

# Neighbor Does Matter: Density-Aware Contrastive Learning for Medical Semi-supervised Segmentation

Feilong Tang<sup>1\*</sup>, Zhongxing Xu<sup>1\*</sup>, Ming Hu<sup>1</sup>, Wenxue Li<sup>1</sup>, Peng Xia<sup>3</sup>,  
Yiheng Zhong<sup>2</sup>, Hanjun Wu<sup>2</sup>, Jionglong Su<sup>2†</sup>, Zongyuan Ge<sup>1†</sup>

<sup>1</sup>AIM Lab, Faculty of IT, Monash University

<sup>2</sup>Xi'an Jiaotong-Liverpool University

<sup>3</sup>UNC-Chapel Hill

Feilong.Tang@monash.edu

## Abstract

In medical image analysis, multi-organ semi-supervised segmentation faces challenges such as insufficient labels and low contrast in soft tissues. To address these issues, existing studies typically employ semi-supervised segmentation techniques using pseudo-labeling and consistency regularization. However, these methods mainly rely on individual data samples for training, ignoring the rich neighborhood information present in the feature space. In this work, we argue that supervisory information can be directly extracted from the geometry of the feature space. Inspired by the density-based clustering hypothesis, we propose using feature density to locate sparse regions within feature clusters. Our goal is to increase intra-class compactness by addressing sparsity issues. To achieve this, we propose a *Density-Aware Contrastive Learning (DACL)* strategy, pushing anchored features in sparse regions towards cluster centers approximated by high-density positive samples, resulting in more compact clusters. Specifically, our method constructs density-aware neighbor graphs using labeled and unlabeled data samples to estimate feature density and locate sparse regions. We also combine label-guided co-training with density-guided geometric regularization to form complementary supervision for unlabeled data. Experiments on the Multi-Organ Segmentation Challenge dataset demonstrate that our proposed method outperforms state-of-the-art methods, highlighting its efficacy in medical image segmentation tasks.

## Introduction

Multi-organ segmentation in medical images is essential for many clinical applications, such as computer-aided intervention (Qu et al. 2024; Tang et al. 2023; Wang et al. 2022; Xu et al. 2024b; Ye et al. 2022, 2024b; Chen et al. 2023b; Chen, Pan, and Xia 2023; Chen et al. 2024), surgical navigation (Zheng et al. 2024; Hu et al. 2024, 2025; Tang et al. 2024a), and radiotherapy (Mofid et al. 2024). However, complex morphological structures of multiple organs, low contrast of soft tissues, and blurred boundaries between adjacent organs make dense annotation of multi-organ images

extremely expensive and challenging. The scarcity of annotations has become a bottleneck that limits the performance of deep learning-based multi-organ segmentation models. As a solution, semi-supervised learning (SSL) (Huang et al. 2024; Chi et al. 2024; Liu et al. 2024; Ye et al. 2024a) enhances labeling efficiency by leveraging a limited amount of labeled data alongside a large amount of unlabeled data. SSL aims to extract effective training signals from unlabeled images to enhance model generalization.

To handle unlabeled data, most existing SSL methods are based on the common density clustering assumption (Wang et al. 2023), *i.e.*, samples located in the same high-density area are likely to belong to the same class, and the decision boundary should be located in the low-density area. In other words, by achieving inter-class separation based on density, semi-supervised models can output highly confident predictions without the supervision of corresponding labels. Current multi-organ segmentation methods (Wong et al. 2024) mainly focus on separating low-density areas. These methods make invariant predictions to perturbations around each unlabeled data point or encourage nearby samples to make the same predictions, including performing consistency regularization (Chen et al. 2023a) or generating pseudo-labels (Basak and Yin 2023). Despite making some progress, current methods mainly utilize single data samples and ignore the rich neighborhood information for learning more discriminative features. Therefore, these methods face challenges with regards to data scarcity and feature sparsity.

Recent advances in manifold learning have established two properties (Dosovitskiy et al. 2020): (1) allowing similar features to be closer to each other, while dissimilar features move further apart; (2) providing superior manifold metrics to measure discrepancies with different statistical and geometrical properties. Based on these, we mine effective neighborhood information from the geometry of the feature space. As shown in Fig. 1 (a), learning that incorporates neighborhood information can effectively capture the complex shapes of various organs (*e.g.*, large liver and small pancreas) and produce more compact and discriminative embeddings (Fig. 1 (b)). In contrast, directly using pixel-level features in contrastive learning methods can misidentify some organ pixels as background, particularly

\*These authors contributed equally.

