A Semantic Knowledge Complementarity based Decoupling Framework for Semi-supervised Class-imbalanced Medical Image Segmentation

Anonymous CVPR submission

Paper ID 3650

1. More Comparison with Existing Methods. To better demonstrate the effectiveness and robustness of our method, we compare our method with several state-of-the-art methods on different proportions of labeled data. As show in Tabs. 1 to 3, we can observe that our method achieves the best result in terms of average Dice at 10% labeled Synapse dataset, 40% labeled Synapse dataset and 2% labeled AMOS dataset. At 10% labeled AMOS dataset, as show in Tab. 4, our method still achieves the second best result on average Dice and average ASD. It is worth noting that methods such as GnericSSL [9] and DHC [8] can only achieve better results under some of these experiments, while our method has good performance under almost all experiments, which proves that our method has better stability and robustness.

2. The Effectiveness of Semantic Knowledge Complementarity Module. We introduce channel-wise cross-attention to achieve semantic knowledge complementarity, which raises the question of whether it is only the introduction of attention that leads to the performance improvement. In order to dispel this doubt, as shown in the Tab. 6, we introduce cross-attention and self-attention respectively into the model, and carry out comparative experiments. We can see that after the introduction of self-attention, the effect of the model even decreased compared to baseline, while our method show steady improvement on almost all organs, which proves the effectiveness of our semantic knowledge complementarity module.

3. Architecture Analysis of Semantic Knowledge Complementarity Module. As shown in Fig. 1, we design different architectures for the semantic knowledge complementarity module, the corresponding results are in Tab. 5. We can observe that when we adopt the structure of (a), (b) and (c), the performance of the model improves to varying degrees. However, when we adopt the structure of (d) and (e), the performance of the model decreases. This may be because, in the structure of (d), the model first adopts the global channel-wise cross-attention, destroying the precise local information of the data, and in the structure of (e), the local information and the global information do not blend well together. Our final structure enables labeled and unlabeled data to learn precise local information from each other first, and then global information, which results in an average Dice of 61.59% for the model.

4. The Motivation of Data Flow Decoupling Framework. As shown in Fig. 2, we visualize the feature maps of both the encoder and the decoder, respectively. It is observable that the encoder's features are primarily utilized for learning the texture of the image and distinguishing between the foreground and background of the image, with these features exhibiting clear distinctions from the labels. In contrast, the features extracted by the decoder possess more intricate detail, which demonstrates that, compared to the encoder, inaccurate labels exert a more pronounced negative effect on the decoder.

5. More Visualizations. More visualizations are shown in Fig. 3. From more visualization results, we can observe that our method exhibits superior segmentation performance, particularly on small organs. Furthermore, compared to other methods, our segmentation results are more stable and robust.

References

[1] Hritam Basak, Sagnik Ghosal, and Ram Sarkar. Addressing class imbalance in semi-supervised image segmentation: A study on cardiac mri. In *MICCAI'22*, pages 224–233, 2022. 2, 3

[2] Baixu Chen, Junguang Jiang, Ximei Wang, Pengfei Wan, Jianmin Wang, and Mingsheng Long. Debiased self-training for semi-supervised learning. *NeurIPS'22*, 35:32424–32437, 2022. 2, 3

[3] Hao Chen, Yue Fan, Yidong Wang, Jindong Wang, Bernt Schiele, Xing Xie, Marios Savvides, and Bhiksha Raj. An embarrassingly simple baseline for imbalanced semi-supervised learning. *arXiv preprint arXiv:2211.11086*, 2022. 2, 3

[4] Xiaokang Chen, Yuhui Yuan, Gang Zeng, and Jingdong Wang. Semi-supervised semantic segmentation with cross pseudo supervision. In *CVPR'21*, pages 2613–2622, 2021. 2, 3

