.
- Specifically, we propose a **Gated Axial Attention** module to help train transformers for low-data regime.
- We also propose a **Local-Global (LoGo)** training strategy to extract both global and local features effectively.
- We successfully improve the performance for medical image segmentation tasks over convolutional networks and fully attention architectures on three different datasets.

Method

Medical Transformer (MedT) architecture

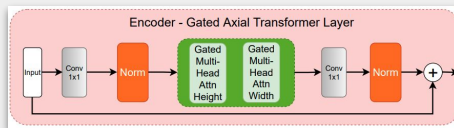


Gated Axial Attention Module



$$y_{ij} = \sum_{w=1}^W \text{softmax} \left(q_{ij}^T k_{iw} + G_Q q_{ij}^T r_{iw}^q + G_R k_{iw}^T r_{iw}^k \right) (G_{V1} v_{iw} + G_{V2} r_{iw}^v)$$

Encoder Transformer Block

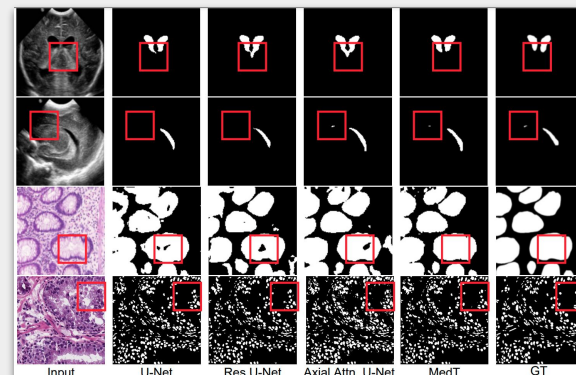


Experiments and Results

Quantitative Results

Type	Network	Brain US		GlaS		MoNuSeg	
		F1	IoU	F1	IoU	F1	IoU
Convolutional Baselines	FCN [1]	82.79	75.02	66.61	50.84	28.84	28.71
	U-Net [17]	85.37	79.31	77.78	65.34	79.43	65.99
	U-Net++ [31]	86.59	79.95	78.03	65.55	79.49	66.04
	Res-UNet [27]	87.50	79.61	78.83	65.95	79.49	66.07
	Axial Attention U-Net [24]	87.92	80.14	76.26	63.03	76.83	62.49
Fully Attention Baseline	Gated Axial Attn.	88.39	80.7	79.91	67.85	76.44	62.01
	LoGo	88.54	80.84	79.68	67.69	79.56	66.17
	MedT	88.84	81.34	81.02	69.61	79.55	66.17

Qualitative Results



Acknowledgement: This work was supported by the NSF grant 1910141.