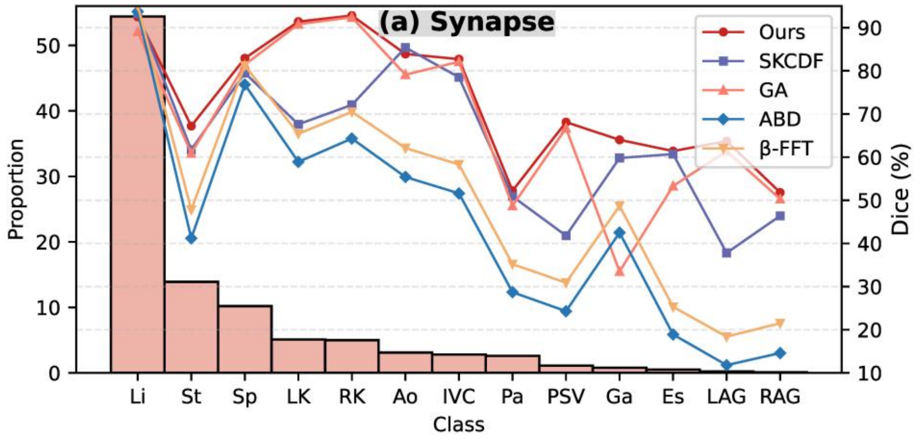
上海交通大學
SHANGHAI JIAO TONG UNIVERSITY

CVPR JUNE 3-7, 2026 DENVER COLORADO

Divide, Conquer, and Aggregate: Asymmetric Experts for Class-Imbalanced Semi-Supervised Medical Image Segmentation

YAJUN LIU
liuyajun@sjtu.edu.cn
SCHOOL OF INTEGRATED CIRCUITS, SHANGHAI JIAO TONG UNIVERSITY
SHANGHAI, CHINA

Existing SSL Methods



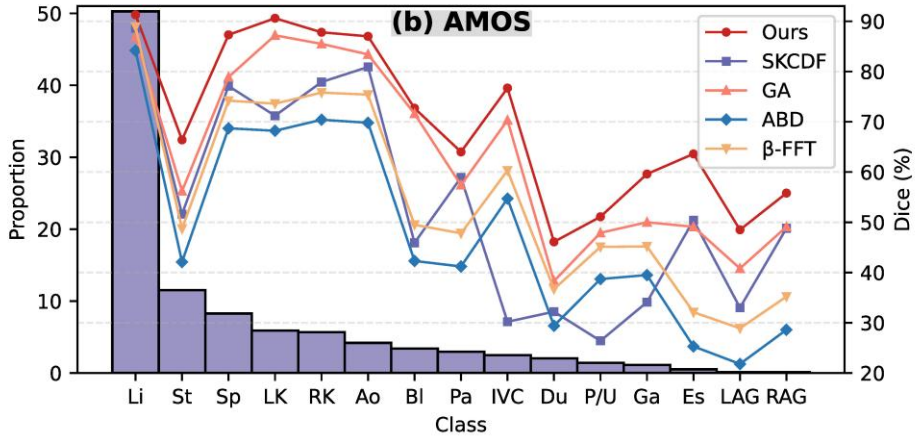


Fig. 1 Class-wise occupancy rankings and Dice comparisons on **Synapse** (**20%**) and **AMOS** (**5%**) datasets.

- Problem 1: Multi-organ datasets are highly class-imbalanced: large organs dominate voxel occupancy.
- Problem 2: Single-decoder SSMIS models are easily biased toward majority organs.

Motivation

- **Finding:** Despite minor ranking fluctuations in Synapse and AMOS dataset, class stratification remains stable and aligns with anatomical priors.

- **Idea:** Organs are stratified into three classes, with three category-specific decoders designed to address their respective segmentation tasks.

- **Motivation:** Stable organ-size hierarchy naturally supports a divide-and-conquer design.

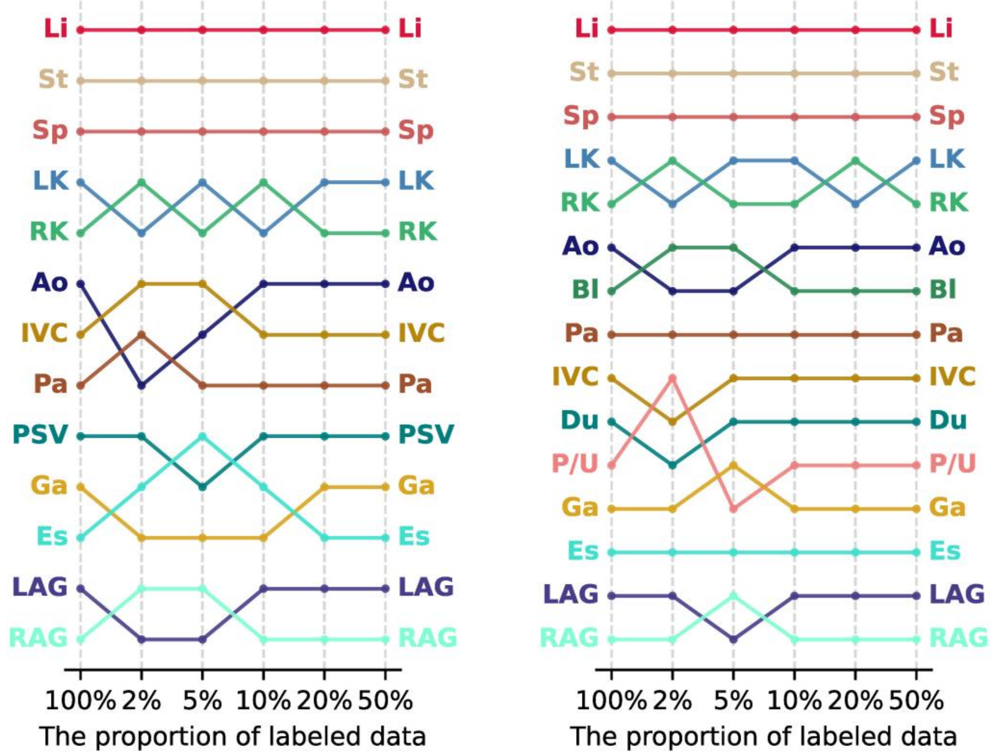


Fig. 2 Class proportion rankings in labeled subsets across varying labeling ratios for **Synapse** and **AMOS** datasets.

