DyCON: Dynamic Uncertainty-aware Consistency and Contrastive Learning for Semi-supervised Medical Image Segmentation

Supplementary Material

This supplementary material is organized as follows:

- **Appendix** 1 provides more details about the datasets.
- **Appendix** 2 provides the gradient analysis of UnCL loss function with loss surface and grad-cam visualizations.
- **Appendix** 3 provides more ablation analysis of different properties of both loss functions.
- **Appendix** 4 provides more qualitative results from different datasets.
- **Appendix** 5 provides additional experiments on multi-organ segmentation task using BTCV dataset [4].

All of the above experimental analysis are conducted with a model trained using 10% labeled data for both ISLES'22 and BraTS'19 datasets to ensure consistency.

1. Datasets

ISLES-2022: The ISLES-2022 dataset [6][1] focuses on segmenting acute and sub-acute ischemic stroke lesions using 3D multi-channel MRI scans (DWI, ADC, and FLAIR). Unless otherwise specified, we use the DWI modality only, as it is highly sensitive to detecting acute ischemic lesions by excluding FLAIR due to registration complexities. The dataset consists of 250 skull-stripped cases (three cases {150, 151, 170} without lesion masks). We use 200 (80%) images for training and the rest 50 (20%) images for validation and testing. This dataset presents a challenge due to the small and scattered nature of stroke lesions, making it well-suited for evaluating DyCON's ability to handle uncertain and imbalanced regions.

Fig. 1 highlights the severe class imbalance in this dataset by revealing the distribution of lesion counts, sizes and scatterness. The figure emphasizes the challenge posed by the dominance of non-lesion and small lesions and the scarcity of large ones, which complicates model training and impacts segmentation performance.

BraTS2019: The BraTS2019 dataset [7] consists of 335 preoperative MRI scans (T1, T1ce, T2, T2-FLAIR) with annotations for brain tumors. We use the T2-FLAIR sequence for whole tumor segmentation, since it enhances the visibility of peritumoral edema critical for accurately capturing the entire tumor region. The dataset is split into 250/25/60 scans for training, validation, and testing. The focus of DyCON on uncertainty and contrastive learning is particularly valuable for this dataset, such as highly heterogeneous tumor regions, which poses challenges in segmentation.

LA dataset: The Left Atrium (LA) dataset [11] includes

[1] https://isles22.grand-challenge.org/isles22/

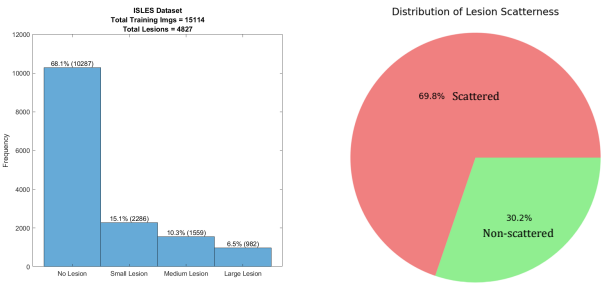


Figure 1. Characteristics of ISLES-2022 dataset. (Left) Lesion size distribution, and (Right) Lesion scatterness distribution.

100 3D gadolinium-enhanced MRI (GE-MRIs) images with manually annotated left atrial segmentation masks. Each scan has an isotropic resolution of $0.625 \times 0.625 \times 0.625 mm^3$. Following [12], we used the 80/20 train/test split, with images cropped to the heart region and normalized to zero mean and unit variance. This dataset poses a class imbalance problem, with the left atrium occupying a small portion of the overall image.

Pancreas-NIH: The Pancreas-NIH dataset [9] comprises 82 3D CT images with pancreas annotations. For comparison, we follow split settings from BCP [1]. This dataset is particularly challenging due to the pancreas's proximity to other organs and its unclear boundaries. DyCON's ability to adapt to uncertain regions makes it a strong candidate for accurate boundary delineation in this dataset.