†Corresponding author: Jionglong Su; Zongyuan Ge  
Copyright © 2025, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

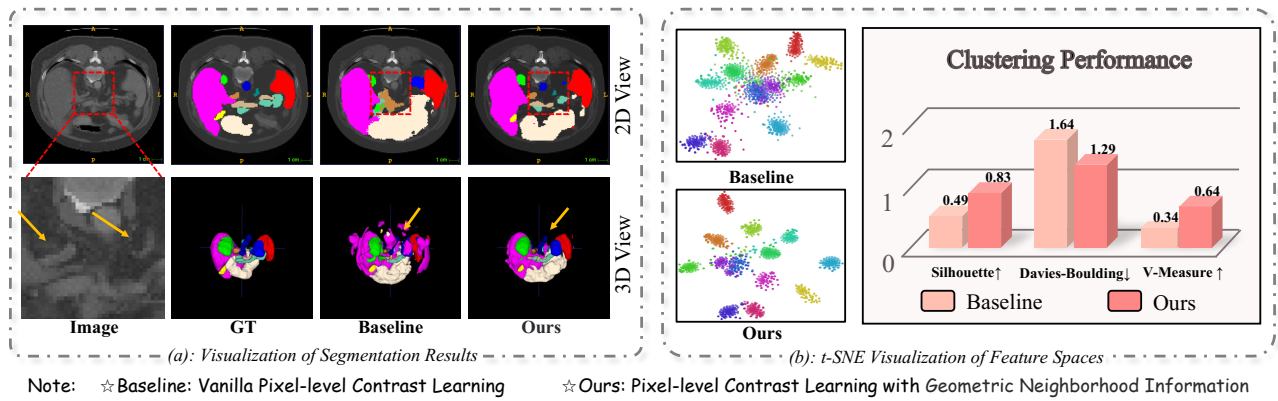


Figure 1: Left (a): Qualitative comparison of abdominal organs. As observed in the red boxes, the contrastive learning method that incorporates geometric neighborhood information achieves a more plausible appearance and a more complete structure compared to the conventional SSL method. Right (b): t-SNE (Van der Maaten and Hinton 2008) reveals that our method produces more compact class clusters with clearer separation between each class compared to the baseline. The clustering performance is evaluated using Silhouette (Rousseeuw 1987), Davies-Boulding (Davies and Bouldin 1979) and V-Measure (Rosenberg and Hirschberg 2007) score. Our method outperforms the baseline in clustering performance metrics.

for smaller organs (Fig. 1 (a)).

In this work, we propose a learning strategy called **Density-Aware Contrastive Learning (DACL)**. Within the feature clusters of both labeled and unlabeled data, we construct density-aware neighbor graphs by measuring local density based on the average distance between the target feature and its nearest neighbors. To achieve robust perception, we introduce a category memory bank to overcome the limitations of small batches, enabling global estimation of intra-class density in a feature-to-bank manner. Simultaneously, we use density-aware maps to comprehensively estimate the final density and locate sparse regions. We also propose **Soft Density-Guided Contrastive Learning (SDCL)** to explicitly shrink sparse regions for more compact clusters. Specifically, sparsity search is conducted within class feature clusters using density-aware neighbor graphs to locate low-density features as anchors. Meanwhile, features from dense areas are selected as positive keys to approximate cluster centers. Subsequently, we calculate the positiveness score of each low-density feature with the corresponding positive key and incorporate it into the contrastive loss calculation. A lower score indicates greater sample distance. Through contrastive learning, anchors are drawn closer to positive keys, enhancing feature clustering.

We evaluate our proposed method on Automatic Cardiac Diagnosis Challenge (ACDC) (Bernard et al. 2018) and the Synapse multi-organ segmentation (Landman et al. 2015) under various semi-supervised settings, where our method achieves state-of-the-art performances. The main contributions of our work are summarized as follows:

- We propose a Density-Aware Contrastive Learning (DACL) strategy that focuses on effectively utilizing the geometric structure of the feature space to address the challenge of data scarcity and feature sparsity in semi-supervised multi-organ segmentation.
- We propose mining effective neighborhood information

in the feature space and integrating it into contrastive learning, consequently enhancing intra-class compactness in the feature space.

- Experimental results on benchmark datasets demonstrate that our method significantly improves upon the efficacy of previous state-of-the-art methods.

