Fuchen Zheng[12], Xuhang Chen[123], Weihuang Liu[1], Haolun Li[1], Yingtie Lei[1], Jiahui He[24], Chi-Man Pun[1*], and Shoujun Zhou[2*]

[1]University of Macau
[2]Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences
[3]Huizhou University
[4]University of Nottingham Ningbo China

IEEE BIBM2024 Paper B606 Report

SMAFormer: Synergistic Multi-Attention Transformer for Medical Image Segmentation

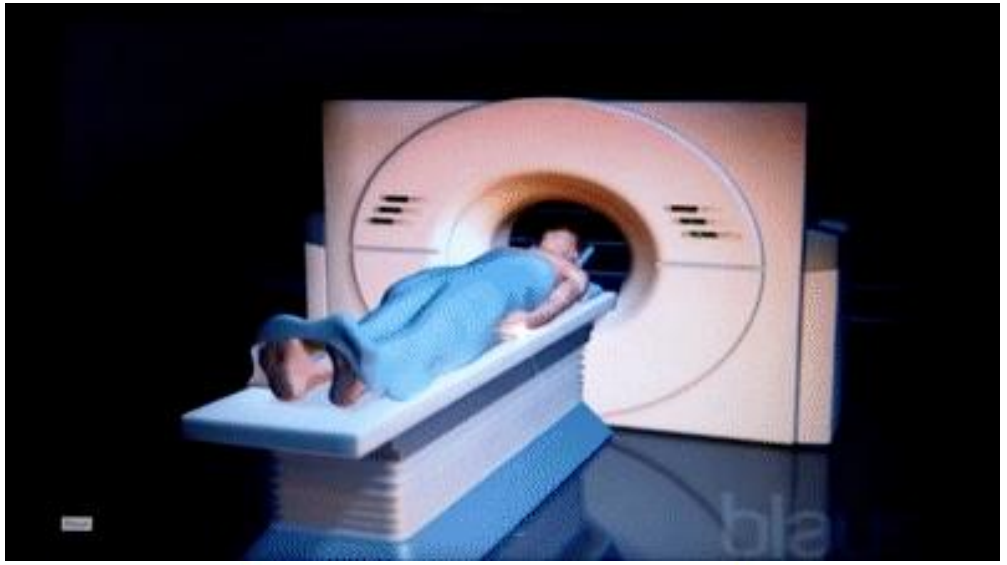
Contents

BASIC INFORMATION OF THE PAPER

01

- Medical imaging: Medical imaging is the process of visualizing abnormal locations based on physical phenomena. It mainly uses technologies such as X-rays, CT scans, and MRI to display the anatomical structure information inside the human body.

- Image Segmentation: Segmenting structures or tissues in medical images into different regions for locating and identifying targets of interest.



RESEARCH BACKGROUND

02

- Medical image segmentation is the process of dividing structures or tissues in medical images into different regions or objects. Traditional medical image segmentation methods have certain **limitations** in accurately segmenting **small and irregularly shaped tumors and organs**.
- (1) Difficulty in Assigning Attention to Relevant Regions: Transformers, especially when not fine-tuned for medical images, often struggle to focus attention on medically relevant regions, hindering their performance in multi-organ or multi-tumor segmentation tasks.
- (2) Limited Capture of Local Context: Local context plays a crucial role in accurately segmenting small structures like organs or tumors. Traditional Transformers, with their global receptive fields, often fail to adequately capture this local information.

RESEARCH IDEAS

03

In order to simultaneously combine the advantages of traditional convolutional neural networks (CNNs) and the multi-head self-attention mechanism in Transformer to achieve better results in medical image segmentation tasks. This study proposes a new Transformer architecture called SMAFormer for efficient and accurate medical image segmentation. Image segmentation. The research ideas of this study mainly include the following aspects:

(1) SMAFormer Architecture: A novel residual U-shaped Transformer model integrating attention mechanisms, U-shaped architecture, and residual connections for efficient and effective medical image segmentation.
(2) Learnable Segmentation Modulator: An embeddable module for multi-scale feature fusion, enhancing the synergy between different attention mechanisms.
(3) State-of-the-art Performance: Extensive experiments demonstrate SMAFormer achieves state-of-the-art results on various medical image segmentation datasets.

