

Learning Beyond Pixels: Structured Knowledge Topology Consistency for Semi-supervised 2D/3D Medical Image Segmentation

Yu Chen^{1*}, Qiang Xu^{1*}, Ke Wang^{1†}, Fan Yang², and Honghao Wei¹

¹ College of Information Science and Technology, Jinan University, China

² Guangdong Institute of Smart Education, Jinan University, Guangzhou, China
{biabuluo, 1002xu}@stu2024.jnu.edu.cn, wangke@jnu.edu.cn

Abstract. Semi-supervised learning (SSL) has become a pivotal strategy for medical image segmentation by harnessing unlabeled data. However, most consistency-based paradigms focus on aligning pixel-level outputs, failing to capture the high-order structural knowledge and diagnostic logic essential for complex anatomical interpretation. Motivated by the hierarchical reasoning process in clinical diagnosis—where radiologists progressively integrate local cues, global context, and anatomical priors—we propose **SKTC-Net**, a novel framework centered on **Structured Knowledge Topology Consistency**. We reformulate the semi-supervised objective by mimicking three critical aspects of clinical diagnosis: (i) **Where to look:** *Spatial Attention Consistency* aligns saliency maps to ensure consensus on salient anatomical regions; (ii) **How to compare:** *Relational Structure Consistency* preserves the global topological distribution among learning samples to mimic comparative diagnosis; and (iii) **How to reason:** *Evolutionary Flow Consistency* constrains the inter-layer feature transitions to maintain a stable diagnostic reasoning path. Evaluations on ACDC (2D) and Pancreas (3D) datasets demonstrate that SKTC-Net consistently performs superior. Notably, SKTC-Net achieves consistent gains in low-label settings (e.g., +3.47% Dice at 10% Pancreas), attributed to the complementary supervision from multi-level structural constraints. The source code and dataset are publicly available at <https://github.com/SSonnyboy/STKC-Net>.

Keywords: Semi-supervised learning · Structured consistency · Clinical diagnostic logic · Medical image segmentation.

1 Introduction

Accurate segmentation of anatomical structures is a cornerstone of computer-aided diagnosis. Despite the remarkable success of deep convolutional neural networks (CNNs) like U-Net [14] and its modern variants [16, 4], their performance heavily relies on large-scale, pixel-level annotated datasets. In the medical domain, acquiring such data is exceptionally costly, requiring expert-level

* Equal contribution. † Corresponding author.

knowledge and strict privacy protection [25, 12]. Consequently, semi-supervised medical image segmentation (SSMIS) has emerged as a strategic solution to harvest descriptive cues from vast amounts of unlabeled data, thereby reducing the reliance on manual expertise [2, 24].

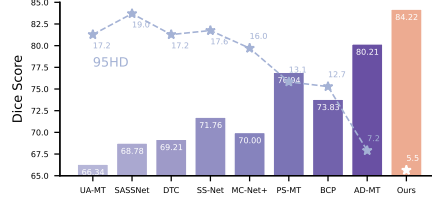
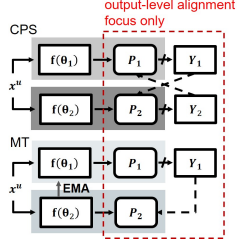


Fig. 1: Basic framework. Fig. 2: Result of SKTC-Net on 10% Pancreas.

Current SSMIS frameworks predominantly rely on consistency regularization (CR), which encourages invariant predictions under diverse perturbations [17, 26]. While influential models like Mean-Teacher (MT) [17] and UA-MT [25] have set strong baselines, most existing methods focus on **output-level alignment**, i.e., forcing the final probability maps to be similar (as shown in Fig. 1). Unlike standard supervision that focuses on output fidelity, clinical diagnosis inherently involves multi-dimensional reasoning: identifying salient landmarks, comparing global case relationships, and maintaining coherent inference paths. We argue that enforcing consistency on these intermediate structural properties better captures the diagnostic logic embedded in the latent space.