	Method	Avg. Dice	Avg. ASD	Dice of Each Class Sp	RK	LK	Ga	Es	Li	St	Ao	IVC	PSV	Pa	RAG	LAG
	VNet (fully)	62.09 ± 1.2	10.28 ± 3.9	84.6	77.2	73.8	73.3	38.2	94.6	68.4	72.1	71.2	58.2	48.5	17.9	29.0
General	UA-MT[13]	18.07 ± 1.2	57.64 ± 1.8	27.1	7.1	17.0	24.4	0.0	80.6	15.6	39.3	16.7	4.4	2.7	0.0	0.0
	URPC[7]	26.37 ± 1.5	53.95 ± 11.3	51.7	35.1	26.4	7.3	0.0	83.8	21.3	69.0	41.0	1.9	5.2	0.0	0.0
	CPS[4]	21.96 ± 1.2	55.42 ± 4.6	37.9	31.8	19.0	31.9	0.0	65.1	15.5	44.8	29.6	4.3	5.5	0.0	0.0
	SS-Net[12]	17.5 ± 3.0	66.17 ± 8.0	45.6	11.6	42.3	2.4	0.0	74.5	6.0	32.6	2.8	0.0	0.0	3.8	5.8
	DST[2]	20.91 ± 5.9	61.33 ± 18.3	43.3	32.8	16.0	24.9	0.0	75.8	22.4	27.6	19.4	3.8	5.4	0.3	0.0
	DePL[10]	21.01 ± 3.3	58.42 ± 6.3	34.2	32.2	17.3	27.2	0.0	65.7	16.8	40.8	29.3	2.8	6.8	0.0	0.0
Imbalance	Adsh[5]	22.8 ± 0.9	46.18 ± 4.0	36.0	35.7	20.0	31.0	0.0	74.7	18.3	32.3	27.8	11.7	7.3	1.7	0.0
	CReST[11]	26.56 ± 2.9	36.17 ± 1.0	37.3	46.5	25.2	27.1	1.7	66.3	14.2	45.2	35.8	11.2	6.8	24.2	3.8
	SimiS[3]	25.05 ± 3.1	43.93 ± 2.4	42.0	38.6	27.2	19.7	0.0	74.2	16.5	51.7	35.0	13.6	5.4	0.0	1.8
	Basak et al.[1]	25.3 ± 2.2	50.02 ± 5.7	40.9	42.3	19.2	35.2	0.0	75.7	19.2	44.7	32.8	5.0	10.4	3.5	0.0
	CLD[6]	22.49 ± 1.6	49.74 ± 4.1	39.3	43.9	25.6	12.8	0.0	73.3	14.3	41.1	25.7	8.8	6.1	0.2	1.1
	DHC[8]	31.64 ± 0.9	21.82 ± 1.0	45.1	47.4	33.1	**36.6**	7.1	71.4	17.8	58.9	34.4	16.5	9.3	21.8	12.0
	GenericSSL[9]	46.24 ± 0.8	**7.78 ± 2.13**	**79.0**	**67.0**	58.4	25.7	12.0	86.3	**30.3**	77.8	65.5	18.2	14.2	**42.2**	**24.4**
	Ours	**48.45 ± 0.6**	7.87 ± 3.47	73.5	66.7	**64.2**	24.4	**27.2**	**90.0**	**30.3**	**79.0**	**68.5**	**33.7**	**18.1**	37.3	17.1

Table 1. Quantitative comparison between our approach and SSL segmentation methods on **10% labeled Synapse dataset**. 'General' or 'Imbalance' indicates whether the methods consider the class imbalance issue or not. Sp: spleen, RK: right kidney, LK: left kidney, Ga: gallbladder, Es: esophagus, Li: liver, St: stomach, Ao: aorta, IVC: inferior vena cava, PSV: portal & splenic veins, Pa: pancreas, RAG: right adrenal gland, LAG: left adrenal gland. Results of 3-times repeated experiments are reported in the 'mean±std' format. Best results are boldfaced, and 2nd best results are underlined.