Contributions

Divide → Conquer → Aggregate

- Divide:
  - Log-Gap Analysis partitions foreground organs into Head, Medium, and Tail sets.
- Conque:
  - Three asymmetric expert decoders specialize on different organ groups using label-split supervision.
- Aggregate
  - Logit stitching produces expert-fused pseudo-labels, while DFAM dynamically aggregates expert features.

Instead of forcing one decoder to handle all scales, DCA assigns scale-aware experts and aggregates their complementary strengths.

Method Overview: DCA Divide → Conquer → Aggregate

CVPR JUNE 3-7, 2026 DENVER COLORADO

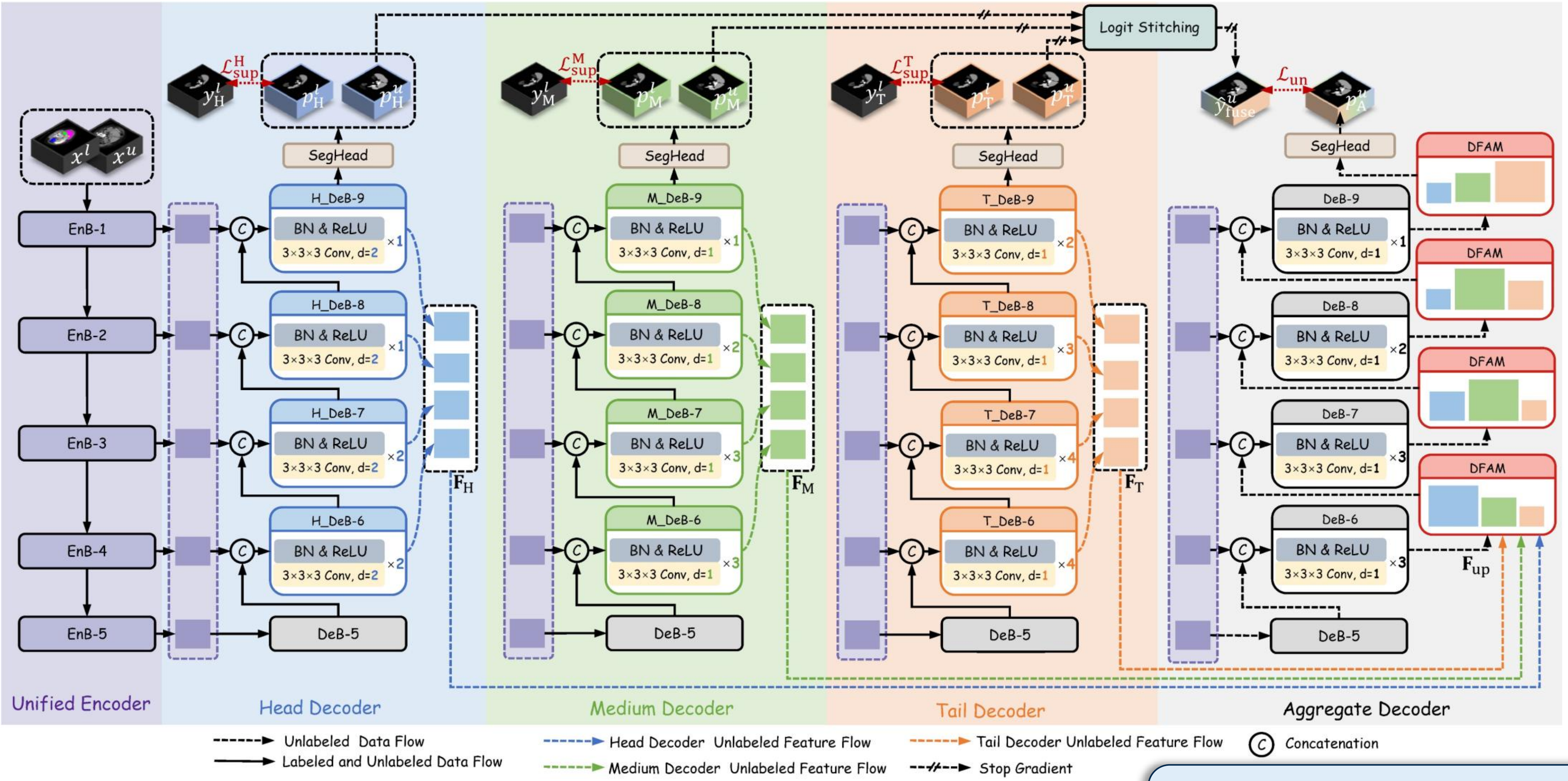


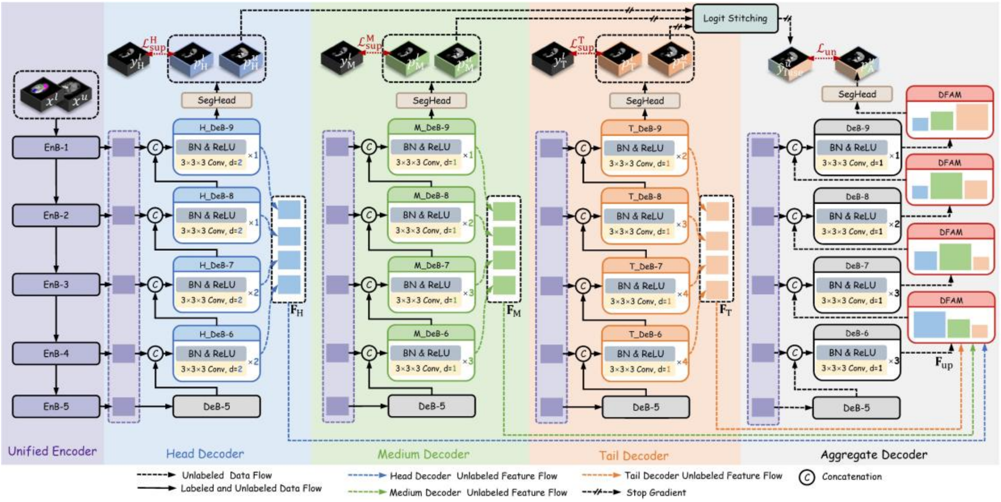
Fig. 3 Overview of Our DCA Framework.