2. Gradient Analysis of UnCL Loss

According to the DyCON framework in Fig.2, the teacher model f_θ^t evolves via an EMA of the student model f_θ^s, and only the student model is optimized using UnCL loss:

$$\mathcal{L}_{\text{UnCL}} = \frac{1}{N}\sum_{i=1}^{N} \frac{\mathcal{L}\left(p_i^s, p_i^t\right)}{\exp(\beta \cdot H_s(p_i^s)) + \exp(\beta \cdot H_t(p_i^t))} + \frac{\beta}{N}\sum_{i=1}^{N}\left(H_s(p_i^s) + H_t(p_i^t)\right) \tag{1}$$

where:

- $p^s = \sigma(f_\theta^s(x_i^s))$ are the student predictions (trainable).
- $p^t = \sigma(f_\theta^t(x_j^t))$ are the teacher predictions (non-trainable).
- $H_s(p_i^s) = -\sum_c p_{i,c}^s \log p_{i,c}^s$ is the entropy of the student predictions,
- $H_t(p_i^t) = -\sum_c p_{i,c}^t \log p_{i,c}^t$ is the entropy of the teacher predictions,

- $\beta > 0$ is a hyperparameter controlling the influence of entropy regularization.

Gradient Computation: Since f_θ^t is updated via EMA, the gradient computations focus solely on the student model f_θ^s. Therefore, the gradient of $\mathcal{L}_{\text{UnCL}}$ with respect to the student model parameters θ^s is:

$$\frac{\partial \mathcal{L}_{\text{UnCL}}}{\partial \theta^s} = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{\partial}{\partial \theta^s}\left(\frac{\mathcal{L}(p_i^s, p_i^t)}{\exp(\beta \cdot H_s(p_i^s)) + \exp(\beta \cdot H_t(p_i^t))}\right) + \frac{\partial}{\partial \theta^s}\left(\beta H_s(p_i^s)\right)\right]. \tag{2}$$

The consistency loss term (*e.g.* MSE) is defined as:

$$\mathcal{L}(p_i^s, p_i^t) = \|p_i^s - p_i^t\|^2. \tag{3}$$

The gradient of $\mathcal{L}(p_i^s, p_i^t)$ with respect to p_i^s is:

$$\frac{\partial \mathcal{L}(p_i^s, p_i^t)}{\partial p_i^s} = 2(p_i^s - p_i^t). \tag{4}$$

Using the chain rule, the gradient with respect to θ^s is:

$$\frac{\partial \mathcal{L}(p_i^s, p_i^t)}{\partial \theta^s} = 2(p_i^s - p_i^t) \cdot \frac{\partial p_i^s}{\partial \theta^s}. \tag{5}$$

Therefore, the denominator $\exp(\beta \cdot H_s(p_i^s)) + \exp(\beta \cdot H_t(p_i^t))$ inversely scales the gradient by predictive uncertainty. Its derivative with respect to θ^s is:

$$\frac{\partial}{\partial \theta^s}\left(\exp(\beta \cdot H_s(p_i^s)) + \exp(\beta \cdot H_t(p_i^t))\right) = \beta \exp(\beta \cdot H_s(p_i^s)) \times \frac{\partial H_s(p_i^s)}{\partial p_i^s} \times \frac{\partial p_i^s}{\partial \theta^s}. \tag{6}$$

Here, the entropy gradient $\frac{\partial H_s(p_i^s)}{\partial p_i^s}$ is:

$$\frac{\partial H_s(p_i^s)}{\partial p_{i,c}^s} = -(\log p_{i,c}^s + 1). \tag{7}$$

Thus, the gradient of the consistency loss is modulated by both the alignment term $p_i^s - p_i^t$ and the entropy-based scaling factor.

The second term $\frac{\beta}{N}\sum_{i=1}^{N} H_s(p_i^s)$ in Eq.(1) involves only the student entropy:

$$\frac{\partial H_s(p_i^s)}{\partial p_{i,c}^s} = -(\log p_{i,c}^s + 1), \tag{8}$$

and thus:

$$\frac{\partial}{\partial \theta^s}\left(H_s(p_i^s)\right) = \frac{\partial H_s(p_i^s)}{\partial p_i^s} \cdot \frac{\partial p_i^s}{\partial \theta^s}. \tag{9}$$