	Method	Avg. Dice	Avg. ASD	Dice of Each Class Sp	RK	LK	Ga	Es	Li	St	Ao	IVC	PSV	Pa	RAG	LAG
	VNet (fully)	62.09 ± 1.2	10.28 ± 3.9	84.6	77.2	73.8	73.3	38.2	94.6	68.4	72.1	71.2	58.2	48.5	17.9	29.0
General	UA-MT[13]	17.09 ± 2.97	91.86 ± 7.93	4.2	57.8	32.2	0.0	0.0	91.0	37.0	0.0	0.0	0.0	0.0	0.0	0.0
	URPC[7]	24.83 ± 8.19	74.44 ± 17.01	42.6	44.8	51.4	0.0	0.0	86.7	37.8	25.8	27.0	0.0	6.6	0.0	0.0
	CPS[4]	33.07 ± 1.07	60.46 ± 2.25	68.4	72.7	64.2	0.0	0.0	91.9	42.1	66.2	22.3	0.0	2.2	0.0	0.0
	SS-Net[12]	32.98 ± 10.99	71.18 ± 20.77	49.0	68.9	71.4	22.9	0.0	92.0	34.7	51.7	38.1	0.0	0.0	0.0	0.0
	DST[2]	35.57 ± 1.54	55.69 ± 1.43	73.8	73.2	64.2	0.0	0.0	92.1	41.3	71.8	40.7	0.0	5.2	0.0	0.0
	DePL[10]	36.16 ± 2.08	56.14 ± 7.61	72.7	72.4	64.4	13.3	0.0	91.7	42.8	63.8	47.0	0.0	1.9	0.0	0.0
Imbalance	Adsh[5]	35.91 ± 6.17	53.7 ± 6.95	66.8	72.5	64.4	19.1	0.0	91.8	43.8	62.0	39.8	0.0	6.5	0.0	0.0
	CReST[11]	41.6 ± 2.49	27.82 ± 5.07	53.8	69.5	58.1	35.3	17.7	85.0	36.0	60.3	45.2	21.5	24.2	23.3	10.8
	SimiS[3]	47.09 ± 2.33	33.46 ± 1.75	75.4	66.9	69.0	62.6	0.0	81.6	53.1	80.4	56.2	29.9	37.1	0.0	0.0
	Basak et al.[1]	35.03 ± 3.68	60.69 ± 6.57	69.1	72.8	67.5	0.0	0.0	91.6	45.0	62.2	39.1	0.0	8.0	0.0	0.0
	CLD[6]	48.23 ± 1.02	28.79 ± 3.64	78.3	73.1	**73.8**	57.0	0.0	87.6	51.7	82.8	50.5	38.0	31.8	1.5	0.8
	DHC[8]	57.13 ± 0.8	11.66 ± 2.7	**82.5**	72.8	73.5	**69.8**	10.7	71.9	41.2	83.7	66.1	53.8	47.4	36.8	32.7
	GenericSSL[9]	61.63 ± 1.49	3.14 ± 1.11	75.0	69.1	65.7	52.8	66.0	88.9	53.3	81.6	74.4	48.2	50.5	46.9	28.7
	Ours	**68.01 ± 1.14**	**2.3 ± 1.61**	80.1	**74.5**	69.7	58.0	**68.0**	**94.1**	**59.8**	**86.7**	**81.8**	**58.2**	**61.0**	**47.0**	**45.2**

Table 2. Quantitative comparison between our approach and SSL segmentation methods on **40% labeled Synapse dataset**. 'General' or 'Imbalance' indicates whether the methods consider the class imbalance issue or not. Sp: spleen, RK: right kidney, LK: left kidney, Ga: gallbladder, Es: esophagus, Li: liver, St: stomach, Ao: aorta, IVC: inferior vena cava, PSV: portal & splenic veins, Pa: pancreas, RAG: right adrenal gland, LAG: left adrenal gland. Results of 3-times repeated experiments are reported in the 'mean±std' format. Best results are boldfaced, and 2nd best results are underlined.