Recent advances have begun to explore beyond pixel-level consistency. For instance, Zhu *et al.* [28] introduced adversarial masked image modeling to unleash the potential of transformers in SSMIS by constructing auxiliary masked domains with adversarial training. Wang *et al.* [18] proposed graph-based clustering to explicitly model data structural information via graph convolutional networks, achieving significant improvements by leveraging pairwise affinities between local features. Liu *et al.* [9] presented translation consistent co-training (TraCoCo) that perturbs input views by varying spatial context, enabling the model to learn segmentation patterns from foreground objects rather than spatial contexts. Peng *et al.* [13] developed a difficulty-adapted masked transformer that modulates reconstruction difficulty through region-based and class-based masking strategies, demonstrating superior performance on multiple benchmarks. Despite these promising directions, existing SSMIS methods—ranging from early uncertainty-aware frameworks [25] to recent advanced paradigms such as bi-directional copy-paste (BCP) [2], cross-pseudo supervision (CPS) [3], and multi-student collaborative learning [10, 20]—still primarily prioritize the consistency of pixel-level probability distributions. While effective, this focus may unintentionally overlook the high-order structural integrity and global topological

constraints essential for complex medical scans. Consequently, without explicit guidance on internal feature representations, these models remain susceptible to boundary blurring and topological fragmented errors, particularly when pseudo-labels are contaminated by noise or bias accumulation in low-data regimes [1, 19, 6].

To transcend pixel-level output consistency, we propose **SKTC-Net**, which enforces **Structured Knowledge Topology Consistency** on intermediate representations to capture high-order anatomical dependencies. Unlike existing works that introduce redundant network branches to provide diversity [27], we shift the paradigm toward a systematic mimicry of the clinical diagnostic process across three orthogonal dimensions:

1. **Where to Look (Spatial Attention Consistency, SAC):** Inspired by the varying attention levels experts allocate across different image regions, SAC enforces consistency of network attention maps, which ensures geometric consensus on organ contours and critical landmarks, extending beyond simple pixel-wise consistency.
2. **How to compare (Relational Structure Consistency, RSC):** Clinical decisions are often refined by comparing current cases with previous experience. RSC preserves the global topological relationships between samples in a batch by aligning their similarity matrices, mimicking comparative diagnostic logic to enhance discrimination against complex backgrounds.
3. **How to reason (Evolutionary Flow Consistency, EFC):** Expertise lies in the logical transition from low-level textures to high-level semantic structures. EFC captures the “flow” of feature abstraction between successive layers via inter-layer Gram matrices, ensuring that the model’s internal reasoning path remains stable under perturbations.

The synergy of these modules allows SKTC-Net to learn a geometrically-aware and logically consistent representation. We evaluate our method on two representative tasks: ACDC (2D cardiac) and Pancreas (3D small organ). Experimental results (Fig. 2) demonstrate that SKTC-Net significantly outperforms competitive frameworks [2, 27, 8, 25, 21, 6, 22, 11].

2 Methodology

The proposed SKTC-Net aims to mimic the multi-dimensional knowledge transfer in clinical apprenticeship. As illustrated in Fig. 3, our framework adopts a dual-student architecture consisting of two networks S_1 and S_2 with different initializations or perturbations. Given a labeled dataset $\mathcal{D}_L = \{(x_i, y_i)\}_{i=1}^{N_L}$ and an unlabeled dataset $\mathcal{D}_U = \{x_i\}_{i=N_L+1}^{N_L+N_U}$, the total objective is defined as:

$$\mathcal{L} = \mathcal{L}_{sup}(\mathcal{D}_L) + \lambda \mathcal{L}_{con}(\mathcal{D}_U) + \alpha \mathcal{L}_{SKTC}(\mathcal{D}_{L \cup U}) \quad (1)$$

where λ adjusts the semi-supervised loss weight (consistent with AD-MT [27]), and α is the hyperparameter of our new loss term. $\mathcal{L}_{sup}$ is the standard supervised