As a result, this term encourages the student model to produce more confident predictions in unambiguous regions as training progresses. Overall, the gradient of $\mathcal{L}_{\text{UnCL}}$ with respect to θ^s is:

$$\frac{\partial \mathcal{L}_{\text{UnCL}}}{\partial \theta^s} = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{2(p_i^s - p_i^t) \cdot \frac{\partial p_i^s}{\partial \theta^s}}{\exp(\beta H_s(p_i^s)) + \exp(\beta H_t(p_i^t))} - \frac{\beta \cdot (p_i^s - p_i^t)^2 \cdot \exp(\beta H_s(p_i^s)) \cdot \frac{\partial H_s(p_i^s)}{\partial p_i^s} \cdot \frac{\partial p_i^s}{\partial \theta^s}}{\left(\exp(\beta H_s(p_i^s)) + \exp(\beta H_t(p_i^t))\right)^2} + \beta \cdot \frac{\partial H_s(p_i^s)}{\partial p_i^s} \cdot \frac{\partial p_i^s}{\partial \theta^s}\right]. \tag{10}$$

Convergency Dynamics: The teacher model f_θ^t evolves as a smoothed EMA of the student model f_θ^s, acting as a stable reference. **(a) Alignment with Teacher Predictions:** The gradient of $\mathcal{L}_{\text{UnCL}}$ encourages the student predictions p^s to align with the teacher predictions p^t, serving as a denoised target. **(b) Uncertainty Focus:** The exponentiated entropy-based weighting $\exp(\beta \cdot H_s(p_i^s)) + \exp(\beta \cdot H_t(p_i^t))$ focuses the optimization on regions with high uncertainty, such as lesion boundaries. **(c) Confidence Calibration:** The entropy regularization term $\frac{\beta}{N}\sum_{i=1}^{N} H_s(p_i^s)$ prevents overconfidence, promoting generalization.

Implications for Medical Image Segmentation: (a) Boundary Refinement: High-entropy regions (e.g., lesion boundaries) are emphasized during optimization to improve segmentation accuracy. **(b) Generalization:** The combination of alignment and uncertainty regularization ensures the model balances sharpness and coverage in segmentation.

2.1. Loss Surface Visualization

Fig. 2 illustrates the loss surface visualization over course of training on ISLES-22 and BraTS-19 datasets, which highlights the dynamics of UnCL's optimization process. The sharp spikes and valleys indicate regions of high uncertainty, corresponding to ambiguous areas such as subtle lesion boundaries or scattered lesions. By leveraging dual-entropy from both teacher and student models, UnCL dynamically modulates the consistency loss in these regions, mitigating the impact of noisy gradients and overconfident predictions. This results in a smoother and more balanced optimization trajectory, promoting stability and better segmentation accuracy in complex lesion distributions.

2.2. Grad-CAM Visualization

In this section, we visualize class-wise activation maps from the penultimate decoder layer of 3D-UNet to further gain interpretable insights into the inherent properties of UnCL loss as shown in Fig.3. The Grad-CAM visualizations effectively highlight DyCON's capability to focus on uncertain regions over the course of training. In early epochs

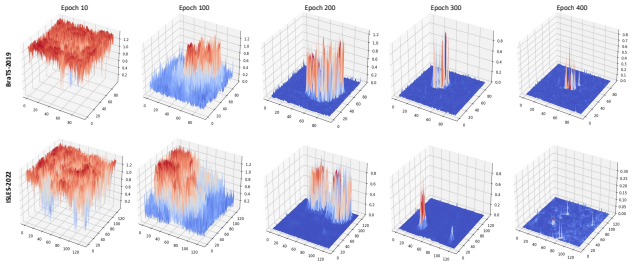


Figure 2. Loss landscape visualization of UnCL loss over training epochs for samples in ISLES'22 and BraTS'19 datasets. Notably, the loss surface is steadily converging toward better alignment and lower uncertainty across both datasets.