- **Divide:** Log-Gap Analysis partitions foreground classes into Head / Medium / Tail.
- **Conquer:** Three asymmetric expert decoders specialize in different organ groups.
- **Aggregate:** DFAM dynamically fuses expert features and produces final segmentation.

DCA decomposes class-imbalanced segmentation into expert specialization and adaptive aggregation.

DCA: Divide — Log-Gap Analysis

Divide : Data-Driven Head / Medium / Tail Partition



1. Voxel Count:

$$V_k = \sum_{i=1}^{N_L} \sum_{\mathbf{p}=(h,w,d)}^{(H,W,D)} \mathbb{I}(y_i^l(\mathbf{p}) = k)$$

Count foreground voxels for each organ class.

2. Foreground Proportion:

$$P_k = \frac{V_k}{V_{fg}}, V_{fg} = \sum_{k=1}^{K} V_k$$

Normalize by total foreground voxels to avoid background dominance.

3. Log-Gap:

$$G_j = \log(P_{(j)}) - \log(P_{(j+1)}) = \log\left(\frac{P_{(j)}}{P_{(j+1)}}\right)$$

Large log gaps reveal natural cliffs between organ-size groups.

Log-Gap Analysis gives a parameter-free, anatomy-aware class partition.

4. Breakpoints:

$$k_{\mathrm{HM}} = \arg \max_{j\in\{1,\ldots,K-1\}} (G_j)$$
$$k_{\mathrm{MT}} = \arg \max_{j\in\{k_{\mathrm{HM}}+1,\ldots,K-1\}} (G_j)$$

5. Sets:

$$S_{\mathrm{H}} = \{c_{(j)} | 1 \le j \le k_{\mathrm{HM}}\}$$
$$S_{\mathrm{M}} = \{c_{(j)} | k_{\mathrm{HM}} + 1 \le j \le k_{\mathrm{MT}}\}$$
$$S_{\mathrm{T}} = \{c_{(j)} | k_{\mathrm{MT}} + 1 \le j \le K\}$$

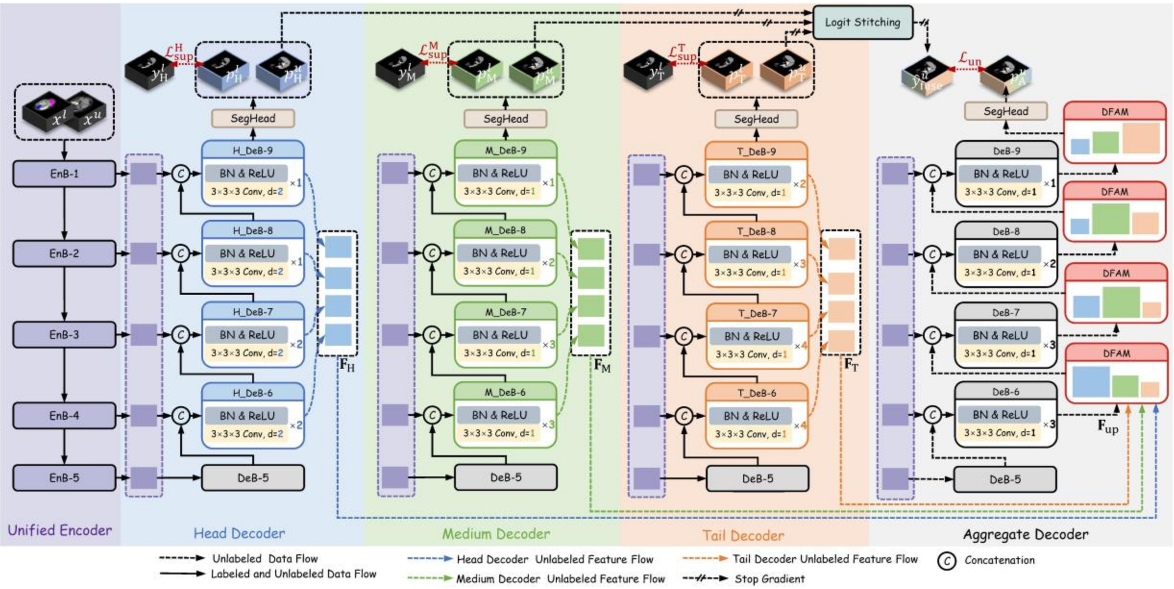
DCA: Conquer — Three Asymmetric Expert Decoders

Conquer: Scale-Aware Expert Decoders

Head Expert
- $\mathcal{D}_{\mathrm{H}}$, `dilation = 2,   n_stages = {2,2,1,1}`
- Large organs: larger receptive field, fewer layers.

Medium Expert
- $\mathcal{D}_{\mathrm{M}}$, `dilation = 1,   n_stages = {3,3,2,1}`
- Medium organs: standard V-Net decoder.

Tail Expert
- $\mathcal{D}_{\mathrm{T}}$, `dilation = 1,   n_stages = {4,4,3,2}`
- Small organs: deeper decoder for fine details.



Label-split:

$$y_i^l(\mathbf{p}) = \begin{cases} y^l(\mathbf{p}) & \text{if } y^l(\mathbf{p}) \in S_{\mathrm{H}} \\ 0 & \text{otherwise} \end{cases}$$

Each expert only learns its own organ group; all other organs are remapped to background.