[5] Lan-Zhe Guo and Yu-Feng Li. Class-imbalanced semi-supervised learning with adaptive thresholding. In *ICML'22*, pages 8082–8094, 2022. 2, 3

[6] Yiqun Lin, Huifeng Yao, Zezhong Li, Guoyan Zheng, and Xiaomeng Li. Calibrating label distribution for class-imbalanced barely-supervised knee segmentation. In *MICCAI'22*, pages 109–118, 2022. 2, 3

[7] Xiangde Luo, Wenjun Liao, Jieneng Chen, Tao Song, Yinan Chen, Shichuan Zhang, Nianyong Chen, Guotai Wang, and Shaoting Zhang. Efficient semi-supervised gross target volume of nasopharyngeal carcinoma segmentation via uncertainty rectified pyramid consistency. In *MICCAI'21*, pages 318–329, 2021. 2, 3

[8] Haonan Wang and Xiaomeng Li. Dhc: Dual-debiased heterogeneous co-training framework for class-imbalanced semi-supervised medical image segmentation. In *MICCAI'23*, pages 582–591, 2023. 1, 2, 3

[9] Haonan Wang and Xiaomeng Li. Towards generic semi-supervised framework for volumetric medical image segmentation. *NeurIPS'24*, 36, 2024. 1, 2, 3

[10] Xudong Wang, Zhirong Wu, Long Lian, and Stella X Yu. Debiased learning from naturally imbalanced pseudo-labels.

	Method	Avg. Dice	Avg. ASD	Sp	RK	LK	Ga	Es	Li	St	Ao	IVC	Pa	RAG	LAG	Du	Bl	P/U
	VNet (fully)	76.50	2.01	92.2	92.2	93.3	65.5	70.3	95.3	82.4	91.4	85.0	74.9	58.6	58.1	65.6	64.4	58.3
General	UA-MT[13]	33.96	22.43	**62.5**	61.7	**59.8**	17.5	13.8	73.4	39.4	34.6	32.4	26.5	12.1	6.5	15.3	32.4	21.7
General	URPC[7]	38.39	37.58	60.8	57.7	56.5	**34.6**	0.0	**78.4**	**41.4**	53.3	49.6	**40.4**	0.0	0.0	30.1	42.5	**30.6**
General	CPS[4]	31.78	39.23	55.9	46.9	53.1	27.7	0.0	66.4	25.2	41.8	45.2	29.4	0.1	0.0	22.1	38.7	24.2
General	SS-Net[12]	17.47	59.05	37.7	20.1	26.3	9.0	3.3	57.1	25.1	28.4	28.2	0.0	0.0	0.0	0.0	26.5	0.2
General	DST[2]	31.94	39.15	50.9	52.4	56.9	24.6	0.0	59.4	31.5	41.8	43.1	26.2	0.0	0.0	23.8	42.6	25.9
General	DePL[10]	31.56	40.70	57.1	49.3	54.3	26.6	0.1	69.2	26.2	41.1	46.7	23.9	0.0	0.0	16.7	40.3	21.8
Imbalance	Adsh[5]	30.30	42.48	53.9	45.1	51.2	28.5	0.0	62.1	27.0	41.4	42.7	25.0	0.0	0.0	20.3	35.8	21.6
Imbalance	CReST[11]	34.13	20.15	57.9	51.5	49.1	22.7	13.2	66.2	34.4	39.4	40.4	24.6	17.2	10.2	24.4	36.5	24.4
Imbalance	SimiS[3]	36.89	26.16	57.8	58.6	58.6	22.9	0.0	70.9	38.0	52.0	47.0	32.4	20.2	11.5	18.1	39.9	25.5
Imbalance	Basak et al.[1]	29.87	35.55	50.7	47.7	44.1	21.1	0.0	61.8	27.7	38.1	40.4	21.8	9.6	9.5	14.6	36.5	24.5
Imbalance	CLD[6]	36.23	27.63	55.2	55.8	59.1	23.9	0.0	69.9	38.2	50.1	44.5	32.3	18.9	9.2	18.8	42.2	24.9
Imbalance	DHC[8]	38.28	20.34	62.1	59.5	57.8	25.0	20.5	66.0	38.2	51.3	47.9	26.8	25.4	7.0	17.8	**43.2**	24.8
Imbalance	GenericSSL[9]	34.56	21.05	44.1	55.0	49.3	22.5	18.8	57.0	36.9	44.1	47.2	33.7	16.1	12.1	**32.5**	29.6	19.4
Imbalance	Ours	**41.80**	**16.52**	62.0	**65.5**	**59.8**	26.9	**25.9**	70.4	38.5	**59.0**	**51.6**	36.7	**27.1**	**15.4**	23.7	41.1	23.4