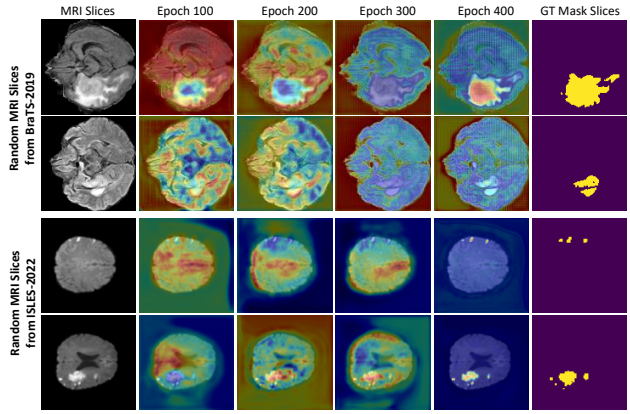


Figure 3. Grad-CAM visualizations over training epochs on ISLES'22 and BraTS'19 datasets. Progressively, the model's attention focuses on lesion regions, showing improved alignment with ground truth masks as training progresses.

(100), the heatmaps show a broader and less focused activation across the brain, indicating high uncertainty in identifying lesion regions, aligned to the ground truth masks. As training progresses (late 400 epochs), the activations become more refined and concentrated around the true lesion areas. This refinement reflects the model's robustness to accurately localize and segment subtle lesions, enhancing lesion boundary precision and overall segmentation accuracy.

3. More Ablation Study

3.1. The Essence of Dual-entropy in UnCL Loss

In this subsection, we analyze the role of dual-entropy (entropy from student and teacher models) in enhancing the performance of the UnCL loss. Dual-entropy introduces a principled mechanism for jointly capturing uncertainty from both models, which proves particularly effective in lesion segmentation under class imbalance and variability. Specifically, we trained DyCON with single-entropy from the teacher and student models without modifying other components. Table 1 highlights the performance improvement of using dual-entropy over single-entropy baselines on

Strategy	Dice (%)	IoU (%)	HD95↓	ASD↓
Teacher-Entropy	63.14	48.85	15.23	2.15
Student-Entropy	64.42	50.24	14.25	1.64
Dual-Entropy	65.75	51.20	13.30	0.76

Table 1. The effect of using single-entropy and dual-entropy in UnCL loss using 10% labeled data.

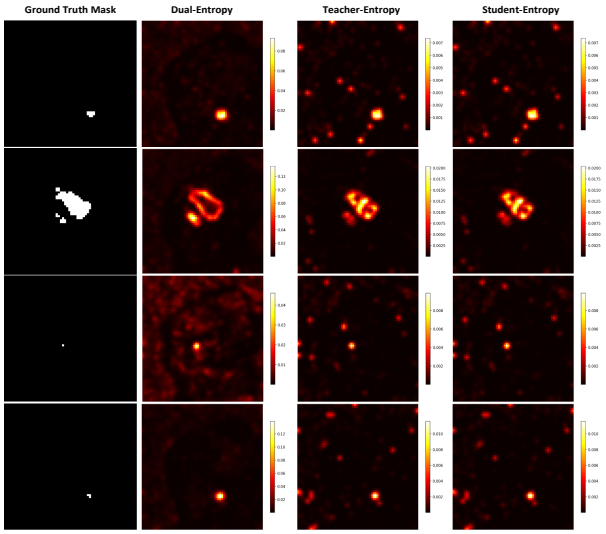


Figure 4. Illustration of uncertainty maps generated using dual-entropy and single-entropy strategies from the teacher and student models in UnCL Loss.

the ISLES-22 dataset. For example, the Dice score improves from 63.14% (teacher entropy only) and 64.42% (student entropy only) to 65.75% with dual-entropy. This result reflects the importance of leveraging uncertainties from both models.

Fig. 4 further demonstrates the impact of single-entropy and dual-entropy strategies when combined with the UnCL loss. The visualizations reveal that relying solely on either teacher or student entropy results in overconfident predictions in regions of high ambiguity, such as small lesions. It struggle to reduce uncertainty, exhibiting higher noise levels while still vaguely highlighting ground truth regions. This overconfidence occurs because a single-entropy approach fails to adequately capture the uncertainty inherent in such challenging areas. In contrast, the dual-entropy effectively reduces uncertainty over challenging lesion regions, as seen by the sharper focus around ground truth areas with minimal noise. This shows that dual-entropy provides more stable and precise guidance, improving the model's confidence and segmentation performance in ambiguous areas.