Supervised loss:

$$\mathcal{L}_{\mathrm{sup}}^{\mathrm{H}} = \mathcal{L}_{\mathrm{seg}}\left(p_{\mathrm{H}}^l, y_{\mathrm{H}}^l\right),$$
$$\mathcal{L}_{\mathrm{sup}}^{\mathrm{M}} = \mathcal{L}_{\mathrm{seg}}\left(p_{\mathrm{M}}^l, y_{\mathrm{M}}^l\right),$$
$$\mathcal{L}_{\mathrm{sup}}^{\mathrm{T}} = \mathcal{L}_{\mathrm{seg}}\left(p_{\mathrm{T}}^l, y_{\mathrm{T}}^l\right)$$

Expert specialization reduces the burden of one decoder handling all organ scales.

DCA: Aggregate — Logit Stitching for Pseudo-Labels

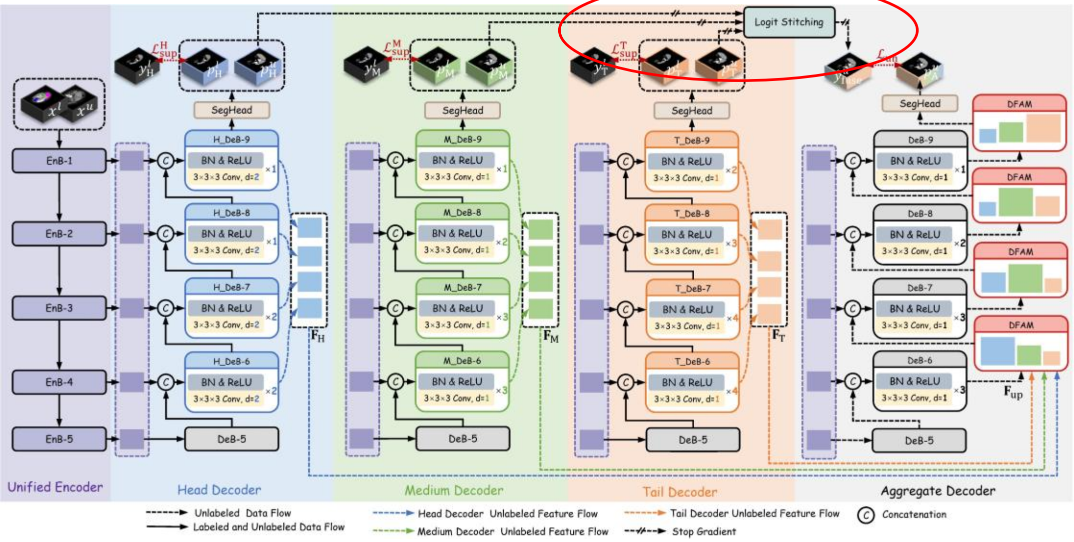
Aggregate: Expert-Fused Pseudo-Labels by Logit Stitching

Foreground Stitching:

- $p^u_{\text{fuse}}(\mathbf{p})[c] = \begin{cases} p^u_{\text{H}}(\mathbf{p})[c] & \text{if } c \in S_{\text{H}} \\ p^u_{\text{M}}(\mathbf{p})[c] & \text{if } c \in S_{\text{M}} \\ p^u_{\text{T}}(\mathbf{p})[c] & \text{if } c \in S_{\text{T}} \end{cases}$
- Each foreground class takes logits only from its responsible expert.

Background Stitching:

- $p^u_{\text{fuse}}(\mathbf{p})[0] = \frac{1}{3}(p^u_{\text{H}}(\mathbf{p})[0] + p^u_{\text{H}}(\mathbf{p})[0] + p^u_{\text{H}}(\mathbf{p})[0])$
- Background is shared by all experts, so we average their background logits.



Pseudo-label:

- $\hat{y}^u_{\text{fuse}} = \arg\max(p^u_{\text{fuse}})$

Logit stitching avoids conflicting softmax averaging and preserves expert specialization.

DCA: Aggregate — DFAM and Final Objective

CVPR JUNE 3-7, 2026 DENVER COLORADO

Aggregate: Dynamic Feature Aggregation Module

Expert Feature Concatenation:

- $\mathbf{F}_{\text{con}} = \text{Concat}(\underset{1\times1\times1}{\text{Conv}}(\mathbf{F}_{\text{H}}), \underset{1\times1\times1}{\text{Conv}}(\mathbf{F}_{\text{M}}), \underset{1\times1\times1}{\text{Conv}}(\mathbf{F}_{\text{T}}))$
- Unify and concatenate same-level expert features.

Attention Maps:

- $\{\mathbf{A}_{\text{H}}, \mathbf{A}_{\text{M}}, \mathbf{A}_{\text{T}}\} = \text{Softmax}\left(\underset{1\times1\times1}{\text{Conv}}\left(\text{Relu}(\underset{3\times3\times3}{\text{Conv}}(\mathbf{F}_{\text{con}}))\right)\right)$
- $\mathbf{A}_{\text{H}}(\mathbf{p}) + \mathbf{A}_{\text{M}}(\mathbf{p}) + \mathbf{A}_{\text{T}}(\mathbf{p}) = \mathbf{1}$
- Generate voxel-wise softmax attention over three experts.