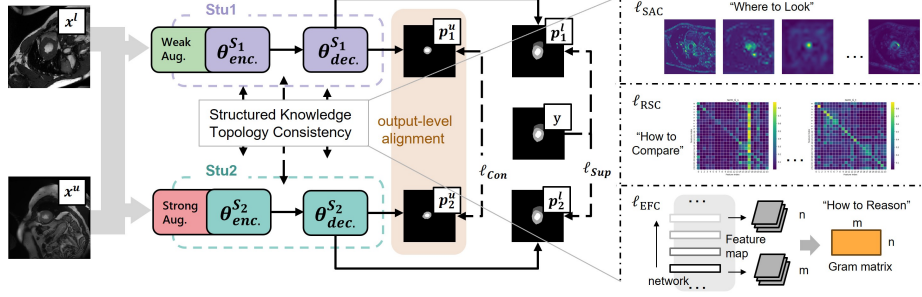


Fig. 3: The pipeline of our SKTC-Net framework.

loss, $\mathcal{L}_{con}$ is the conventional two-teacher consistency loss (based on output), and $\mathcal{L}_{SKTC}$ is our core contribution, formulated as the ensemble of three diagnostic-inspired constraints:

$$\mathcal{L}_{SKTC} = \mathcal{L}_{SAC} + \mathcal{L}_{RSC} + \mathcal{L}_{EFC} \quad (2)$$

SAC: Directing “Where to Look.” Radiological interpretation prioritizes salient anatomical regions over background noise. To embed this diagnostic prior, **Spatial Attention Consistency (SAC)** enforces agreement between the spatial saliency maps of the dual student networks. For a feature map $F_{S_m}^{l,b} \in \mathbb{R}^{C_l \times H_l \times W_l}$ extracted from layer l of student S_m ($m \in \{1, 2\}$) for the b -th sample, we generate a 2D attention map via channel-wise aggregation with power factor p :

$$\mathcal{A}(F_{S_m}^{l,b}; p) = \sum_{c=1}^{C_l} |F_{S_m,c}^{l,b}|^p, \quad (3)$$

where we set $p = 2$ to emphasize dominant activations. To eliminate scale variance across student perspectives, we normalize each attention map to unit L_2 norm over its spatial dimensions (H_l, W_l):

$$\hat{\mathcal{A}}(F_{S_m}^{l,b}) = \frac{\mathcal{A}(F_{S_m}^{l,b}; p)}{\|\mathcal{A}(F_{S_m}^{l,b}; p)\|_{2, (H_l, W_l)} + \epsilon}, \quad (4)$$

with $\epsilon = 10^{-6}$ for numerical stability. The SAC loss is then defined as the mean squared error between normalized attention maps across all supervised bottleneck layers Ω :

$$\mathcal{L}_{SAC} = \frac{1}{|\Omega|} \sum_{l \in \Omega} \frac{1}{B} \sum_{b=1}^B \frac{1}{H_l W_l} \left\| \hat{\mathcal{A}}(F_{S_1}^{l,b}) - \hat{\mathcal{A}}(F_{S_2}^{l,b}) \right\|_{2, (H_l, W_l)}^2, \quad (5)$$

where H_l and W_l denote the spatial dimensions of layer l , $|\Omega|$ is the number of supervised layers, and B is the batch size. This formulation ensures that both

students reach a geometric consensus on the spatial localization of discriminative regions—effectively teaching them *where to look* for organ boundaries—while maintaining scale invariance and numerical robustness.

RSC: Mimicking “How to Compare.” Experienced radiologists refine their judgment by comparing the current case with a mental database of prior cases. We formalize this comparative logic as **Relational Structure Consistency (RSC)**. For a batch of B unlabeled samples, we extract global representation vectors from each supervised bottleneck layer $l \in \Omega$ by spatially flattening the feature maps: $\mathbf{f}_m^{l,b} = \text{flatten}(F_{S_m}^{l,b}) \in \mathbb{R}^{D_l}$ where $D_l = C_l \times H_l \times W_l$ and $m \in \{1, 2\}$ denotes the student index. Each vector is L_2 -normalized to unit norm before similarity computation:

$$\bar{\mathbf{f}}_m^{l,b} = \frac{\mathbf{f}_m^{l,b}}{\|\mathbf{f}_m^{l,b}\|_2 + \epsilon}, \quad \epsilon = 10^{-6} \quad (6)$$