3.2. Effects of Hyperparameters in FeCL

To analyze the impact of the focal weighting factor γ and the top-k hard negative selection in FeCL, we conduct a de-

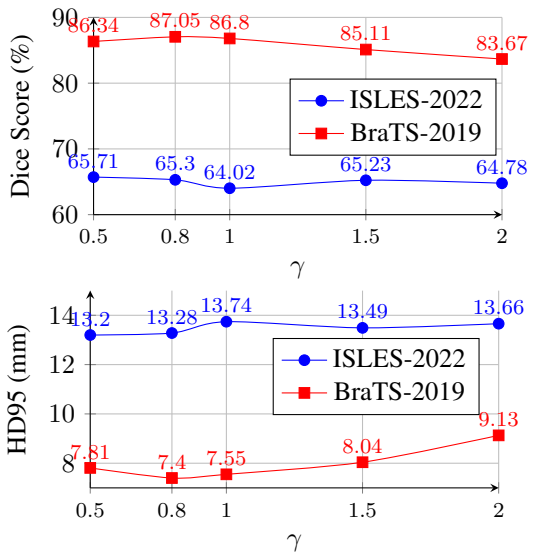


Figure 5. Performance comparison of Dice scores and HD95 values for ISLES'22 and BraTS'19 datasets across different λ values.

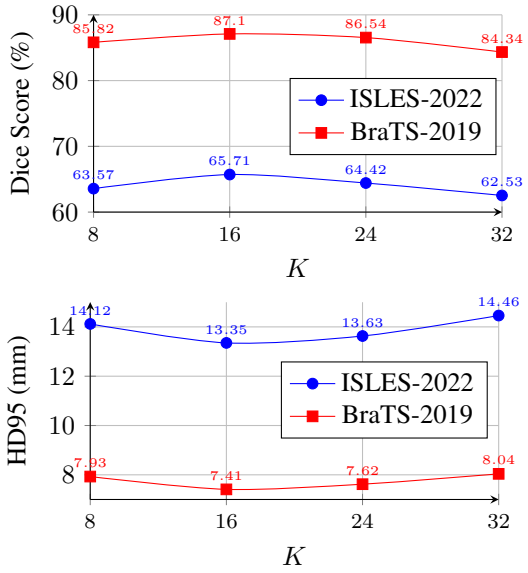


Figure 6. Performance comparison of Dice and HD95 values for ISLES-2022 and BraTS-2019 datasets across different K values.

tailed ablation study across two datasets, focusing on their respective roles in improving lesion segmentation performance. These parameters are crucial in addressing class imbalance and enhancing the model's ability to discriminate subtle lesion boundaries under challenging conditions.

Effect of Focus Controller γ: The parameter γ modulates the contribution of hard positive and negative patch pairs by emphasizing samples that are difficult to distinguish. A higher γ value increases the weight on hard-to-distinguish patch pairs, while lower values reduce the emphasis, treating all samples more uniformly. In this experiment, we vary $\gamma \in \{0.5, 0.8, 1.0, 1.5, 2.0\}$ and evaluate segmentation accuracy using Dice and HD95 metrics. All experiments are conducted on both datasets with 10% labeled data, keeping the other FeCL components constant. The results for various γ are reported in Fig. 5.

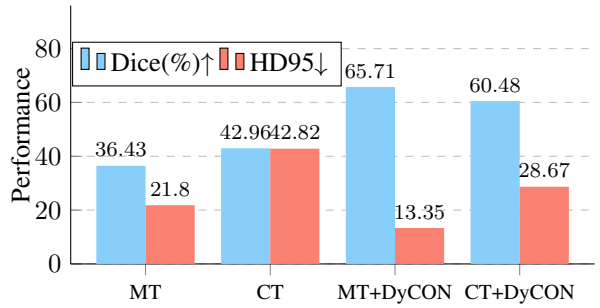
The focal weighting factor γ in FeCL demonstrates optimal performance at $\gamma = 0.8$ for BraTS'19 and $\gamma = 0.5$ for ISLES'22, achieving a balance between emphasizing hard positive and negative patch pairs and avoiding over-amplification of noise. Increasing γ beyond this value leads to marginal performance improvements, which indicates the model already effectively captures complex lesion characteristics with moderate γ, such as variability in size, shape, and spatial distribution. This phenomenon highlights the robustness of FeCL to lesion complexity in handling small, irregular, and scattered lesions. The results suggest that $\gamma = \{0.8, 0.5\}$ provides the optimal trade-off between emphasizing subtle lesion patterns and maintaining training stability. In this way, FeCL ensures consistent feature discrimination across diverse and challenging datasets.