RESEARCH METHODS

04

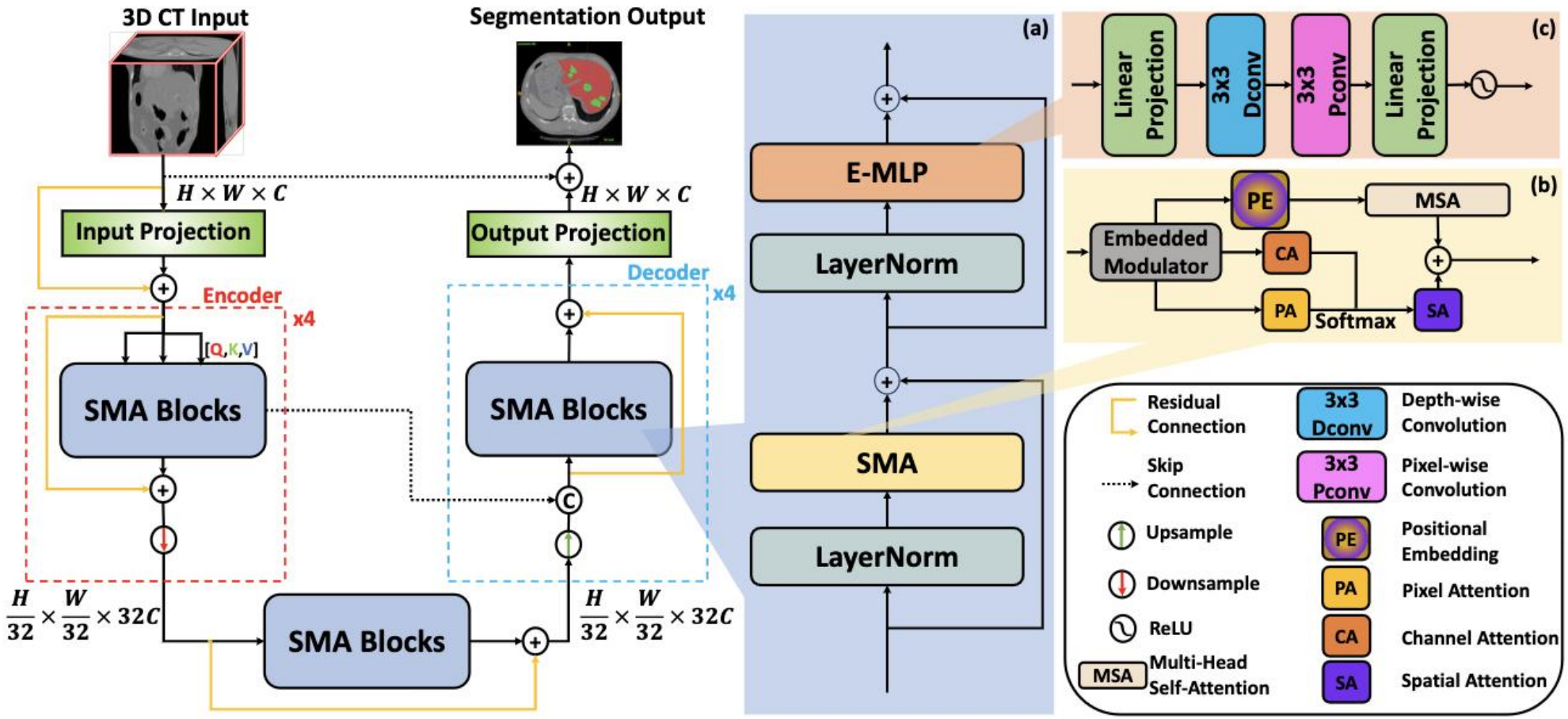


Fig. 1. This figure provides an overview of the SMAFormer architecture. The figure details (a) the SMA Transformer block, (b) the SMA Part within the SMA Transformer block, and (c) the E-MLP Part within the SMA Transformer block.

Positioned at each scale of the network, the modulator serves three primary functions:

- Positional Encoding: The modulator embeds positional information into the feature maps, compensating for the lack of inherent positional awareness in the Transformer architecture.
- Trainable Bias: The modulator incorporates a trainable bias term, which is added to the output of the multi-head self-attention mechanism. This bias term, updated during training, helps fine-tune the attention maps and improves the model's ability to focus on relevant regions.
- Facilitating Multi-Attention Computations: The modulator assists in performing the necessary transpositions and matrix multiplications required for the three different attention mechanisms within the SMA block. This ensures efficient computation and seamless integration of the multi-attention module within the overall architecture.