## Related Work

**Semi-supervised Medical Image Segmentation.** Recent efforts in semi-supervised segmentation have been focused on incorporating unlabeled data into CNNs (Tang et al. 2024b; Zhao et al. 2024a; Liu et al. 2022; Wu et al. 2024b,a; Yang et al. 2023), which can be broadly categorized into five groups: self-training (Chi et al. 2024), co-training (Ma et al. 2024), deep adversarial learning (Xu et al. 2024a) and self-ensembling ( $\pi$ -model (Deng et al. 2024) and Mean-Teacher model (Zhao et al. 2024b)). For instance, Liu et al. (Liu et al. 2023) achieved Deep Co-Training by learning two classifiers on two views. Liu et al. (Liu, Li, and Yuan 2024) designed a Deep Adversarial Network, which enforced the segmentation of unlabeled data to be similar to the labeled ones. Yu et al. (Yu et al. 2019) proposed an Uncertainty-Aware Mean-Teacher, which enables the student to learn from reliable targets. Nevertheless, considering most regions with low uncertainty tend to be easy for a segmentation network to make correct predictions (Jiang et al. 2024), low prediction quality for hard regions such as boundaries remains a challenge for performance improvement (Wu et al. 2022a). Therefore, we aim to improve the general quality of the learning target directly based on a co-training framework.

**Contrastive Learning.** Image-level contrastive learning has achieved significant results in self-supervised learning. Typical methods such as SimCLR (Chen et al. 2020) have proposed pipelines for learning augmentation-invariant representations, which outperform traditional supervised pre-training methods (Wu, Zhuang, and Chen 2024). Building

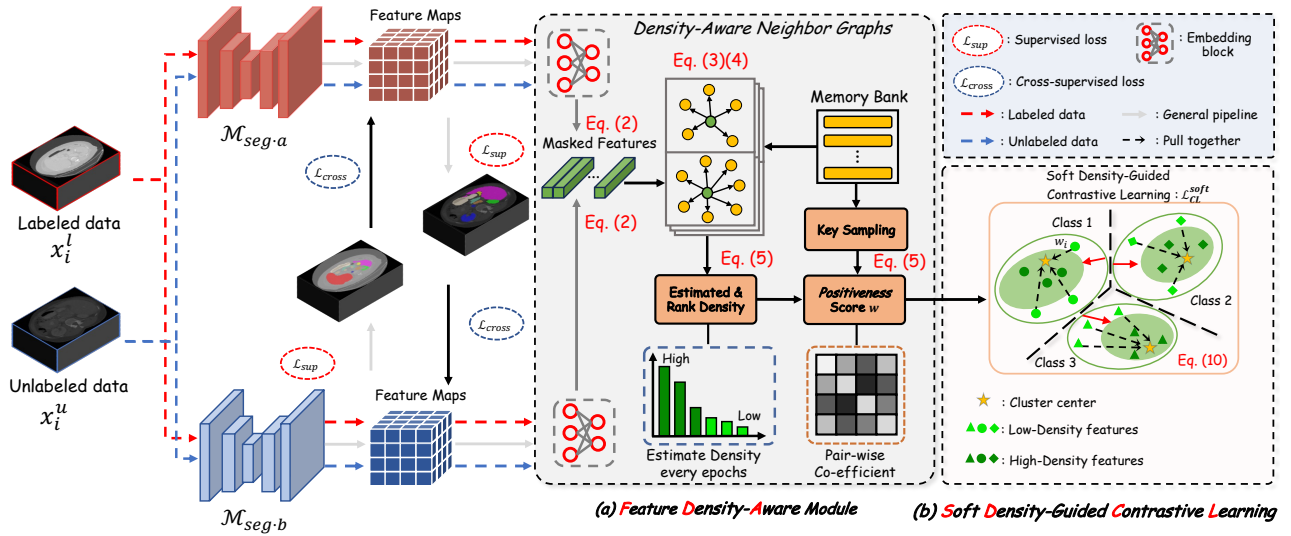


Figure 2: Overview of the proposed unified learning framework. (a) shows the feature density-aware module. (b) shows our Soft Density-guided Contrastive Learning strategy. For labeled image  $x_i^l$ , we apply the commonly-used supervised loss  $\mathcal{L}_{sup}$  to update model parameters. For unlabeled image  $x_i^u$ , the model is optimized with the cross-supervised consistency loss  $\mathcal{L}_{cross}$  and one manifold constraint  $\mathcal{L}_{CL}^{soft}$  to explore the geometry of the feature space.

on this, we guide pixel-level feature learning by incorporating label information and transforming feature density into training signals, enhancing model performance.

## Methodology

Fig. 2 illustrates the overall architecture of our proposed DACL framework, which builds upon the co-teaching paradigm. This framework aims to mine effective supervision from unlabeled images by leveraging a limited set of annotations. It employs a density-guided contrastive learning strategy to regularize the structure of clusters in the feature space. Within each mini-batch, we evaluate the density to locate features in sparse regions as anchors, while selecting features from dense regions as positive keys approximating the cluster center. By pushing low-density anchors towards the approximated cluster center, we aim to minimize contrastive loss, dynamically shrink cluster volume, and increase intra-cluster compactness. Furthermore, we softly calculate the positiveness score for each low-density feature with its corresponding positive key and incorporate it into the contrastive loss calculation.