Effect of Top-k Hard Negative Selection: The top-k parameter controls the number of hard negatives incorporated from the teacher model, enriching the negative sampling process in the contrastive loss. In this experiment, we test $K \in \{8, 16, 24, 32\}$ while keeping the optimal values of λ fixed for BraTS-2019 and ISLES-2022 datasets. For each setting, we measure Dice and HD95 to evaluate segmentation performance. Fig. 6 shows the Dice and HD95 scores of various K values. The key observations reveal that increasing k up to $k = 16$ enhances the diversity of negative samples, improving feature separability and segmentation performance, as evidenced by a Dice score increase from 63.57% (k=8) to 65.71% (k=16) on ISLES-22. Beyond k=16, the performance plateaus or slightly decreases across both datasets due to the inclusion of redundant or overly similar negatives, which introduces noise into the loss.

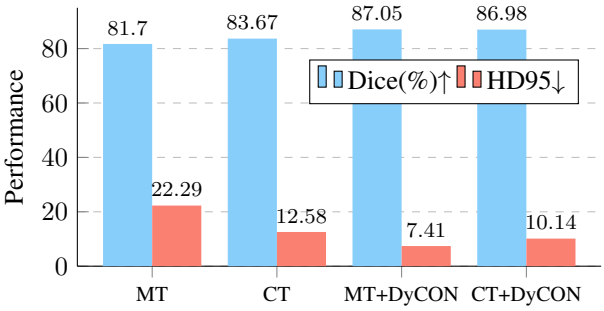
3.3. Integrating DyCON with SSL Frameworks

To demonstrate the versatility and effectiveness of DyCON, we integrate it into two representative semi-supervised learning frameworks: Mean-Teacher (MT)[10] and Co-training (CT)[3]. These frameworks utilize distinct mechanisms for leveraging unlabeled data, providing a robust foundation for evaluating the adaptability of DyCON.

In this regard, MT leverages a student-teacher paradigm where DyCON's UnCL improves global consistency by dynamically weighting uncertain regions, while FeCL enhances local feature discrimination for nuanced lesion segmentation, addressing class imbalance effectively. Similarly, Co-Training benefits from DyCON by using UnCL to align predictions from independent sub-networks through dual-entropy, and FeCL to refine local features, capturing subtle lesion details. DyCON significantly improves the performance of both frameworks on ISLES'22 and BraTS'19 datasets.



(a) ISLES'22 Results.



(b) BraTS'19 Results.

Figure 7. Comparative performance (Dice (%) and HD95) when integrating DyCON into MT and CT frameworks with 10% labeled data. DyCON enhances segmentation accuracy consistently across ISLES'22 and BraTS'19 datasets without modifying the underlying architectures or training pipelines.

Our experiments demonstrate that DyCON consistently improves both MT and CT frameworks, as shown in Fig. 7. By enhancing global and local feature consistency, DyCON achieves superior segmentation results without requiring modifications to the underlying architectures. These findings highlight DyCON's adaptability and potential for broad adoption in SSL-based medical image segmentation.

4. More Visualization Results

Fig. 8 presents brain tumor and lesion segmentation visualizations from BraTS'19 and ISLES'22 datasets. Moreover, Fig. 9 illustrates left atrium organ segmentation visualizations from LA dataset. These results demonstrate that, DyCON persistently delivers more accurate segmentation of complex lesions and organs compared to the SOTA methods across all datasets. Closely note that, in each of the visualizations the False Negatives (blues) are produced due to invariance of the lesion or tumor boundaries with the healthy tissues. In this case, DyCON has achieved a remarkable delineation of these challenging tumors and lesions regardless of their morphological difficulties.