Table 3. Quantitative comparison between our approach and SSL segmentation methods on **2% labeled AMOS dataset**. Sp: spleen, RK: right kidney, LK: left kidney, Ga: gallbladder, Es: esophagus, Li: liver, St: stomach, Ao: aorta, IVC: inferior vena cava, Pa: pancreas, RAG: right adrenal gland, LAG: left adrenal gland, Du: duodenum, Bl: bladder, P/U: prostate/uterus.

	Method	Avg. Dice	Avg. ASD	Sp	RK	LK	Ga	Es	Li	St	Ao	IVC	Pa	RAG	LAG	Du	Bl	P/U
	VNet (fully)	76.50	2.01	92.2	92.2	93.3	65.5	70.3	95.3	82.4	91.4	85.0	74.9	58.6	58.1	65.6	64.4	58.3
General	UA-MT[13]	40.60	38.45	61.0	75.4	58.8	0.1	0.0	84.4	45.2	72.8	61.6	36.2	0.0	0.0	30.7	46.5	36.3
General	URPC[7]	49.09	29.69	81.7	77.5	77.2	38.1	0.0	87.7	57.9	75.0	62.7	52.1	0.0	0.0	35.9	48.8	41.7
General	CPS[4]	54.51	7.84	80.7	79.8	74.3	35.2	44.4	90.5	51.1	76.1	65.6	48.6	31.6	21.8	33.6	47.0	37.3
General	SS-Net[12]	38.91	53.43	73.4	73.4	72.2	42.4	0.0	83.5	46.7	74.1	69.6	0.0	0.0	0.0	0.0	48.3	0.2
General	DST[2]	52.24	17.66	81.7	80.2	78.6	39.5	41.0	89.8	52.8	78.5	65.9	51.1	4.3	0.1	34.2	48.8	37.2
General	DePL[10]	56.76	6.70	81.9	80.6	79.5	41.0	42.6	89.3	57.6	79.1	66.0	53.2	34.6	21.8	34.9	48.4	40.9
Imbalance	Adsh[5]	54.92	8.07	81.6	78.5	76.6	40.1	43.4	90.1	53.0	76.7	64.4	48.3	25.9	24.2	34.7	48.7	37.7
Imbalance	CReST[11]	60.74	4.65	85.3	84.5	84.0	43.2	50.8	89.9	58.7	84.7	73.0	54.2	41.8	31.6	41.0	52.8	35.8
Imbalance	SimiS[3]	57.48	4.46	83.1	80.9	80.0	39.6	45.9	90.0	57.1	78.0	66.3	54.1	35.8	26.9	39.9	49.3	35.4
Imbalance	Basak et al.[1]	53.66	8.50	80.3	78.2	79.0	36.3	40.3	88.6	53.2	76.8	65.6	46.8	23.9	16.1	31.4	49.7	38.6
Imbalance	CLD[6]	61.55	4.21	86.0	85.3	84.8	44.5	51.9	90.8	59.7	83.7	73.1	55.7	40.2	37.2	41.4	53.0	36.1
Imbalance	DHC[8]	**64.16**	3.51	87.4	**86.6**	87.1	45.8	**57.0**	89.8	64.7	86.0	**75.0**	**62.5**	39.8	36.8	44.0	**56.5**	**43.6**
Imbalance	GenericSSL[9]	60.28	**3.11**	79.8	84.4	85.7	39.6	55.8	86.3	61.2	80.0	70.2	60.6	39.8	**40.4**	**48.0**	40.6	31.9
Imbalance	Ours	63.58	3.36	**87.5**	86.2	**87.3**	**47.1**	54.8	**91.9**	**64.9**	**86.1**	74.8	59.2	**42.7**	40.2	43.1	49.2	38.6