Feature Aggregation:

- $\mathbf{F}_{\text{expert}} = \mathbf{A}_{\text{H}} \otimes \mathbf{F}_{\text{H}} + \mathbf{A}_{\text{M}} \otimes \mathbf{F}_{\text{M}} + \mathbf{A}_{\text{T}} \otimes \mathbf{F}_{\text{T}}$

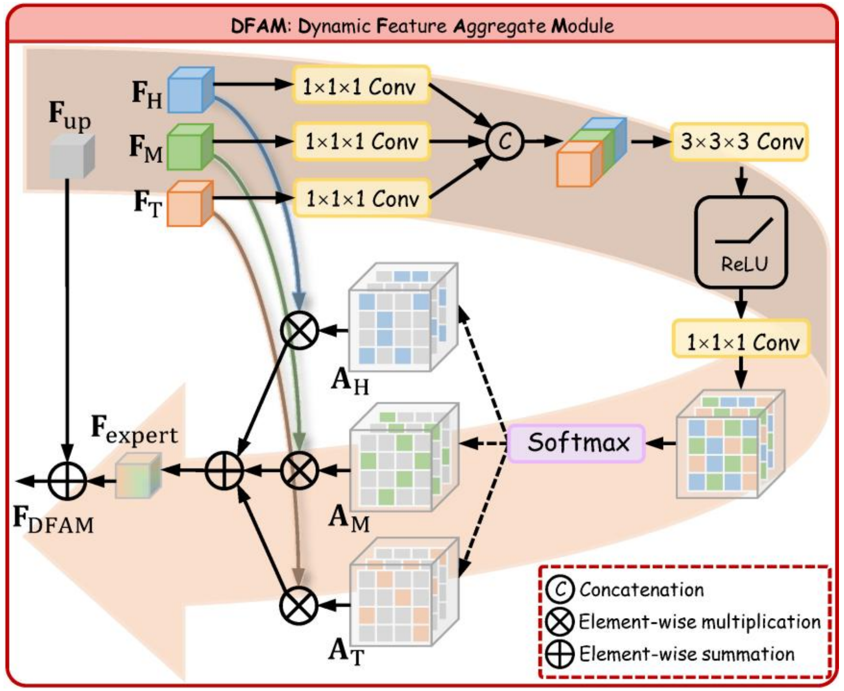


Fig. 4 DFAM: Dynamic Feature Aggregate Module.

Residual Fusion:

- $\mathbf{F}_{\text{DFAM}} = \mathbf{F}_{\text{expert}} + \mathbf{F}_{\text{up}}$

Unsupervised Loss:

- $\mathcal{L}_{\text{un}} = \mathcal{L}_{\text{seg}}(p^u_{\text{A}}, \hat{y}^u_{\text{fuse}})$

Total Objective :

- $\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{seg}} + \lambda\, \mathcal{L}_{\text{un}}$

DFAM adaptively selects expert priors at each voxel and produces unbiased final predictions.

Experiments — Synapse

	Methods	Avg. Dice	Avg. ASD	Average Dice of Each Class Sp	RK	LK	Ga	Es	Li	St	Ao	IVC	PSV	Pa	RAG	LAG
	V-Net (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 [41]	20.26±2.2	71.67±7.4	48.2	31.7	22.2	0.0	0.0	81.2	29.1	23.3	27.5	0.0	0.0	0.0	0.0
	URPC [23]	25.68±5.1	72.74±15.5	66.7	38.2	56.8	0.0	0.0	85.3	33.9	33.1	14.8	0.0	5.1	0.0	0.0
	CPS [8]	33.55±3.7	41.21±9.1	62.8	55.2	45.4	35.9	0.0	91.1	31.3	41.9	49.2	8.8	14.5	0.0	0.0
	SS-Net [38]	35.08±2.8	50.81±6.5	62.7	67.9	60.9	34.3	0.0	89.9	20.9	61.7	44.8	0.0	8.7	4.2	0.0
	Co-BioNet [25]	58.83±2.7	7.50±5.8	82.8	90.0	86.5	11.6	19.5	92.3	47.7	77.5	77.7	51.3	30.3	47.5	50.2
	BCP [1]	50.23±1.5	18.45±2.3	72.4	58.1	52.7	38.2	12.5	92.3	35.6	48.9	45.2	18.7	22.1	8.4	5.2
	MagicNet [5]	60.57±2.5	22.48±6.3	82.5	91.0	89.5	11.2	0.0	89.4	62.7	77.6	79.0	66.1	47.3	36.8	54.3
	ABD [9]	55.67±1.2	15.32±1.8	76.8	64.3	58.9	42.5	18.9	93.7	41.2	55.4	51.6	24.3	28.7	14.6	11.8
	β-FFT [14]	59.12±0.9	12.18±1.1	81.2	70.5	65.4	48.7	25.3	95.1	47.8	62.1	58.3	30.9	35.2	21.5	18.4
Imbalance	Adsh [12]	35.29±0.5	39.61±4.6	55.1	59.6	45.8	52.2	0.0	89.4	32.8	47.6	53.0	8.9	14.4	0.0	0.0
	CReST [34]	38.33±3.4	22.85±9.0	62.1	64.7	53.8	43.8	8.1	85.9	27.2	54.4	47.7	14.4	13.0	18.7	4.6
	SimiS [6]	40.07±0.6	32.98±0.5	62.3	69.4	50.7	61.4	0.0	87.0	33.0	59.0	57.2	29.2	11.8	0.0	0.0
	Basak *et al.* [3]	33.24±0.6	43.78±2.5	57.4	53.8	48.5	46.9	0.0	87.8	28.7	42.3	45.4	6.3	15.0	0.0	0.0
	CLD [21]	41.07±1.2	32.15±3.3	62.0	66.0	59.3	61.5	0.0	89.0	31.7	62.8	49.4	28.6	18.5	0.0	5.0
	DHC [32]	48.61±0.9	10.71±2.6	62.8	69.5	59.2	66.0	13.2	85.2	36.9	67.9	61.5	37.0	30.9	31.4	10.6
	A&D [33]	60.88±0.7	2.52±0.4	85.2	66.9	67.0	52.7	62.9	89.6	52.1	83.0	74.9	41.8	43.4	44.8	27.2
	GA [26]	68.43±0.5	3.11±0.2	81.4	92.4	90.8	33.5	53.3	89.1	60.9	79.1	82.1	66.7	48.7	50.3	61.4
	SKCDF [43]	64.27±1.4	1.45±0.1	79.5	72.1	67.6	59.8	60.7	93.3	61.7	85.4	78.5	41.8	50.9	46.4	37.8
	DCA (Ours)	**73.20±0.3**	**1.78±0.1**	**82.9**	**92.8**	**91.4**	**64.0**	**61.4**	**92.5**	**67.2**	**83.9**	**82.7**	**68.1**	**52.2**	**51.8**	**63.7**

Tab. 1 Quantitative results on **20%** labeled **Synapse** dataset. 'General' or 'Imbalance' indicate whether the methods consider class-imbalance issue or not. The top-3 results are highlighted as first, second, and third.