5. Experiments on Multi-Organ Segmentation

To further assess the versatility and robustness of DyCON, we conducted additional experiments on the BTCV

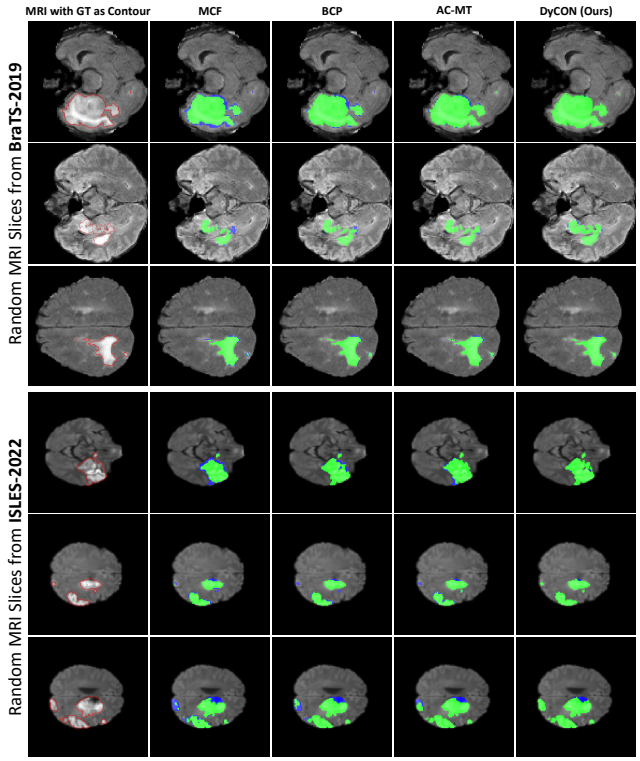


Figure 8. Comparison of brain tumor and lesion segmentation with SOTA methods. The comparison results are illustrated in terms of True Positives (green), and False Negatives (blue) in segmenting challenging lesions of MRI scans. The baseline methods struggle with detecting subtle lesion boundaries across both datasets, resulting in higher instances of FNs (blue) while DyCON effectively reduces these errors, leading to a significant number of correctly identified lesion voxels (green).

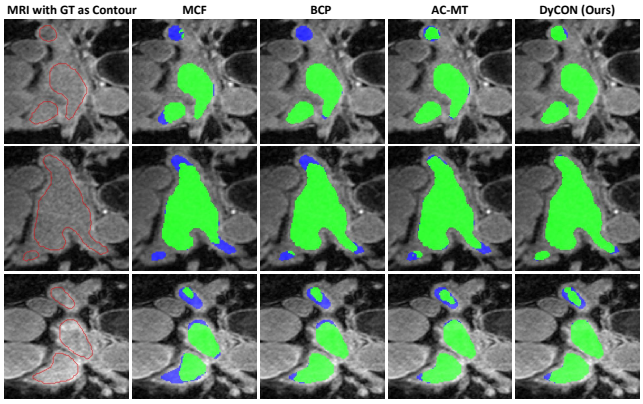


Figure 9. Comparison of challenging left atrium organ segmentation results with SOTA methods using 10% labeled data on LA dataset. DyCON consistently produces more accurate segmentation of left atrial organs.

(Synapse) dataset for multi-organ segmentation, following the evaluation protocol established by existing SOTA methods such as DHC [5], and GA [8] combined with Magic-

Methods	Average		Average Dice of each class using 20% labeled data												
	Dice	ASD	Sp	RK	LK	Ga	Es	Li	St	Ao	IVC	PSV	PA	RAG	LAG
DHC [5]	48.6	10.7	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
GA+MagicNet [8]	68.4	**3.1**	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**
DyCON+GA+MagicNet	**69.5**	3.5	**88.5**	85.9	**92.6**	49.6	**60.9**	**90.6**	58.7	**85.7**	81.1	62.4	**51.9**	**55.9**	39.64

Table 2. DyCON outperforms all methods in some organ classes while it remains comparable with other classes.

Net [2] (GA+MagicNet). The BTCV dataset encompasses segmentation tasks across various abdominal organs, each posing unique challenges, including considerable variability in organ size, shape, boundary ambiguity, and proximity to neighboring structures. Therefore, we integrated our proposed losses, UnCL and FeCL into the GA+MagicNet framework, creating the DyCON+GA+MagicNet variant. Table 2 illustrates the comparative performance of DyCON against the previous SOTA DHC and GA+MagicNet across thirteen different organ classes.