The batch-wise cosine similarity matrix $G_m^l \in \mathbb{R}^{B \times B}$ for student S_m is then constructed as:

$$G_{m,ij}^l = \bar{\mathbf{f}}_m^{l,i} \cdot (\bar{\mathbf{f}}_m^{l,j})^\top \quad (7)$$

which, due to the prior normalization, directly yields cosine similarity without explicit denominator terms. The RSC loss enforces consistent relational topology between the dual students across all supervised layers:

$$\mathcal{L}_{\text{RSC}} = \frac{1}{|\Omega|} \sum_{l \in \Omega} \frac{1}{B^2} \sum_{i=1}^B \sum_{j=1}^B (G_{1,ij}^l - G_{2,ij}^l)^2 \quad (8)$$

This formulation—equivalent to the mean squared error between normalized similarity matrices—preserves the intrinsic geometric structure of the latent space. By aligning the relative distances between patient cases across student perspectives, the model learns to cluster anatomically similar features without explicit supervision, effectively mimicking the comparative diagnostic reasoning of radiologists.

EFC: Modeling “How to Reason.” The transition from recognizing low-level textures to synthesizing high-level semantic organs represents the diagnostic reasoning path. We propose **Evolutionary Flow Consistency (EFC)** to stabilize this hierarchical abstraction process. For adjacent layers l and $l+1$ with feature maps $F_{S_m}^l \in \mathbb{R}^{B \times C_l \times H_l \times W_l}$ and $F_{S_m}^{l+1} \in \mathbb{R}^{B \times C_{l+1} \times H_{l+1} \times W_{l+1}}$ from student S_m , we first align their spatial dimensions via adaptive average pooling to a common resolution (H^*, W^*) . The inter-layer evolutionary flow for sample b is then modeled by a Gram matrix $\mathcal{G}_m^{l,b} \in \mathbb{R}^{C_{l+1} \times C_l}$:

$$\mathcal{G}_{m,nm}^{l,b} = \frac{1}{H^* W^*} \sum_{h=1}^{H^*} \sum_{w=1}^{W^*} F_{S_m,m}^{l,b,h,w} \cdot F_{S_m,n}^{l+1,b,h,w} \quad (9)$$

where $m \in \{1, \dots, C_l\}$ and $n \in \{1, \dots, C_{l+1}\}$ are channel indices of the shallow and deep layers respectively. This formulation captures how activations in shallow channels *flow into* and influence deep channels during hierarchical reasoning.

Table 1: Performance comparison with the other methods on the **ACDC (2D)**, in the semi-supervised setting of 5% and 10% labeled data.

Method	ACDC (5% / 3 labeled data)				ACDC (10% / 7 labeled data)			
	Dice $\uparrow$	Jaccard $\uparrow$	95HD $\downarrow$	ASD $\downarrow$	Dice $\uparrow$	Jaccard $\uparrow$	95HD $\downarrow$	ASD $\downarrow$
UNet (SupOnly)	47.83	37.01	31.16	12.62	79.41	68.11	9.35	2.70
UA-MT [25] (MICCAI'19)	46.04	35.97	20.08	7.75	81.65	70.64	6.88	2.02
SS-Net [21] (MICCAI'22)	65.82	55.38	6.67	2.28	86.78	77.67	6.07	1.40
BCP [2] (CVPR'23)	87.59	78.67	1.90	0.67	88.84	80.62	3.98	1.17
AD-MT [27] (ECCV'24)	88.75	80.41	1.48	0.50	89.46	81.47	1.51	0.44
M^3HL [7] (MICCAI'25)	88.31	80.62	1.83	0.56	90.47	83.23	1.43	0.34
ALHVR [5] (ICCV'25)	82.25	72.92	4.11	1.00	90.56	83.21	2.57	0.78
SKTC-Net (Ours)	89.74$\pm$0.001	81.85$\pm$0.001	1.28$\pm$0.218	0.34$\pm$0.053	90.76$\pm$0.001	83.59$\pm$0.001	1.27$\pm$0.135	0.32$\pm$0.028

The EFC loss enforces consistency of this flow pattern between dual students across all adjacent layer pairs $\mathcal{K}$:

$$\mathcal{L}_{\text{EFC}} = \frac{1}{|\mathcal{K}|} \sum_{k \in \mathcal{K}} \frac{1}{B \cdot C_{l_{k+1}} \cdot C_{l_k}} \sum_{b=1}^B \left\| \mathcal{G}_1^{k,b} - \mathcal{G}_2^{k,b} \right\|_F^2 \quad (10)$$

where $\|\cdot\|_F$ denotes the Frobenius norm. This layer-wise constraint—equivalent to mean squared error between per-sample channel correlation matrices—ensures that both students develop a consistent diagnostic reasoning trajectory as information propagates through the network hierarchy.