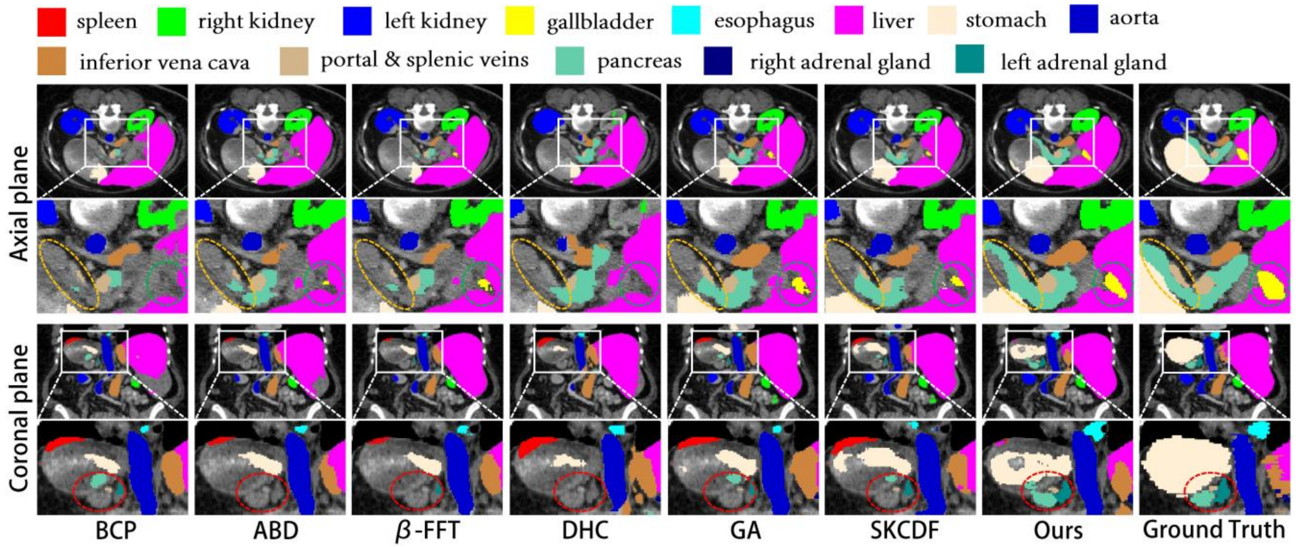


Fig. 5 Qualitative comparison between our method and the SOTA methods on **20%** labeled **Synapse** dataset.

Qualitative results show fewer under-segmentation errors and clearer small-organ boundaries.

Experiments — AMOS

	Methods	Avg. Dice	Avg. ASD	Sp	RK	LK	Ga	Es	Li	St	Ao	IVC	Pa	RAG	LAG	Du	Bl	P/U
	V-Net (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 [41]	42.16	15.48	59.8	64.9	64.0	35.3	34.1	77.7	37.8	61.0	46.0	33.3	26.9	12.3	18.1	29.7	31.6
	URPC [23]	44.93	27.44	67.0	64.2	67.2	36.1	0.0	83.1	45.5	67.4	54.4	46.7	0.0	29.4	35.2	44.5	33.2
	CPS [8]	41.08	20.37	56.1	60.3	59.4	33.3	25.4	73.8	32.4	65.7	52.1	31.1	25.5	6.2	18.4	40.7	35.8
	SS-Net [38]	33.88	54.72	65.4	68.3	69.9	37.8	0.0	75.1	33.2	68.0	56.6	33.5	0.0	0.0	0.0	0.2	0.2
	Co-BioNet [25]	48.32	26.04	76.6	82.1	75.1	41.5	38.2	87.9	40.4	75.2	53.7	40.8	4.8	0.0	25.1	64.2	19.2
	BCP [1]	45.67	28.12	62.3	66.1	62.5	34.8	18.7	79.4	38.2	64.3	49.1	35.6	22.4	15.2	22.8	35.1	32.9
	MagicNet [5]	54.08	29.03	80.0	84.5	86.1	47.9	0.0	85.1	50.7	81.7	69.3	57.2	46.0	0.0	40.8	62.9	19.2
	ABD [9]	50.23	22.45	68.7	70.4	68.2	39.5	25.3	84.2	42.1	69.8	54.7	41.2	28.6	21.8	29.4	42.3	38.7
	β-FFT [14]	54.89	18.76	74.2	75.8	73.6	45.2	32.1	88.9	48.7	75.4	60.3	47.8	35.2	28.9	36.7	49.5	45.1
Imbalance	Adsh [12]	40.33	24.53	56.0	63.6	57.3	34.7	25.7	73.9	30.7	65.7	51.9	27.1	20.2	0.0	18.6	43.5	35.9
	CReST [34]	46.55	14.62	66.5	64.2	65.4	36.0	32.2	77.8	43.6	68.5	52.9	40.3	24.7	19.5	26.5	43.9	36.4
	SimiS [6]	47.27	11.51	77.4	72.5	68.7	32.1	14.7	86.6	46.3	74.6	54.2	41.6	24.4	17.9	21.9	47.9	28.2
	Basak *et al.* [3]	38.74	31.76	68.8	59.0	54.2	29.0	0.0	83.7	39.3	61.7	52.1	34.6	0.0	0.0	26.8	45.7	26.2
	CLD [21]	46.10	15.86	67.2	68.5	71.4	41.0	21.0	76.1	42.4	69.8	52.1	37.9	24.7	23.4	22.7	38.1	35.2
	DHC [32]	49.53	13.89	68.1	69.6	71.1	42.3	37.0	76.8	43.8	70.8	57.4	43.2	27.0	28.7	29.1	41.4	36.7
	A&D [33]	50.03	5.21	73.1	76.0	76.5	29.1	44.9	82.5	49.0	72.8	61.7	48.5	30.2	19.7	36.4	32.9	18.2
	GA [26]	63.51	4.58	78.9	85.5	87.2	50.0	49.1	86.9	56.2	83.4	70.3	57.4	49.1	40.8	38.3	71.6	47.9
	SKCDF [43]	53.81	5.97	77.1	77.9	71.2	34.1	50.4	88.6	51.6	80.9	58.9	48.8	33.0	30.2	32.2	45.9	26.4
	DCA (Ours)	**69.90**	**2.66**	**87.3**	**87.8**	**90.6**	**59.6**	**63.6**	**91.3**	**66.4**	**87.0**	**76.7**	**64.0**	**55.8**	**48.5**	**46.1**	**72.7**	**51.1**