DyCON significantly surpasses GA+MagicNet, achieving new SOTA average Dice performance of **69.5%**, attaining a +1.1% improvement. Specifically notable improvements are observed in challenging organ classes characterized by high uncertainty and ambiguous boundaries, such as the Spleen (Sp: **88.5%** vs. 81.4%), Left Kidney (LK: **92.6%** vs. 90.8%), Gallbladder (Ga: **49.6%** vs. 33.5%), Esophagus (Es: **60.9%** vs. 53.3%), and Aorta (Ao: **85.7%** vs. 79.1%). These gains underscore DyCON's capability to dynamically emphasize uncertain and ambiguous boundary regions, enhancing the accuracy of segmentation outcomes in anatomically complex areas.

Although DyCON slightly increases the ASD from 3.1mm to 3.5mm compared to GA+MagicNet, the improvement in Dice scores across critical organs indicates that DyCON prioritizes precision in delineating organ boundaries over minimal surface distance discrepancies, beneficially balancing boundary sharpness and general segmentation robustness. Overall, these results reinforce DyCON's effectiveness as a versatile semi-supervised learning framework, not only for brain lesion segmentation but also extending its applicability and superior performance to complex multi-organ segmentation tasks.

References

[1] Yunhao Bai, Duowen Chen, Qingli Li, Wei Shen, and Yan Wang. Bidirectional copy-paste for semi-supervised medical image segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11514–11524, 2023. 1

[2] Duowen Chen, Yunhao Bai, Wei Shen, Qingli Li, Lequan Yu, and Yan Wang. Magicnet: Semi-supervised multi-organ segmentation via magic-cube partition and recovery. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 23869–23878, 2023. 6

[3] Xiaokang Chen, Yuhui Yuan, Gang Zeng, and Jingdong Wang. Semi-supervised semantic segmentation with cross pseudo supervision. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2613–2622, 2021. 4

[4] Landman et al. Miccai multi-atlas labeling beyond the cranial vault–workshop and challenge. In *MICCAI*, 2020. 1

[5] Wang et al. Dhc: Dual-debiased heterogeneous co-training framework for class-imbalanced ssmis. In *MICCAI*, 2023. 5, 6

[6] Moritz R Hernandez Petzsche, Ezequiel de la Rosa, Uta Hanning, Roland Wiest, Waldo Valenzuela, Mauricio Reyes, Maria Meyer, Sook-Lei Liew, Florian Kofler, Ivan Ezhov, et al. Isles 2022: A multi-center magnetic resonance imaging stroke lesion segmentation dataset. *Scientific data*, 9(1): 762, 2022. 1

[7] Bjoern H Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, et al. The multimodal brain tumor image segmentation benchmark (brats). *IEEE transactions on medical imaging*, 34(10):1993–2024, 2014. 1

[8] Wenbo Qi, Jiafei Wu, and SC Chan. Gradient-aware for class-imbalanced semi-supervised medical image segmentation. In *European Conference on Computer Vision*, pages 473–490. Springer, 2024. 5, 6

[9] Holger R Roth, Le Lu, Amal Farag, Hoo-Chang Shin, Jiamin Liu, Evrim B Turkbey, and Ronald M Summers. 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*, pages 556–564. Springer, 2015. 1

[10] Antti Tarvainen and Harri Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *Advances in neural information processing systems*, 30, 2017. 4

[11] Zhaohan Xiong, Qing Xia, Zhiqiang Hu, Ning Huang, Cheng Bian, Yefeng Zheng, Sulaiman Vesal, Nishant Ravikumar, Andreas Maier, Xin Yang, 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. 1

[12] Zhe Xu, Yixin Wang, Donghuan Lu, Xiangde Luo, Jiangpeng Yan, Yefeng Zheng, and Raymond Kai-yu Tong. Ambiguity-selective consistency regularization for mean-teacher semi-supervised medical image segmentation. *Medical Image Analysis*, 88:102880, 2023. 1