3 Experiments

Datasets and Evaluation Metrics. In terms of experimental datasets, we followed previous works and adopted three widely used benchmark datasets to validate the effectiveness and feasibility of our proposed method: the Automated Cardiac Diagnosis Challenge (ACDC) dataset [23] and the Pancreas-NIH dataset [15]. For evaluation, we used metrics from previous works [27, 2] and used the following metrics: Dice score (%), Jaccard score (%), 95% Hausdorff distance (95HD), and the average surface distance (ASD).

To evaluate the effectiveness of the proposed **SKTC-Net**, we conduct comprehensive evaluations on three benchmark datasets: ACDC, LA, and Pancreas. Our method is compared against a range of representative semi-supervised segmentation frameworks, including established paradigms like UA-MT [25], as well as more recent high-performance models such as BCP [2], AD-MT [27], M^3HL [7], and ALHVR [5].

Implementation Details. Following previous work [27], the segmentation model for the 2D ACDC dataset was trained for 30,000 iterations with a batch size of 24 (12 labeled and 12 unlabeled samples). For the LA and Pancreas datasets, training was conducted for 15,000 iterations with a batch size of 4 (2 labeled and 2 unlabeled samples). We use Unet and Vnet as the main backbone network for our experiments. Following by [27], we set the maximum loss weight $\lambda_{max} = 2.0$. In subsequent experiments, the SKTC loss weight α was set to 0.1. For our experimental result, we use **three different random seed runs**

Table 2: Performance comparison with other methods on the **Pancreas (3D)**, in the semi-supervised setting of 10% and 20% labeled data.

Method	Pancreas (10% / 6 labeled data)				Pancreas (20% / 12 labeled data)			
	Dice $\uparrow$	Jaccard $\uparrow$	95HD $\downarrow$	ASD $\downarrow$	Dice $\uparrow$	Jaccard $\uparrow$	95HD $\downarrow$	ASD $\downarrow$
VNet (SupOnly)	55.60	41.74	45.33	18.63	72.38	58.26	19.35	5.89
UA-MT[25] (MICCAI'19)	66.34	53.21	17.21	4.57	76.10	62.62	10.84	2.43
SS-Net[21] (MICCAI'22)	71.76	57.05	17.56	5.77	78.98	66.32	8.86	2.01
BCP[2] (CVPR'23)	73.83	59.24	12.71	3.72	82.91	70.97	6.43	2.25
AD-MT [27] (ECCV'24)	80.21	67.51	7.18	1.66	82.61	70.70	4.94	1.38
M^3HL [7] (MICCAI'25)	80.75	68.26	6.54	1.56	82.94	71.08	4.83	1.32
ALHVR [5] (ICCV'25)	79.58	65.32	9.41	2.83	81.42	69.37	6.75	2.31
SKTC-Net (Ours)	84.22 $\pm$ 0.001	73.64 $\pm$ 0.002	5.48 $\pm$ 0.846	1.31 $\pm$ 0.254	84.86 $\pm$ 0.001	74.39 $\pm$ 0.001	4.95 $\pm$ 0.674	1.25 $\pm$ 0.176

to obtain the average. All experiments were implemented using the PyTorch framework and conducted on an NVIDIA GeForce RTX 3090 GPU.

Results on ACDC. Tab. 1 summarizes the performance on the 2D cardiac ACDC dataset. Following the experimental protocol in AD-MT [27], we utilize 5% and 10% labeled data. SKTC-Net achieves favorable performance across most metrics: with 5% labels, it reaches 89.74% Dice, comparing favorably against AD-MT (88.75%) and M^3HL (88.31%). At 10% labels, our method maintains this trend with 90.76% Dice. All results are reported as mean $\pm$ standard deviation over 3 random seeds (the standard deviation for both Dice and Jac is the result of not multiplying by 100%). The consistent improvements suggest that structured knowledge consistency provides beneficial regularization under limited annotation.