Tab. 2 Quantitative results on **5%** labeled **AMOS** dataset. 'General' or 'Imbalance' indicate whether the methods consider class-imbalance issue or not. The top-3 results are highlighted as first, second, and third.

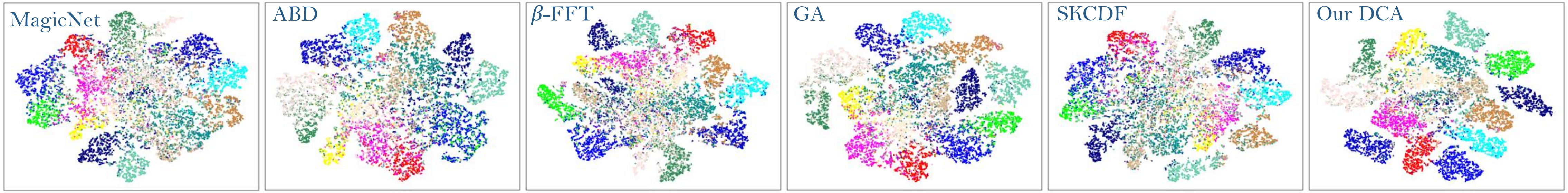


Fig. 6 The t-SNE visualizations of feature representation extracted from SOTA methods and our method on **5%** labeled **AMOS** dataset.
● Sp, ● RK, ● LK, ● Ga, ● Es, ● Li, ● St, ● Ao, ● IVC, ● Pa, ● RAG, ● LAG, ● Du, ● Bl and ● P/U.

t-SNE shows more compact intra-class clusters and clearer inter-class separation.

Ablations — Main Components + Split Strategy

Baseline	3 Experts	LGA	DFAM	Avg. Dice	Avg. ASD	Sp	RK	LK	Ga	Es	Li	St	Ao	IVC	PSV	PA	RAG	LAG
						Average Dice of Each Class												
✓				47.92$\pm$4.3	30.52$\pm$2.4	66.4	74.3	73.2	51.3	0	74.1	53.8	67.2	66.2	54.6	41.8	0	0
✓	✓			55.38$\pm$2.5	15.30$\pm$2.1	62.5	70.0	68.9	48.3	46.3	69.8	50.7	63.3	62.4	51.4	39.4	39.1	48.0
✓	✓	✓		69.89$\pm$1.6	1.37$\pm$0.8	78.9	88.3	87.0	60.9	58.4	88.0	64.0	79.9	78.7	64.8	49.7	49.3	60.6
✓	✓	✓	✓	**73.20**$\pm$**0.3**	**1.78**$\pm$**0.1**	**82.9**	**92.8**	**91.4**	**64.0**	**61.4**	**92.5**	**67.2**	**83.9**	**82.7**	**68.1**	**52.2**	**51.8**	**63.7**

Tab. 3 Ablation study for the effectiveness of each component on **20%** labeled **Synapse** dataset. 3 Experts: Three asymmetric expert decoders, LGA: Logarithmic Gap Analysis.

	Methods	Avg. Dice	Avg. ASD	Li	St	Sp	LK	RK	Ao	Bl	Pa	IVC	Du	P/U	Ga	Es	LAG	RAG
				Average Dice of Each Class														
5% L	Avg.	67.79	3.81	88.6	65.1	83.8	89.7	83.4	85.3	70.5	61.4	75.1	43.8	49.1	58.4	61.7	48.0	53.0
5% L	Fixed Th.	69.90	2.66	91.3	66.4	87.3	90.6	87.8	87.0	72.7	64.0	76.7	46.1	51.1	59.6	63.6	48.5	55.8
5% L	LGA	69.90	2.66	91.3	66.4	87.3	90.6	87.8	87.0	72.7	64.0	76.7	46.1	51.1	59.6	63.6	48.5	55.8
2% L	Avg.	55.53	6.77	74.9	53.0	65.7	69.4	64.9	69.0	56.3	55.2	62.0	41.7	42.0	45.0	44.2	43.3	46.4
2% L	Fixed Th.	57.93	5.05	79.8	54.2	69.3	70.8	70.7	69.7	59.4	55.2	65.3	42.6	43.4	48.0	48.1	45.7	46.8
2% L	LGA	59.88	3.97	81.4	57.0	72.2	73.0	72.1	73.4	61.9	56.9	66.7	44.4	43.8	50.5	49.1	47.1	48.8

Tab. 4 Ablation of different split strategies on **5%** and **2%** labeled **AMOS** dataset. Organs are highlighted by their partitioned category: Head, Medium, and Tail.

LGA is parameter-free and more stable than uniform or fixed-threshold division.