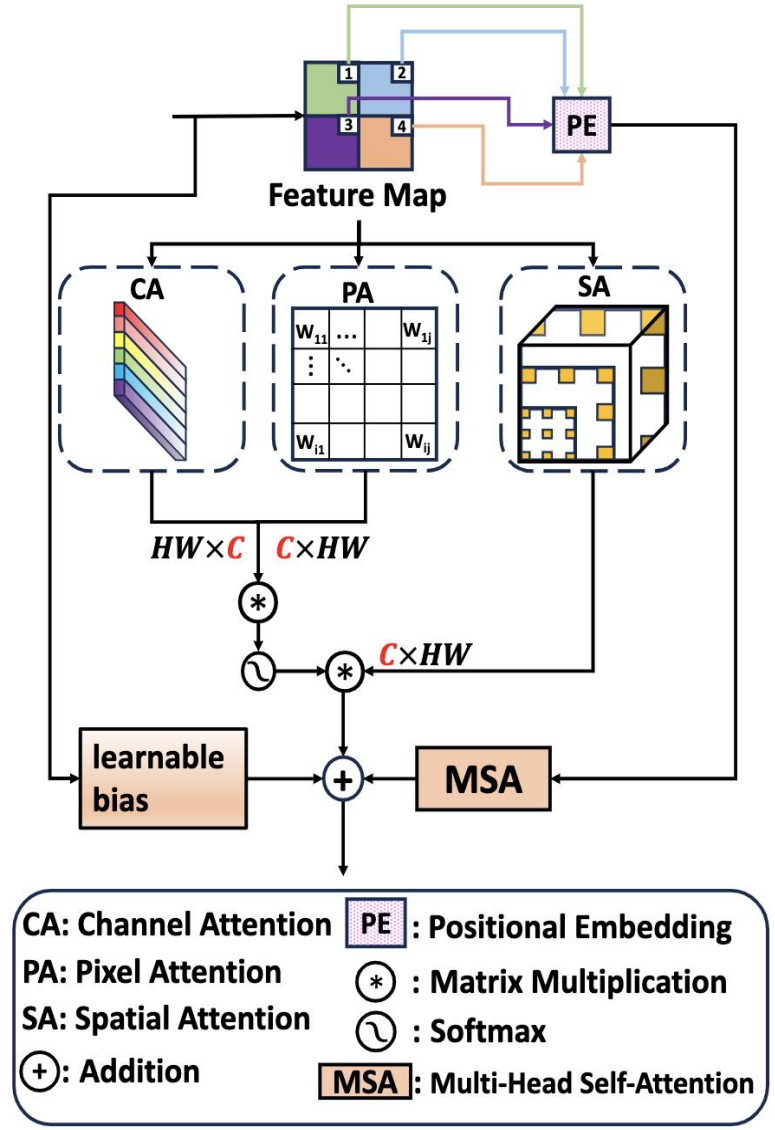


Fig. 2. This figure presents a schematic diagram of the proposed modulator.

RESULTS

05

EXPERIMENTS

- Datasets

This work utilizes three publicly available medical image segmentation datasets: (1) LiTS2017 (Liver and Tumor), (2) ISICDM2019 (Bladder and Tumor) and (3) Synapse (Multi-Organ).

- Evaluation Metrics

We evaluated the segmentation performance using Dice Coefficient Score (DSC) and Mean Intersection over Union (mIoU):

$$DSC = \frac{2 \times |P \cap G|}{|P| + |G|}, \qquad mIoU = \frac{1}{C} \sum_{i=1}^{C} \frac{|P_i \cap G_i|}{|P_i| + |G_i| - |P_i \cap G_i|},$$

Comparisons with State-of-the-Art Methods

TABLE I
COMPARISON WITH STATE-OF-THE-ART MODELS ON THE ISICDM2019 AND LITS2017 DATASETS. THE BEST RESULTS ARE BOLDED WHILE THE SECOND BEST ARE UNDERLINED.