Results on Pancreas. Tab. 2 presents results on the challenging 3D Pancreas dataset with 10% and 20% labeled data, consistent with the ACDC setup. SKTC-Net demonstrates competitive improvements, particularly at 10% labels where it achieves 84.22% Dice, showing a notable margin over M^3HL (80.75%). At 20% labels, our method reaches 84.86% Dice with the lowest ASD (1.25). While the 95HD at 20% is comparable to M^3HL (4.95 vs. 4.83), the overall boundary quality remains robust. These observations indicate that incorporating intermediate-level structural constraints may help enhance segmentation stability for complex 3D organs when labeled data is scarce.

Table 3: Ablation					Table 4: α			Table 5: Plug and play		
Components			Metrics		α	Metrics		Methods	Metrics	
$\mathcal{L}_{SAC}$	$\mathcal{L}_{RSC}$	$\mathcal{L}_{EFC}$	Dice	95HD		Dice	95HD		Dice	95HD
			80.04	7.37	0.01	83.05	5.83	MT	79.85	7.92
✓			81.73	5.64	0.1	84.22	5.48	+SKTC	83.45	6.17
✓	✓		82.02	5.52	1	84.08	5.62	CPS	80.04	7.37
✓	✓	✓	84.22	5.48	10	81.78	6.46	+SKTC	84.22	5.48

Ablation Study. Tab. 3 examines the contribution of each SKTC component on the Pancreas dataset (10% labeled). Starting from the CPS baseline

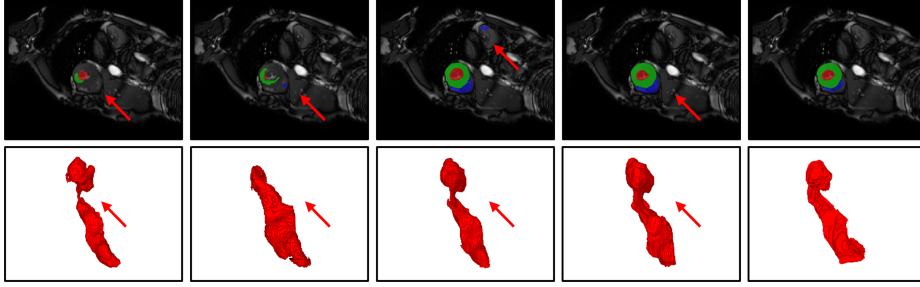


Fig. 4: Qualitative comparison on ACDC (5% labeled, 2D) and Pancreas (10% labeled, 3D). From left to right: UA-MT [25], BCP [2], AD-MT [27], SKTC-Net (Ours), and Ground Truth. Red arrows highlight regions where SKTC-Net produces more anatomically plausible boundaries.

(80.04% Dice), adding $\mathcal{L}_{SAC}$ yields a noticeable gain (+1.69% Dice), suggesting that aligning spatial attention helps the model focus on anatomically relevant regions. Further incorporating $\mathcal{L}_{RSC}$ provides modest additional improvements (+0.29% Dice), indicating that preserving inter-sample relational structures may enhance discrimination in complex backgrounds. The full integration of $\mathcal{L}_{EFC}$ achieves the highest performance (84.22% Dice), which suggests that regularizing cross-layer feature transitions contributes to more stable representation learning. While each component offers incremental benefits, their combination appears to provide complementary supervision for structured feature alignment.

Sensitivity to Loss Weight α . We investigate the influence of the SKTC loss weight α on the Pancreas dataset (10% labeled), as shown in Tab. 4. Performance remains relatively stable across $\alpha \in [0.01, 1]$, with the highest Dice observed at $\alpha = 0.1$. A very small weight ($\alpha = 0.01$) leads to slightly reduced gains, which may indicate insufficient regularization from structured constraints. Conversely, a large weight ($\alpha = 10$) shows more noticeable degradation, possibly due to over-emphasis on feature-level consistency at the expense of task-specific supervision. These observations suggest that SKTC-Net is reasonably robust to moderate variations in α , and a balanced weighting (e.g., 0.1) may help harmonize pixel-level and structure-level objectives.

