

# Perspective+ Unet: Enhancing Segmentation with Bi-Path Fusion and Efficient Non-Local Attention for Superior Receptive Fields

## Supplementary Material

Jintong Hu<sup>1</sup>, Siyan Chen<sup>2</sup>, Zhiyi Pan<sup>1</sup>, Sen Zeng<sup>1</sup>, and Wenming Yang<sup>1</sup>(✉)

<sup>1</sup> Shenzhen International Graduate School, Tsinghua University

<sup>2</sup> College of Electronics and Information Engineering, Shenzhen University

**Table 1.** Detailed information of the two datasets we used.

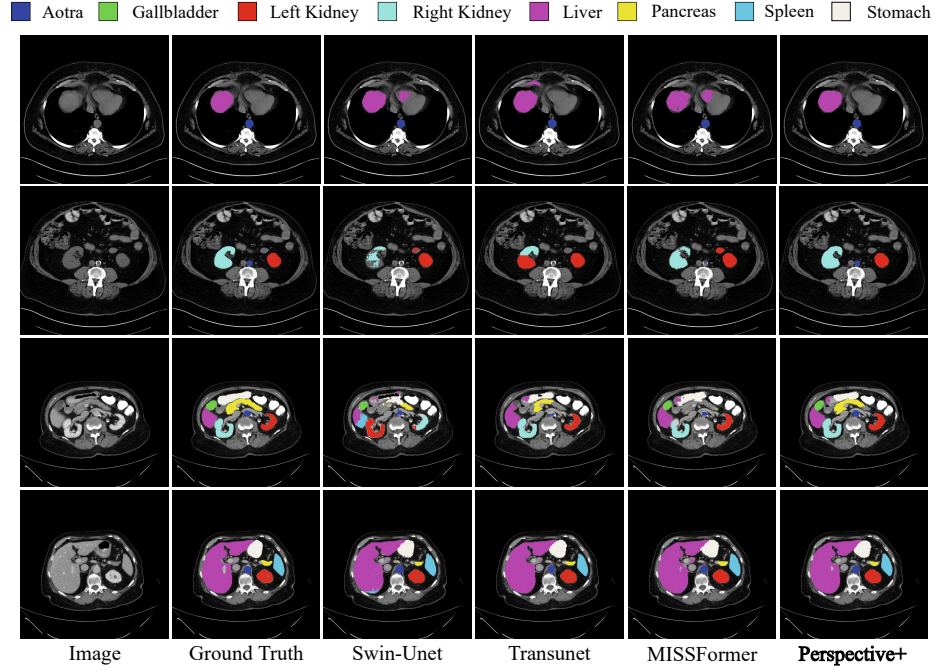
Datasets	Modality	Num. of Class	Train\Valid\Test
Synapse dataset	CT	9	18\0\12
ACDC dataset	MRI	4	70\10\20

**Table 2.** Efficiency comparison of ENLSA. -m is the number of input channel.

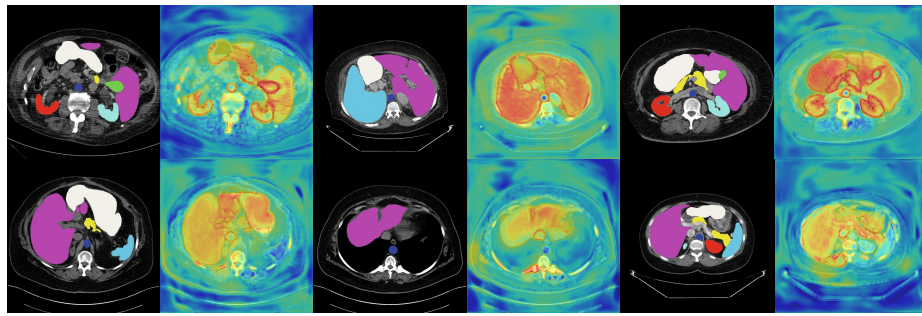
Module	Input size	FLOPs↓	Params↓
NLSA-64	[1, 64, 224, 224]	2.06G	0.04M
ENLSA-64	[1, 64, 224, 224]	<b>0.62G</b>	<b>0.01M</b>
NLSA-128	[1, 128, 112, 112]	2.06G	0.16M
ENLSA-128	[1, 128, 112, 112]	<b>0.62G</b>	<b>0.05M</b>
NLSA-256	[1, 256, 56, 56]	2.06G	0.66M
ENLSA-256	[1, 256, 56, 56]	<b>0.62G</b>	<b>0.20M</b>
NLSA-512	[1, 512, 28, 28]	2.06G	2.62M
ENLSA-512	[1, 512, 28, 28]	<b>0.62G</b>	<b>0.79M</b>
NLSA-1024	[1, 1024, 14, 14]	2.06G	10.49M
ENLSA-1024	[1, 1024, 14, 14]	<b>0.62G</b>	<b>3.15M</b>

**Table 3.** Inputs and outputs size for each stage of the BPRB and ENLTB modules.

Stage	BPRB		ENLTB	
	Input size	Output size	Input size	Output size
1	[1, 3, 224, 224]	[1, 64, 224, 224]	-	-
2	[1, 64, 224, 224]	[1, 128, 112, 112]	-	-
3	[1, 128, 112, 112]	[1, 256, 56, 56]	[1, 256, 56, 56]	[1, 256, 56, 56]
4	[1, 256, 56, 56]	[1, 512, 28, 28]	[1, 512, 28, 28]	[1, 512, 28, 28]
5	[1, 512, 28, 28]	[1, 1024, 14, 14]	[1, 1024, 14, 14]	[1, 1024, 14, 14]



**Fig. 1.** Visualized segmentation results of different methods on the Synapse multi-organ CT dataset. Our method (the last column) exhibits the smoothest boundaries and the most accurate segmentation outcomes.



**Fig. 2.** Visualization of attention heat maps from the intermediate layers of the network. Highlighting areas are closely aligned with segmentation labels, demonstrating our Perspective+ Unet's accuracy in feature identification and localization.