Table 4. Quantitative comparison between our approach and SSL segmentation methods on **10% labeled AMOS dataset**. Sp: spleen, RK: right kidney, LK: left kidney, Ga: gallbladder, Es: esophagus, Li: liver, St: stomach, Ao: aorta, IVC: inferior vena cava, Pa: pancreas, RAG: right adrenal gland, LAG: left adrenal gland, Du: duodenum, Bl: bladder, P/U: prostate/uterus.

Architecture	a	b	c	d	e
Avg. Dice	**61.59**	58.91	59.39	55.48	56.10
Avg. ASD	**10.98**	11.48	11.20	13.47	12.54

Table 5. Ablation study on the **20% labeled Synapse dataset** of different channel-wise cross-attention architectures with respect to those in Fig. 1 .

In *CVPR'22*, pages 14647–14657, 2022. 2, 3

[11] Chen Wei, Kihyuk Sohn, Clayton Mellina, Alan Yuille, and Fan Yang. Crest: A class-rebalancing self-training framework for imbalanced semi-supervised learning. In *CVPR'21*, pages 10857–10866, 2021. 2, 3

[12] Yicheng Wu, Zhonghua Wu, Qianyi Wu, Zongyuan Ge, and Jianfei Cai. Exploring smoothness and class-separation for semi-supervised medical image segmentation. In *MICCAI'22*, pages 34–43, 2022. 2, 3

[13] Lequan Yu, Shujun Wang, Xiaomeng Li, Chi-Wing Fu, and Pheng-Ann Heng. Uncertainty-aware self-ensembling model for semi-supervised 3d left atrium segmentation. In *MICCAI'19*, pages 605–613, 2019. 2, 3

CVPR #3650

Method	Avg. Dice	Avg. ASD	Dice of Each Class												
			Sp	RK	LK	Ga	Es	Li	St	Ao	IVC	PSV	Pa	RAG	LAG
Baseline	58.68	11.55	79.0	71.4	69.2	52.6	0.0	86.1	**67.6**	83.8	**80.5**	42.8	50.8	46.8	32.0
SA	58.22	11.44	**80.5**	66.7	61.5	**54.2**	0.0	**92.8**	59.4	84.5	79.1	**50.3**	40.8	**50.7**	36.6
CA	**59.39**	**11.20**	**80.5**	**72.9**	**70.2**	40.1	0.0	92.6	65.2	**84.9**	77.7	47.7	**53.4**	47.3	**39.3**

Table 6. Quantitative comparison between channel-wise cross-attention and channel-wise self-attention on **20% labeled Synapse dataset**. SA: channel-wise self-attention, CA: channel-wise cross-attention.