Plug-and-Play Compatibility. To assess the generality of SKTC, we integrate it into two representative SSL frameworks (MT and CPS) on Pancreas (10% labeled), as summarized in Tab. 5. Adding SKTC consistently improves both baselines (+3.60% Dice for MT, +4.18% for CPS), suggesting that the proposed structured consistency constraints can complement existing consistency-based paradigms without architectural modifications. This plug-and-play behavior indicates potential for broader applicability across different semi-supervised segmentation frameworks.

Qualitative Analysis. Figure 4 shows visual comparisons under low-label settings. SKTC-Net produces smoother boundaries and fewer fragmented regions compared to baselines, particularly in low-contrast areas of Pancreas. These ob-

servations align with our quantitative results, suggesting that SKTC consistency constraints may help preserve anatomical plausibility when annotations are limited.

4 Conclusion

In this work, We propose **SKTC-Net**, a novel semi-supervised framework for medical image segmentation that enforces **Structured Knowledge Topology Consistency** across spatial attention, relational structure, and evolutionary flow. The method captures high-order anatomical priors by mimicking clinical reasoning in the latent space. Experiments on ACDC and Pancreas datasets demonstrate that SKTC-Net consistently outperforms competitive baselines, especially under extremely limited labeled data. Future work will explore adaptive topology construction and extension to broader medical imaging applications.

Acknowledgments. This work was supported in part by National Key Program of National Natural Science of China (Grant No. 82430108).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Arazo, E., Ortego, D., Albert, P., O’Connor, N.E., McGuinness, K.: Pseudo-labeling and confirmation bias in deep semi-supervised learning. In: 2020 International joint conference on neural networks (IJCNN). pp. 1–8. IEEE (2020)
2. Bai, Y., Chen, D., Li, Q., Shen, W., Wang, Y.: Bidirectional copy-paste for semi-supervised medical image segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 11514–11524 (2023)
3. Chen, X., Yuan, Y., Zeng, G., Wang, J.: Semi-supervised semantic segmentation with cross pseudo supervision. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 2613–2622 (2021)
4. Han, Z., Jian, M., Wang, G.G.: Convunext: An efficient convolution neural network for medical image segmentation. *Knowledge-based systems* **253**, 109512 (2022)
5. Lei, T., Yang, Z., Wang, X., Wang, Y., Wang, X., Sun, F., Nandi, A.K.: Adaptive learning of high-value regions for semi-supervised medical image segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 21450–21459 (2025)
6. Li, S., Zhang, C., He, X.: Shape-aware semi-supervised 3d semantic segmentation for medical images. In: Medical Image Computing and Computer Assisted Intervention. pp. 552–561. Springer (2020)
7. Liu, Y., Zhang, Z., Yue, J., Guo, W., Li, D.: M3HL: Mutual Mask Mix with High-Low Level Feature Consistency for Semi-Supervised Medical Image Segmentation . In: proceedings of Medical Image Computing and Computer Assisted Intervention – MICCAI 2025. vol. LNCS 15961. Springer Nature Switzerland (September 2025)
8. Liu, Y., Tian, Y., Chen, Y., Liu, F., Belagiannis, V., Carneiro, G.: Perturbed and strict mean teachers for semi-supervised semantic segmentation. In: IEEE Conf. Comput. Vis. Pattern Recog. (2022)