### Co-training Framework with Geometric Loss

In a general semi-supervised learning problem, assume that the whole dataset consists of  $N_L$  labeled samples  $\{(x_i^l, y_i)\}_{i=1}^{N_L}$  and  $N_U$  unlabeled samples  $\{x_i^u\}_{i=1}^{N_U}$ , where  $x_i \in \mathbb{R}^{H \times W}$  denotes an image and  $y_i \in \mathbb{R}^{N \times H \times W}$  is the ground-truth annotation with  $N$  classes (including background).  $W$  and  $H$  respectively represent the width and height of the input image. We adopt a co-teaching paradigm such as MCNet (Wu et al. 2021) to enforce two segmentation models  $\mathcal{M}_{seg-a}$  and  $\mathcal{M}_{seg-b}$  to predict the prediction results of its counterpart. To leverage limited annotations and

extract effective supervisory information from unlabeled images, we identify sparse region features as anchors and use representative features as positive keys. Our objective is to minimize the contrastive loss by pushing low-density anchors towards the cluster center, enhancing cluster compactness. A unified loss function  $\mathcal{L}$  optimizes the model:

$$\mathcal{L} = \mathcal{L}_{sup} + \lambda_{cross} \mathcal{L}_{cross} + \lambda_{CL} \mathcal{L}_{CL}^{soft}, \quad (1)$$

where  $\mathcal{L}_{CL}^{soft}$  is the soft density-guided contrastive loss, and  $\mathcal{L}_{sup}$  and  $\mathcal{L}_{cross}$  are the supervised loss and cross-supervised loss, respectively.  $\lambda_{cross}$  and  $\lambda_{CL}$  are used to balance the relationship between losses (specifically, we choose  $\lambda_{cross} = 1$  and  $\lambda_{CL}$  as Gaussian warming up function  $\lambda(t) = 0.1 \times e^{-5(1-t/t_{max})^2}$ , where  $t$  is the current training step and  $t_{max}$  is the maximum training step). During training, a batch of mixed labeled and unlabeled samples are fed into the network. The supervised loss  $\mathcal{L}$  is applied only to labeled data, while all samples are utilized to construct cross-supervised learning. Refer to the Appendix for details on  $\mathcal{L}_{sup}$  and  $\mathcal{L}_{cross}$  losses.

### Feature Density-Aware Module

Before analyzing the density of features within each class, we first approximate the distribution of class features. The features of classes within images or batches can only offer a partial perspective. For each image  $x$ , the input to encoder extract feature maps  $f \in \mathbb{R}^{W \times H \times D}$  with  $D$  channels and the feature maps are mapped to the projection space  $j = z(f)$  by a projection head  $z$  for instance prototyping. Each instance embedding represents the regional semantics of the categories observed in  $x$  based on the corresponding ground truth or pseudo labels  $M_n$ . Specifically, for the  $n$ -th category, its projected features is summarized to a vector

$v_n \in \mathbb{R}^D$  by masked average pooling (MAP) (Siam, Oreshkin, and Jagersand 2019):

$$v_n = \frac{\sum_{x=1}^W \sum_{y=1}^H \mathbf{T}_n(x, y) * j(x, y)}{\sum_{x=1}^W \sum_{y=1}^H \mathbf{T}(x, y)}, \quad (2)$$

where  $\mathbf{T}_n = \mathbf{1}(M_n > \phi) \in \{0, 1\}^{W \times H}$  is a binary mask, emphasizing only strongly-activated pixels of class  $n$  in its activation map.  $\mathbf{1}(\cdot)$  is an indicator function, and  $\phi$  is a hyper-parameter denoting the threshold of the reliability score.  $*$  represents element-wise multiplication. Here,  $v_n$  is compact and lightweight, allowing feasible exploration of its relationships with numerous other samples.

We propose a dynamic memory bank to aggregate class features from the entire dataset, enabling more comprehensive modeling of class features. Specifically, we extract embeddings using two segmentation models. The embeddings  $v \in \mathbb{R}^{B \times D}$  are projection features with batch size  $B$ . These embeddings are stored in the memory bank. For each class  $n$ , the feature memory  $\mathcal{G}_n$  is updated following a First-In, first-Out strategy while maintaining a fixed size  $L$ . In this module, we adopt a feature-to-bank approach to construct density-aware graphs. Within a mini-batch, the features for class  $n$  are denoted as  $\mathcal{V}_n = \{v_i\}_{i=1}^B$ . As shown in Fig. 2 (a), we measure the density of each  $v$  by constructing a  $k$ -nearest neighbor graph  $\text{NN}(v) = \{v_j\}_{j=1}^k$  within  $\mathcal{G}_n \cup \mathcal{V}$ . The density  $d_v$  is computed as the average cosine similarity between  $v$  and its  $k$  neighbors:

$$d_v = \frac{1}{|\text{NN}(v)|} \sum_{r \in \text{NN}(v)} \tilde{\mathbf{f}}_r^T \tilde{\mathbf{f}}_v, \quad (3)$$

where  $\tilde{\mathbf{f}}_r, \tilde{\mathbf{f}}_v$  are the L2-normalized feature embedding of  $r$  and  $v$ . We compute densities  $\{d_v\}$  to identify the low-density and high-density features of  $v$ . Meanwhile, batch features with density values  $\{(v, d_v)\}$  are used to update the corresponding memory bank. Since density estimation with a few neighbors tends to focus on local regions, true cluster centers fail to be captured. To achieve a more accurate estimation, we introduce multi-scale nearest neighbor graphs:

$$\text{NN}_{\text{multi-scale}} = \{\text{NN}_{k_i}\}_{i=1}^N \quad (4)$$

for the target feature vector  $v$ , where  $k_1 < k_2 < \dots < k_n$  and  $n$  denotes the number of different neighbor scales used. We compute density values  $d^1, d^2, \dots, d^n$  using different graphs according to Eq. 3. The final estimation is calculated as the average density value:

$$d_v = \frac{1}{N} \sum_{i=1}^N d_v^i. \quad (5)$$

### Soft Density-Guided Contrastive Learning

Here, we introduce Soft Density-Guided Contrastive Learning for complementary supervision on unlabeled data. The feature vectors in each mini-batch are ranked based on their density values, and those with low-density features are selected as anchors. Meanwhile, features are extracted from a memory bank as positive keys, and the positiveness scores between low-density features and positive keys are computed and incorporated into the contrastive loss.

**Low-density (sparse) Anchors Sampling.** Density values are used to measure the representativeness of features within a class, indicating the ‘‘hardness’’ of training for each feature. Features with low-density values indicate less representativeness in the class, *i.e.*, insufficient training for the current model state or being trapped in local optima. Therefore, addressing sparse regions of features will benefit the learning process. For all features of a specific class within a mini-batch, we first estimate their densities based on the corresponding memory bank. They are then ranked by their density values. For each class  $n$ , we select the  $N_q$  features with the lowest densities as anchors and store them in  $\mathcal{M}^n$ :

$$\mathcal{M}^n = \left\{ m_i^n : i \in \arg \min_{m_i^n \in \mathcal{V}_n} (d_{m_i^n}), \text{top } N_q \right\}, \quad (6)$$

where  $d_{m_i^n}$  denotes the density value of feature  $m_i^n$  and  $\arg \min$  selects the indices of the  $N_q$  lowest density values in the current mini-batch.

**High-density (dense) Positive Key Sampling.** We approximate the class center using only high-density features to guide the anchors. During training, a total of  $N_p^+$  high-density positive keys are sampled from both the mini-batch and the memory bank. While batch features provide fresh and intra-object contrasts, the global memory bank exhibits more comprehensive and diverse category patterns. Therefore, assuming complementarity between these two sets of features can provide robust center estimation. We select the  $\frac{1}{2}N_p^+$  positive keys with the highest density from the mini-batch and the other  $\frac{1}{2}N_p^+$  from the memory bank for class  $n$  and store them in  $\mathcal{P}_{\text{type}}^{n,+}$  as:

$$\mathcal{P}^{n,+} = \left\{ m_i^n : i \in \arg \max_{\mathbf{p}_i^{n,+} \in \mathcal{V}_n \cup \mathcal{G}_n} (d_{m_i^n}), \text{top } \frac{1}{2}N_p^+ \right\}, \quad (7)$$

where  $\theta_{\text{global/local}}^n$  denotes the  $N_p^+$ -th highest intra-class density value in the current batch/memory bank. Next, we calculate the category cluster center as:

$$\mathbf{p}_{\text{center}}^{n,+} = \frac{1}{|\mathcal{P}^{n,+}|} \sum_{\mathbf{p}^+ \in \mathcal{P}^{n,+}} \mathbf{p}^+. \quad (8)$$

**Negative Key Sampling.** We randomly sample  $N_p^-$  out-of-class features for each class  $n$  from the anchors in the current batch, forming  $\mathcal{P}^{n,-} = \{\mathbf{p}^-\}$  to create negative contrast pairs.