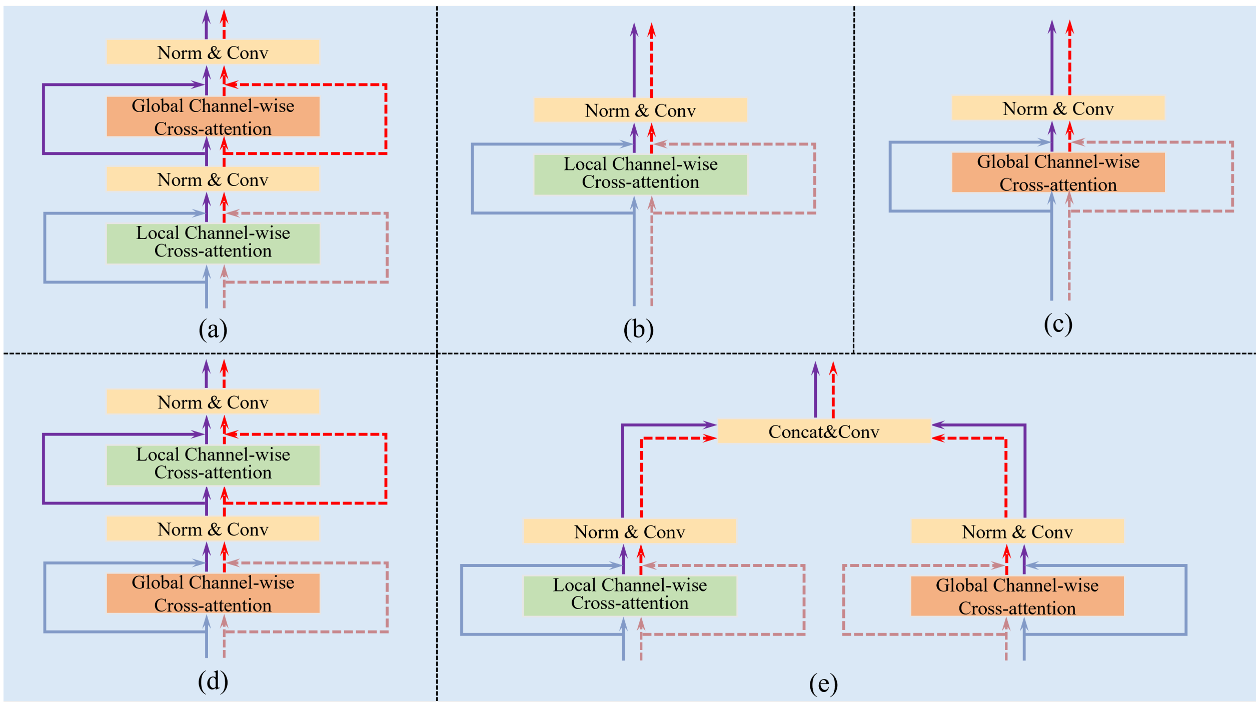

Figure 1. Ablation study on different channel-wise cross-attention architectures. (a) is the final architecture.

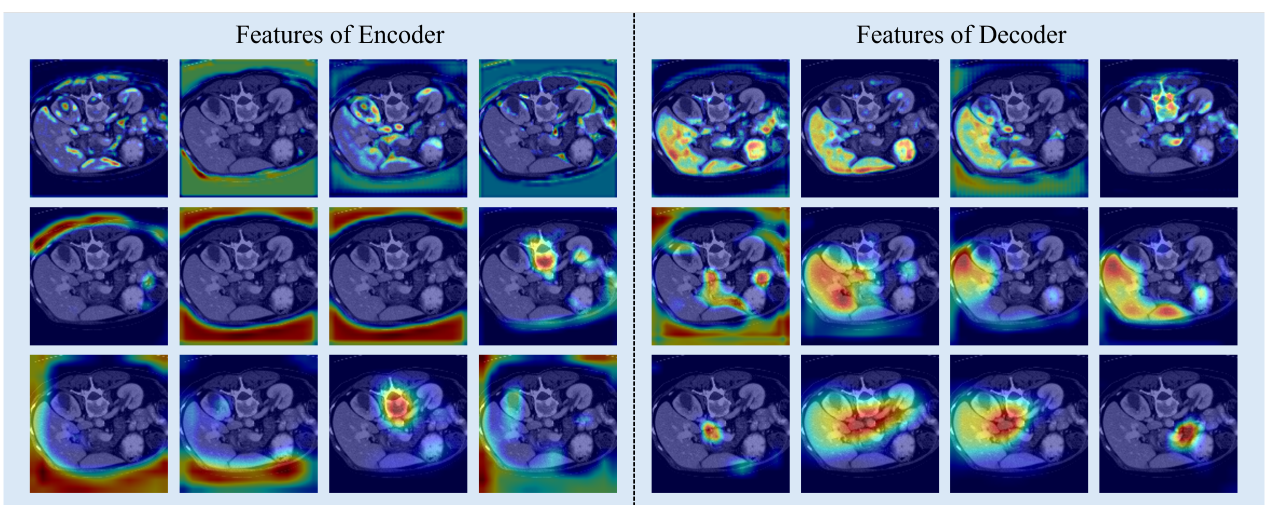

Figure 2. The visualization of feature maps. On the left are the features of the encoder, and on the right are the features of the decoder. Feature maps at the same location have the same channel index.