Method	ISIDM2019				LITS2017			
	Average		Bladder	Tumor	Average		Bladder	Tumor
	DSC(%) ↑	mIoU(%) ↑	DSC(%) ↑	DSC(%) ↑	DSC(%) ↑	mIoU(%) ↑	DSC(%) ↑	DSC(%) ↑
ViT [26]+CUP [29]	88.60	84.40	91.88	85.32	80.33	77.25	83.97	76.69
R50-ViT [26]+CUP [29]	88.77	85.62	92.05	85.49	82.62	79.68	85.83	79.41
ResUNet++ [30]	87.11	83.78	89.90	84.32	75.73	74.19	79.12	72.34
ResT-V2-B [28]	89.26	82.13	93.01	85.50	78.53	75.24	81.22	75.83
TransUNet [29]	94.56	93.60	97.74	91.38	93.28	90.81	95.54	91.03
SwinUNet [40]	91.95	89.77	94.73	89.17	89.68	86.62	93.31	86.04
Swin UNETR [41]	92.60	90.61	95.08	90.12	91.95	90.02	94.73	89.17
UNETR [42]	91.55	88.34	94.83	88.26	89.38	87.46	92.89	85.86
nnFormer [43]	93.54	89.11	96.97	90.41	91.74	89.95	94.57	88.91
SMAFormer(Ours)	**96.07**	**94.67**	**98.57**	**93.56**	**94.11**	**91.94**	**95.88**	**92.34**

TABLE II
COMPARISON WITH STATE-OF-THE-ART MODELS ON THE SYNAPSE MULTI-ORGAN DATASET. THE BEST RESULTS ARE BOLDED WHILE THE SECOND BEST ARE UNDERLINED.

Model	Average DSC(%)↑	Aotra DSC(%)↑	Gallbladder DSC(%)↑	Kidney(Left) DSC(%)↑	Kidney(Right) DSC(%)↑	Liver DSC(%)↑	Pancreas DSC(%)↑	Spleen DSC(%)↑	Stomach DSC(%)↑
ViT [26]+CUP [29]	67.86	70.19	45.10	74.70	67.40	91.32	42.00	81.75	70.44
R50-ViT [26]+CUP [29]	71.29	73.73	55.13	75.80	72.20	91.51	45.99	81.99	73.95
TransUNet [29]	84.36	90.68	71.99	86.04	83.71	95.54	73.96	88.80	**84.20**
SwinUNet [40]	79.13	85.47	66.53	83.28	79.61	94.29	56.58	90.66	76.60
UNETR [42]	79.56	89.99	60.56	85.66	84.80	94.46	59.25	87.81	73.99
Swin UNETR [41]	73.51	82.94	60.96	80.41	71.14	91.55	56.71	77.46	66.94
CoTr [48]	85.72	**92.96**	71.09	85.70	85.71	96.88	81.28	90.44	81.74
nnFormer [43]	85.32	90.72	71.67	85.60	87.02	96.28	**82.28**	87.30	81.69
SMAFormer(Ours)	**86.08**	92.13	**72.03**	**86.97**	**88.60**	**97.71**	81.93	**91.77**	84.15

Ablation Study

TABLE III
ABLATION STUDY OF DIFFERENT MODULES IN SMAFORMER.

SMA	E-MLP	Modulator	ISICDM2019 Average DSC ↑	LiTS2017 Average DSC ↑
✓	✗	✗	82.28%	79.95%
✗	✓	✗	80.54%	75.67%
✗	✗	✓	78.41%	73.20%
✓	✓	✗	89.53%	88.47%
✓	✗	✓	86.31%	84.26%
✓	✓	✓	96.07%	94.61%

