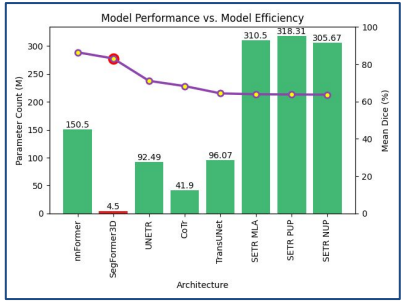
SegFormer3D: an Efficient Transformer for 3D Medical Image Segmentation

Shehan Perera, Pouyan Navard, Alper Yilmaz
Photogrammetric Computer Vision Lab, The Ohio State University

CVPR SEATTLE, WA JUNE 17-21, 2024

Motivation:

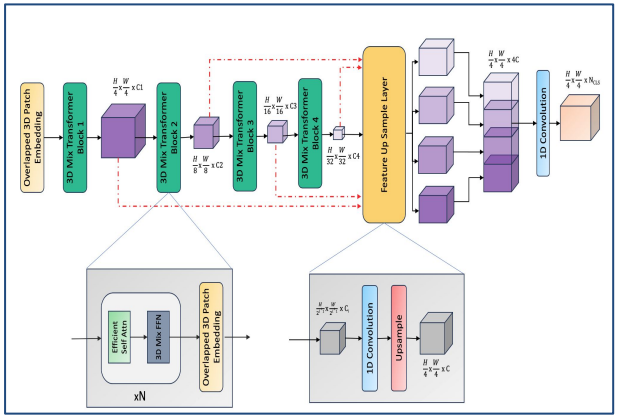
- Adoption of Vision Transformers (ViTs) in 3D Medical Imaging has resulted in significant advancements.
- However, converting 2D ViTs to handle 3D sequences have resulted in extremely large architectures that increases training and deployment complexities.
- In addition, large scale architectures require significant pretraining efforts to perform well in situations with limited datasets which is often seen in medical imaging.
- **Question:** Can we develop a Transformer based solution that is efficient, lightweight and high performing? **YES!**



Segformer3D Highlights:

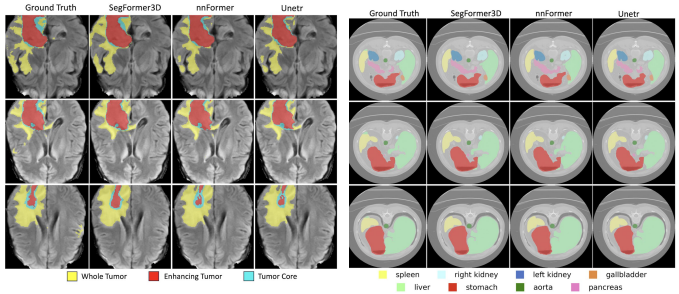
- We introduce a lightweight memory efficient segmentation model that preserves the performance characteristics of larger models for 3D medical imaging.
- With 4.5M parameters and 17 GFLOPs Segformer3D presents a 34x and 13x reduction in parameter count and model complexity compared SOTA and well established architectures.
- Why you should care: We showcase highly competitive results without pretraining against large well established architectures.

Segformer 3D:



- Key Takeaway: The Sequence Reduction Capability of Segformer3D helps reduce long sequence lengths [Sequence, H, W, C] ideal for 3D Medical Imaging, Video Analysis, and instances with Volumetric or Time Series Data

Experimental Results (Qualitative):



Experimental Results (Quantitative):

Methods	Params	Avg % ↑	AOR	LIV	LKID	RKID	GAL	PAN	SPL	STO
nnFormer[32]	150.5	86.57	92.04	96.84	86.57	86.25	70.17	83.35	90.51	86.83
Ours	**4.5**	**82.15**	**90.43**	**95.68**	**86.53**	**86.13**	**55.26**	**73.06**	**89.02**	**81.12**
MISSFormer[12]	–	81.96	86.99	94.41	85.21	82.00	68.65	65.67	91.92	80.81
UNETR[11]	92.49	79.56	89.99	94.46	85.66	84.80	60.56	59.25	87.81	73.99
SwinUNet[3]	–	79.13	85.47	94.29	83.28	79.61	66.53	56.58	90.66	76.60
LeVit-UNet-384[29]	52.17	78.53	87.33	93.11	84.61	80.25	62.23	59.07	88.86	72.76
TransClaw U-Net[4]	–	78.09	85.87	94.28	84.83	79.36	61.38	57.65	87.74	73.55
TransUNet[5]	96.07	77.48	87.23	94.08	81.87	77.02	63.16	55.86	85.08	75.62
R50-ViT+CUP[5]	86.00	71.29	73.73	91.51	75.80	72.20	55.13	45.99	81.99	73.95
ViT+CUP[5]	86.00	67.86	70.19	91.32	74.70	67.40	45.10	42.00	81.75	70.44

Table 3: Synapse comparisons ranked based on average performance across classes. Segformer3D is highly competitive, outperforming well-established solutions and second to only nnformer with 34x parameters.

Methods	Params	Avg % ↑	Whole Tumor ↑	Enhancing Tumor ↑	Tumor Core ↑
nnFormer[32]	150.5	86.4	91.3	81.8	86.0
Ours	**4.5**	**82.1**	**89.9**	**74.2**	**82.2**
UNETR[11]	92.49	71.1	78.9	58.5	76.1
TransBTS[25]	–	69.6	77.9	57.4	73.5
CoTr[28]	41.9	68.3	74.6	55.7	74.8
CoTr w/o CNN Encoder[28]	–	64.4	71.2	52.3	69.8
TransUNet[5]	96.07	64.4	70.6	54.2	68.4
SETR MLA[31]	310.5	63.9	69.8	55.4	66.5
SETR PUP[31]	318.31	63.8	69.6	54.9	67.0
SETR NUP[31]	305.67	63.7	69.7	54.4	66.9

Table 2: BRaTs comparison table ranked based on average performance across all classes. Segformer3D is highly competitive out performing well established solutions across all categories.

Methods	Params	Avg % ↑	RV	Myo	LV
nnFormer [32]	150.5	92.06	90.94	89.58	95.65
Ours	**4.5**	**90.96**	**88.50**	**88.86**	**95.53**
LeViT-UNet-384 [29]	52.17	90.32	89.55	87.64	93.76
SwinUNet [3]	–	90.00	88.55	85.62	95.83
TransUNet [5]	96.07	89.71	88.86	84.54	95.73
UNETR [11]	92.49	88.61	85.29	86.52	94.02
R50-VIT-CUP [5]	86.00	87.57	86.07	81.88	94.75
VIT-CUP [5]	86.00	81.45	81.46	70.71	92.18

Table 4: ACDC comparison ranked based on average performance across classes. Segformer3D is highly competitive outperforming well established solutions and is within 1% of SOTA with 150 million parameters.