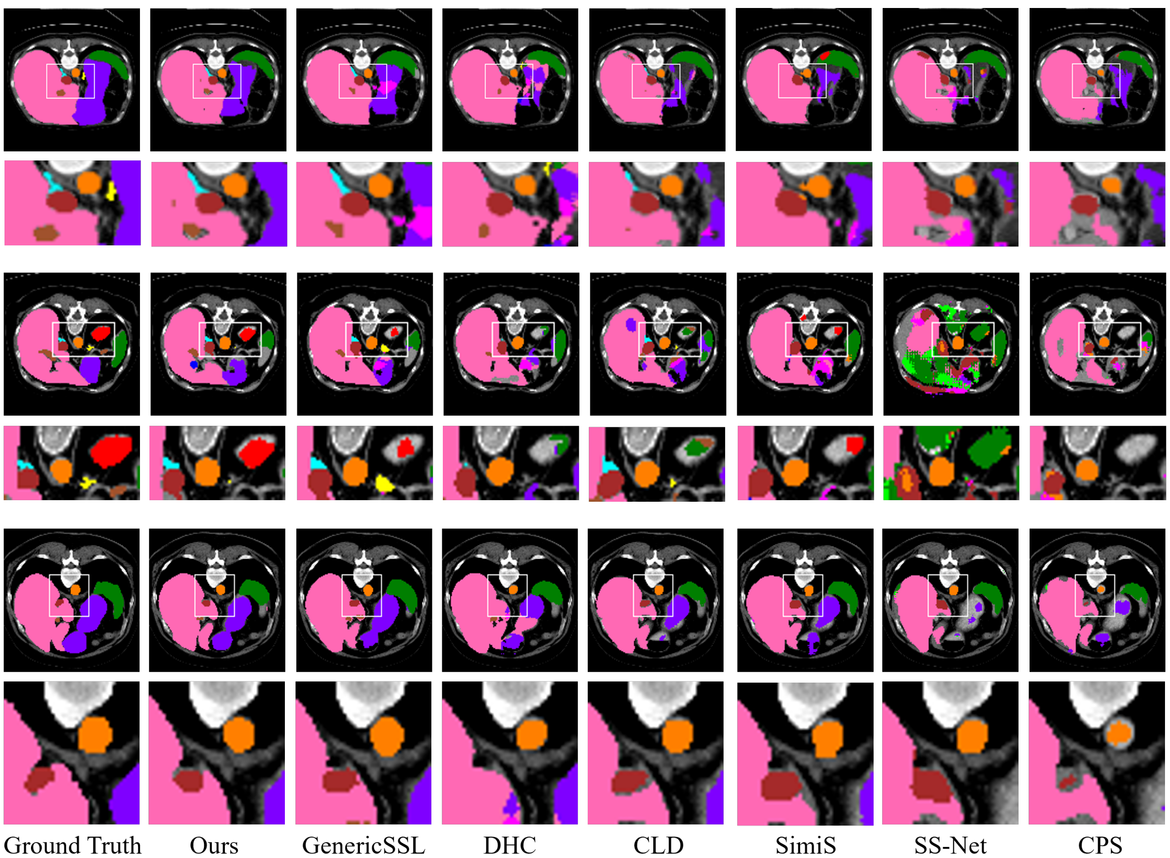

Figure 3. Visual comparison on the 20% labeled Synapse dataset: ■ spleen, ■ right kidney, ■ left kidney, ■ gallbladder, ■ esophagus, ■ liver, ■ stomach, ■ aorta, ■ inferior vena cava, ■ protal & splenic veins, ■ pancreas, ■ right adrenal gland, and ■ left adrenal gland.