9. Liu, Y., Tian, Y., Wang, C., Chen, Y., Liu, F., Belagiannis, V., Carneiro, G.: Translation consistent semi-supervised segmentation for 3D medical images. *IEEE Transactions on Medical Imaging* **43**(12), 3700–3712 (2024). <https://doi.org/10.1109/TMI.2024.3396922>
10. Luo, W., Yang, M.: Semi-supervised semantic segmentation via strong-weak dual-branch network. In: *European Conference on Computer Vision*. pp. 784–800. Springer (2020)
11. Luo, X., Chen, J., Song, T., Wang, G.: Semi-supervised medical image segmentation through dual-task consistency. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. vol. 35, pp. 8801–8809 (2021)
12. Luo, X., Hu, M., Song, T., Wang, G., Zhang, S.: Semi-supervised medical image segmentation via cross teaching between cnn and transformer. In: *International conference on medical imaging with deep learning*. pp. 820–833. PMLR (2022)
13. Peng, Z., et al.: DMformer: Difficulty-adapted masked transformer for semi-supervised medical image segmentation. *IEEE Journal of Biomedical and Health Informatics* **29**(4), 2345–2356 (2025). <https://doi.org/10.1109/JBHI.2025.3534698>
14. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: *Medical image computing and computer-assisted intervention—MICCAI 2015: 18th international conference, Munich, Germany, October 5–9, 2015, proceedings, part III* 18. pp. 234–241. Springer (2015)
15. Roth, H.R., Lu, L., Farag, A., Shin, H.C., Liu, J., Turkbey, E.B., Summers, R.M.: Deeporgan: Multi-level deep convolutional networks for automated pancreas segmentation. In: *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part I* 18. pp. 556–564. Springer (2015)
16. Tang, F., Ding, J., Quan, Q., Wang, L., Ning, C., Zhou, S.K.: Cmunext: An efficient medical image segmentation network based on large kernel and skip fusion. In: *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*. pp. 1–5. IEEE (2024)
17. Tarvainen, A., Valpola, H.: Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *Advances in neural information processing systems* **30** (2017)
18. Wang, M., Li, J., Yin, N., Yang, L., Su, H., Shen, L.: Graphcl: Graph-based clustering for semi-supervised medical image segmentation. In: *International Conference on Machine Learning (ICML)*. PMLR (2025)
19. Wang, Y., Xiao, B., Bi, X., Li, W., Gao, X.: Mcf: Mutual correction framework for semi-supervised medical image segmentation. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp. 15651–15660 (2023)
20. Wang, Z., Zhao, Z., Xing, X., Xu, D., Kong, X., Zhou, L.: Conflict-based cross-view consistency for semi-supervised semantic segmentation. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp. 19585–19595 (2023)
21. Wu, Y., Wu, Z., Wu, Q., Ge, Z., Cai, J.: Exploring smoothness and class-separation for semi-supervised medical image segmentation. In: *International conference on medical image computing and computer-assisted intervention*. pp. 34–43. Springer (2022)
22. Wu, Y., Xu, M., Ge, Z., Cai, J., Zhang, L.: Semi-supervised left atrium segmentation with mutual consistency training. In: *Medical Image Computing and Computer Assisted Intervention—MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part II* 24. pp. 297–306. Springer (2021)

23. Xiong, Z., Xia, Q., Hu, Z., Huang, N., Bian, C., Zheng, Y., Vesal, S., Ravikumar, N., Maier, A., Yang, X., et al.: A global benchmark of algorithms for segmenting the left atrium from late gadolinium-enhanced cardiac magnetic resonance imaging. *Medical image analysis* **67**, 101832 (2021)
24. You, C., Dai, W., Min, Y., Staib, L., Duncan, J.S.: Bootstrapping semi-supervised medical image segmentation with anatomical-aware contrastive distillation. In: *International conference on information processing in medical imaging*. pp. 641–653. Springer (2023)
25. Yu, L., Wang, S., Li, X., Fu, C.W., Heng, P.A.: Uncertainty-aware self-ensembling model for semi-supervised 3d left atrium segmentation. In: *Medical image computing and computer assisted intervention—MICCAI 2019: 22nd international conference, Shenzhen, China, October 13–17, 2019, proceedings, part II* 22. pp. 605–613. Springer (2019)
26. Yuan, J., Liu, Y., Shen, C., Wang, Z., Li, H.: A simple baseline for semi-supervised semantic segmentation with strong data augmentation. In: *Proceedings of the IEEE/CVF international conference on computer vision*. pp. 8229–8238 (2021)
27. Zhao, Z., Wang, Z., Wang, L., Yu, D., Yuan, Y., Zhou, L.: Alternate diverse teaching for semi-supervised medical image segmentation. In: *European Conference on Computer Vision*. pp. 227–243. Springer (2024)
28. Zhu, J., et al.: Advmim: Adversarial masked image modeling for semi-supervised medical image segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Springer (2025)