Ablations— Expert Decoder + DFAM Design

Setting	Decoder	{DeB-6, DeB-7, DeB-8, DeB-9} n_stages	{DeB-6, DeB-7, DeB-8, DeB-9} Dilation	Avg. Dice	Avg. ASD
V1	$\mathcal{D}_{\mathrm{H}}$	{3, 3, 2, 1}	{1, 1, 1, 1}	60.25	4.54
	$\mathcal{D}_{\mathrm{M}}$	{3, 3, 2, 1}	{1, 1, 1, 1}		
	$\mathcal{D}_{\mathrm{T}}$	{3, 3, 2, 1}	{1, 1, 1, 1}		
V2	$\mathcal{D}_{\mathrm{H}}$	**{2, 2, 1, 1}**	{1, 1, 1, 1}	67.33	3.08
	$\mathcal{D}_{\mathrm{M}}$	{3, 3, 2, 1}	{1, 1, 1, 1}		
	$\mathcal{D}_{\mathrm{T}}$	**{4, 4, 3, 2}**	{1, 1, 1, 1}		
V3	$\mathcal{D}_{\mathrm{H}}$	{3, 3, 2, 1}	**{2, 2, 2, 2}**	64.96	4.52
	$\mathcal{D}_{\mathrm{M}}$	{3, 3, 2, 1}	{1, 1, 1, 1}		
	$\mathcal{D}_{\mathrm{T}}$	{3, 3, 2, 1}	{1, 1, 1, 1}		
Ours	$\mathcal{D}_{\mathrm{H}}$	**{2, 2, 1, 1}**	**{2, 2, 2, 2}**	**69.90**	**2.66**
	$\mathcal{D}_{\mathrm{M}}$	{3, 3, 2, 1}	{1, 1, 1, 1}		
	$\mathcal{D}_{\mathrm{T}}$	**{4, 4, 3, 2}**	{1, 1, 1, 1}		

Tab. 5 Ablation study on expert decoder parameters on the **5%** labeled **AMOS** dataset.

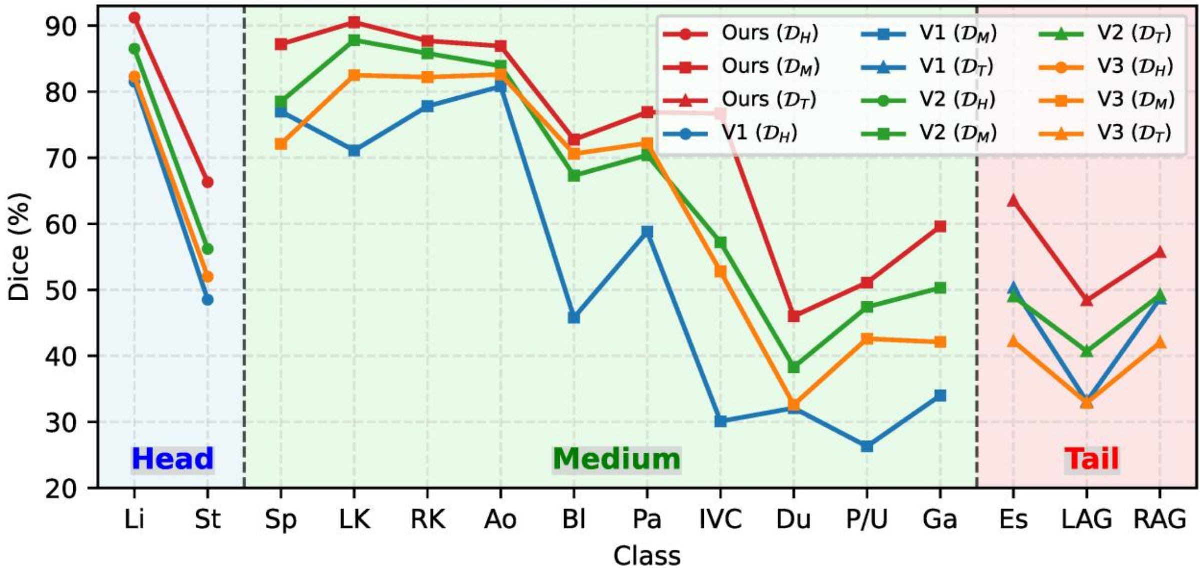


Fig. 7 Performance of parameter-varied H/M/T decoders on their respective categories on the **5%** labeled **AMOS** dataset.

Asymmetric expert design gives the best Avg. Dice and ASD.

Deeper tail decoder improves small-organ detail extraction.
Dilated head decoder improves large-organ context modeling.

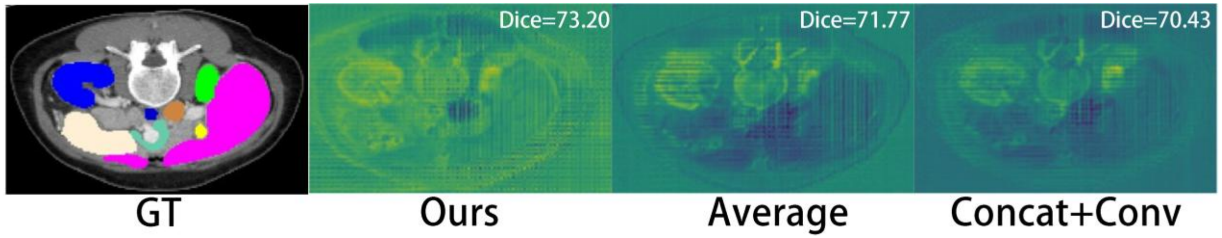


Fig. 8 Fused feature maps of different expert feature fusion strategies on the **20%** labeled **Synapse** dataset. The feature maps are visualized from the final decoder stage (DeB-9).

DFAM activates both head-class and tail-class regions better than average or concat fusion.

Conclusion

CVPR JUNE 3-7, 2026 DENVER COLORADO

DCA tackles class-imbalanced SSMIS by replacing one overloaded decoder with specialized experts and adaptive aggregation.

- DCA significantly improves class-imbalanced SSMIS by reducing majority-class dominance.
- Expert specialization enables more balanced learning across large, medium, and tail organs.
- Dynamic aggregation effectively integrates complementary expert priors for robust final prediction.
- DCA achieves superior performance on Synapse and AMOS, especially improving tail-class organs.

上海交通大學
SHANGHAI JIAO TONG UNIVERSITY

CVPR JUNE 3-7, 2026 DENVER COLORADO

Thanks for your attention!

Project Link