**Positiveness Scores.** We have designed pair-wise positiveness scores to softly measure (in a non-binary form) the relevance between low-density features and positive keys within the same class. For sample pairs  $(m_i^n, \mathbf{p}_{\text{center}}^{n,+})$ , we use the following positiveness score  $w_i$  to adjust the contribution of low-density features in contrastive learning:

$$w_i = \frac{1}{\gamma_i} \text{softmax} \left[ l_1(m_i^n) \times l_2(\mathbf{p}_{\text{center}}^{n,+})^\top \right], \quad m_i^n \in \mathcal{M}^n, \quad (9)$$

where  $l_1(\cdot)$  and  $l_2(\cdot)$  are parameter-free identity mapping layers in feature transformation.  $\gamma_i$  is a scaling factor to adjust  $w_i$ .

Methods	ACDC database					Synapse dataset				
	Scans used		Metrics			Scans used		Metrics		
	Labeled	Unlabeled	Dice(%) $\uparrow$	JC(%) $\uparrow$	ASD(voxel) $\downarrow$	Labeled	Unlabeled	Dice(%) $\uparrow$	JC(%) $\uparrow$	ASD(voxel) $\downarrow$
V-Net	3(5%)	0	43.30 $\pm$ 8.6	37.01 $\pm$ 6.90	12.41 $\pm$ 0.8	2(10%)	0	12.37 $\pm$ 9.7	10.29 $\pm$ 6.8	68.14 $\pm$ 6.9
V-Net	7(10%)	0	79.62 $\pm$ 1.5	68.11 $\pm$ 1.88	2.94 $\pm$ 1.2	4(20%)	0	32.13 $\pm$ 6.4	27.58 $\pm$ 3.4	35.84 $\pm$ 2.2
V-Net	70(All%)	0	91.54 $\pm$ 0.7	84.87 $\pm$ 0.36	0.90 $\pm$ 0.6	20(All%)	0	62.09 $\pm$ 1.2	58.43 $\pm$ 2.6	10.28 $\pm$ 3.9
CPS			57.67 $\pm$ 0.5	44.02 $\pm$ 0.6	3.75 $\pm$ 1.8			18.07 $\pm$ 1.2	14.42 $\pm$ 0.9	57.64 $\pm$ 1.8
MC-Net			61.56 $\pm$ 1.9	49.45 $\pm$ 0.8	2.38 $\pm$ 0.4			21.96 $\pm$ 6.1	19.42 $\pm$ 2.6	55.42 $\pm$ 4.6
URPC			54.69 $\pm$ 1.2	44.78 $\pm$ 0.9	3.47 $\pm$ 0.3			26.37 $\pm$ 1.5	23.62 $\pm$ 1.2	53.95 $\pm$ 9.3
SS-Net			64.12 $\pm$ 1.7	54.32 $\pm$ 0.7	2.03 $\pm$ 1.4			17.5 $\pm$ 3.0	15.73 $\pm$ 0.8	66.17 $\pm$ 8.0
BCP			65.92 $\pm$ 1.2	53.98 $\pm$ 0.8	2.15 $\pm$ 0.8			20.91 $\pm$ 5.9	18.64 $\pm$ 0.7	62.33 $\pm$ 8.5
CAML	3(5%)	67(95%)	66.15 $\pm$ 1.4	55.01 $\pm$ 0.5	2.09 $\pm$ 0.4	2(10%)	18(90%)	19.68 $\pm$ 6.3	17.54 $\pm$ 0.8	60.72 $\pm$ 9.3
DHC			67.19 $\pm$ 0.9	55.65 $\pm$ 1.3	2.01 $\pm$ 0.7			25.64 $\pm$ 0.9	21.43 $\pm$ 1.1	56.82 $\pm$ 4.0
Cross-ALD			80.61 $\pm$ 1.7	67.83 $\pm$ 1.2	1.9 $\pm$ 0.4			26.79 $\pm$ 0.5	22.13 $\pm$ 0.8	54.25 $\pm$ 2.2
ACTION++			88.5 $\pm$ 0.9	74.32 $\pm$ 1.4	0.72 $\pm$ 0.3			27.41 $\pm$ 0.9	24.37 $\pm$ 0.4	53.69 $\pm$ 1.2
<b>DACL (Ours)</b>			<b>90.12<math>\pm</math>0.4</b>	<b>76.12<math>\pm</math>1.6</b>	<b>0.55<math>\pm</math>0.2</b>			<b>29.88<math>\pm</math>0.7</b>	<b>26.54<math>\pm</math>1.2</b>	<b>51.32<math>\pm</math>1.4</b>
CPS			85.71 $\pm$ 0.9	70.53 $\pm$ 0.6	4.00 $\pm$ 0.9			31.46 $\pm$ 2.5	28.13 $\pm$ 2.0	40.74 $\pm$ 6.9
MC-Net			86.44 $\pm$ 1.7	72.64 $\pm$ 1.5	1.84 $\pm$ 0.4			32.46 $\pm$ 3.3	29.64 $\pm$ 1.7	43.21 $\pm$ 7.1
URPC			83.37 $\pm$ 0.2	69.31 $\pm$ 0.8	1.56 $\pm$ 0.2			25.68 $\pm$ 5.1	23.56 $\pm$ 1.2	72.74 $\pm$ 15.5
SS-Net			86.78 $\pm$ 0.7	75.12 $\pm$ 1.3	1.45 $\pm$ 0.2			35.08 $\pm$ 0.6	32.47 $\pm$ 1.3	50.81 $\pm$ 6.5
BCP			87.62 $\pm$ 0.5	75.86 $\pm$ 1.2	1.46 $\pm$ 0.3			35.57 $\pm$ 2.9	31.65 $\pm$ 1.5	36.39 $\pm$ 3.5
CAML	7(10%)	63(90%)	88.48 $\pm$ 0.2	76.12 $\pm$ 1.0	1.35 $\pm$ 0.3	4(20%)	16(80%)	37.35 $\pm$ 2.6	34.61 $\pm$ 1.2	37.42 $\pm$ 2.8
DHC			87.34 $\pm$ 0.6	75.63 $\pm$ 1.0	1.37 $\pm$ 0.4			39.61 $\pm$ 3.9	36.45 $\pm$ 1.8	34.73 $\pm$ 2.6
Cross-ALD			87.52 $\pm$ 0.9	76.42 $\pm$ 1.4	1.6 $\pm$ 0.3			37.93 $\pm$ 1.1	34.76 $\pm$ 1.2	37.24 $\pm$ 1.9
ACTION++			89.4 $\pm$ 1.1	78.16 $\pm$ 1.2	0.59 $\pm$ 0.4			39.43 $\pm$ 0.9	36.48 $\pm$ 1.2	34.19 $\pm$ 1.2
<b>DACL (Ours)</b>			<b>90.91<math>\pm</math>0.5</b>	<b>79.12<math>\pm</math>1.0</b>	<b>0.38<math>\pm</math>0.2</b>			<b>41.32<math>\pm</math>1.2</b>	<b>37.54<math>\pm</math>1.0</b>	<b>32.65<math>\pm</math>1.1</b>