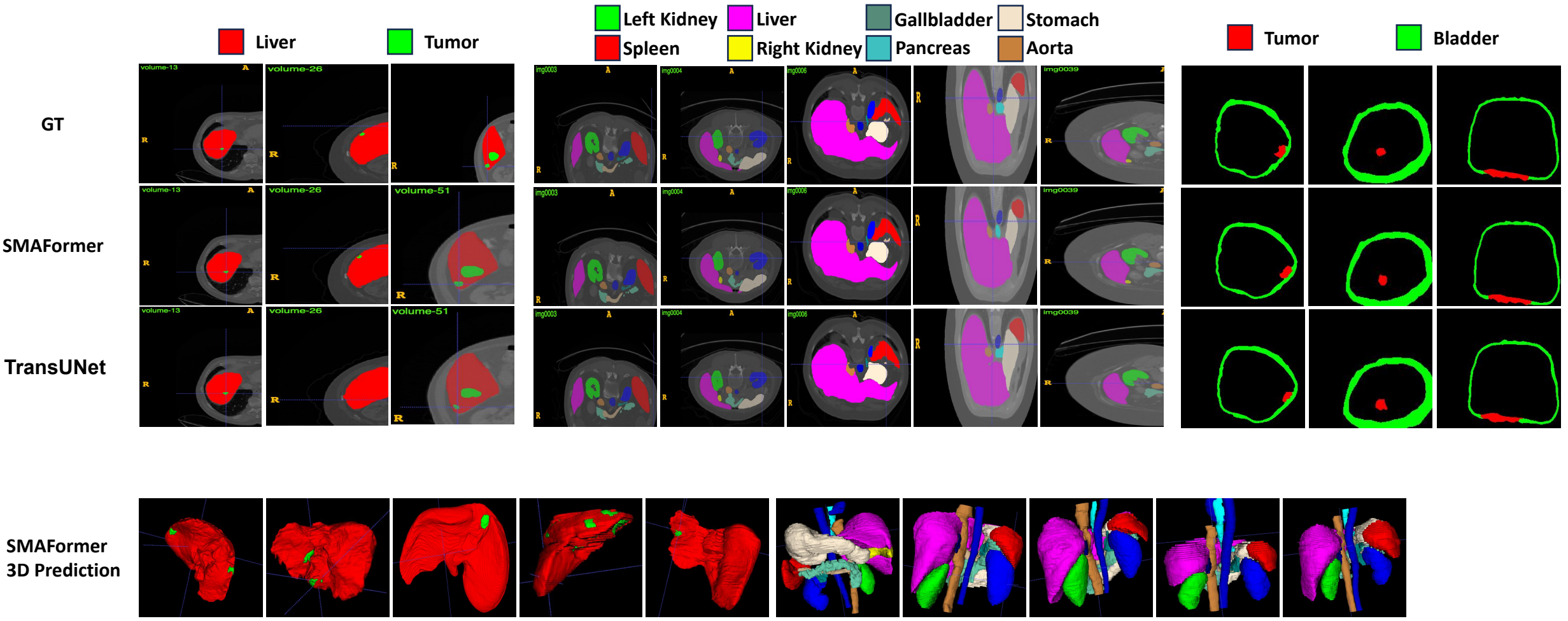
This subsection presents an ablation study to assess the impact of each component within SMAFormer. We conducted experiments on the ISICDM2019 and LiTS2017 datasets, highlighting the contribution of each component to the overall performance.

Effectiveness of SMA:
The synergistic interplay of these attention mechanisms within the SMA block allows for a more nuanced understanding of the input data, leading to improved segmentation accuracy.

Impact of E-MLP:
This outcome underscores the value of incorporating depth-wise and pixel-wise convolutions within the E-MLP. These convolutions enhance the model's ability to capture local context, crucial for accurately delineating the boundaries of small structures like tumors.

Contribution of Multi-Scale Segmentation Modulator: By embedding positional information, providing a trainable bias term, and streamlining multi-attention computations, the modulator contributes significantly to the overall efficacy of the SMAFormer model.

Visualization of Segmentation Results



DISCUSSION AND INSIGHTS

06

According to this study, possible follow-up research directions include:

1. Expand to other medical imaging modalities: Currently, this study mainly focuses on medical image segmentation tasks. SMAFormer can be further applied to other medical imaging modalities, such as MRI, CT, etc., to verify its performance on different image types.

2. Performance evaluation in more challenging clinical environments: This study conducted experiments on public medical image segmentation datasets, but these datasets may not fully represent real clinical situations. Subsequent research can apply SMAFormer to more challenging clinical datasets to evaluate its performance in actual clinical environments.

3. Improvement by combining with other deep learning techniques: Although SMAFormer performs well in medical image segmentation tasks, it is still possible to improve it by combining with other deep learning techniques.

Thank You!

Avenida da Universidade,
Taipa, Macau, China

Email : yc379501@um.edu.mo

Website :
https://github.com/lzeeorno/SMAFormer

澳門大學
UNIVERSIDADE DE MACAU
UNIVERSITY OF MACAU