Table 1: Comparison with state-of-the-art methods on the ACDC dataset (with 5% and 10% label data) and Synapse dataset (with 10% and 20% label data). Metrics reported the mean $\pm$ standard results with three random seeds.

**Soft Density-Guided Contrastive Learning.** We utilize a pixel-level contrastive loss to align low-density anchors  $m_i^n$  with their the cluster center  $\mathbf{p}_{\text{center}}^{n,+}$  approximated by local and global high-density positive keys. This approach is inspired by the SupCon (Khosla et al. 2020). Our methodology includes selective sampling and optimization across all classes within a batch. Soft Density-Guided Contrastive Learning employs the following contrastive loss function, with  $\tau$  as the temperature parameter:

$$\mathcal{L}_{\text{contra}} = - \sum_{n \in \mathcal{N}} \sum_{m_i \in \mathcal{M}^n} \log \frac{w_i \cdot \exp(m_i \cdot \mathbf{p}_{\text{center}}^{n,+} / \tau)}{\sum_{\mathbf{p}^{n,-} \in \mathcal{P}^{n,-}} \exp(m_i \cdot \mathbf{p}^{n,-} / \tau)}. \quad (10)$$

**Claim 1.** Assume we train models using the proposed optimization method,  $\mathbf{p}_{\text{center}}^{n,+}$  and  $\mathcal{M}^n$  are  $n$ -th class high-density positive keys and low-density anchor set, respectively. The optimal value of similarity measure  $s_i^*$  can be expressed as  $\frac{w_i}{\sum_{k=1}^{N_q} w_k}$ , where  $w_i$  is the corresponding positiveness score for the prototype pair  $(\mathbf{p}_{\text{center}}^{n,+}, \mathbf{m}_i \in \mathcal{M}^n)$  in Eq. 10.

## Experiments and Results

### Dataset and Evaluation Metrics

We evaluate our proposed method on four public datasets with different imaging modalities, the Automatic Cardiac Diagnosis Challenge dataset (ACDC) (Bernard et al. 2018) and Synapse Dataset (Landman et al. 2015). We evaluate our model and other methods using four criteria on the same testing set for fair comparisons, including the Dice similarity coefficient (Mean Dice), Jaccard (JC), and the average

surface distance (ASD). We stack the results of each testing case in all folds together and report the mean and standard deviation (in parentheses) of each metric. The best-column results are highlighted in red and bold.

### Implementation Details

All models are trained with the SGD optimizer, where the initial learning rate is 0.01, momentum is 0.9 and weight decay is  $10^{-4}$ . The network converges after 30,000 iterations of training. An exception is made for the first 1,000 iterations, where  $\lambda_{\text{cross}}$  and  $\lambda_{CL}$  are set to 1 and 0, respectively, which prevents model collapse caused by the initialized prototypes. Empirically, the hyperparameter  $N_q$  (the number of anchors per class in each mini-batch) is set to 256. For each anchor, the number of positive keys  $N_p^+$  and negative keys  $N_p^-$  are both set to 512. The temperature coefficient  $\tau$  in Eq. 10 is set to 0.4. More details are in the appendix.

### Comparisons with State-of-the-Art Methods

We evaluate our proposed model on two datasets using different baseline models (Chen et al. 2021; Wu et al. 2021; Luo et al. 2021; Wu et al. 2022b; Bai et al. 2023; Gao et al. 2023; Wang and Li 2023; Nguyen-Duc et al. 2023; You et al. 2023) to verify the effectiveness of segmentation. Our method is compared with fully supervised learning (SL) and seven state-of-the-art SSL methods from three categories.

**ACDC Dataset.** As given in Table 1, DACL surpasses existing methods under all settings and achieves a new state-of-the-art. It also achieves remarkable improvements compared to the second-best results on three partitions (Dice: +1.62%,

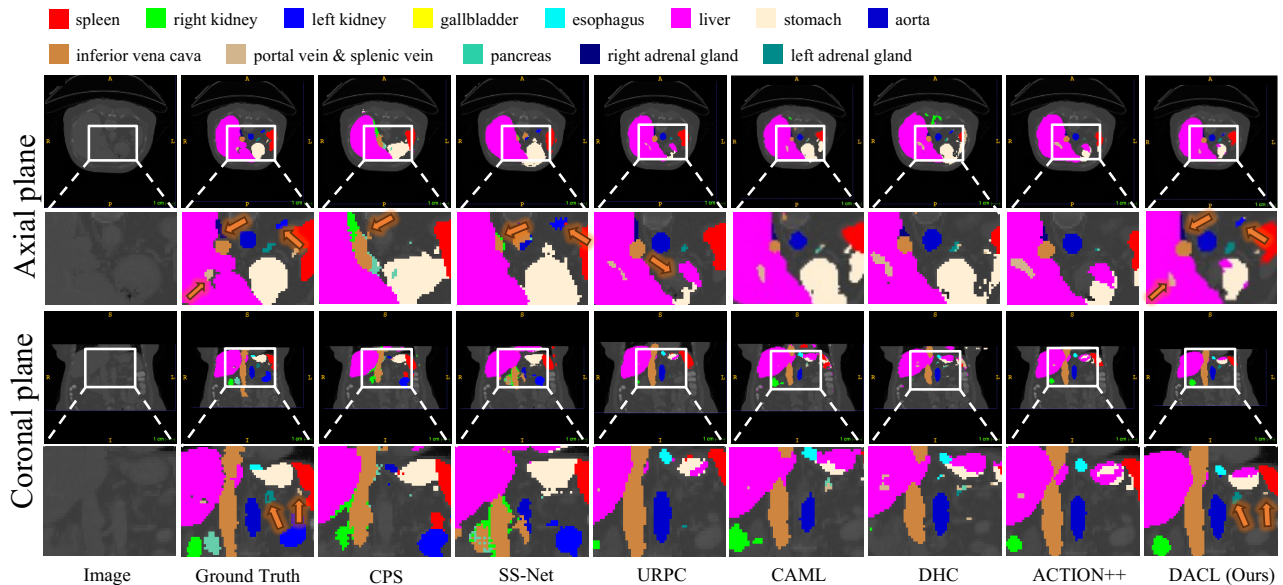


Figure 3: Visualization of the segmentation results from different methods on the Synapse dataset.

	PCL	DA	MS	Bank	$w_i$	ACDC	Synapse
I						$85.71 \pm 0.9$	$18.07 \pm 1.2$
II	✓					$87.50 \pm 0.8$	$24.32 \pm 0.6$
III	✓	✓				$88.60 \pm 0.7$	$25.45 \pm 0.6$
IV	✓	✓	✓			$89.30 \pm 0.6$	$26.30 \pm 0.5$
V	✓	✓	✓	✓		$89.80 \pm 0.5$	$27.12 \pm 0.5$
VI	✓	✓	✓	✓	✓	<b><math>90.91 \pm 0.5</math></b>	<b><math>29.88 \pm 0.7</math></b>

Table 2: Ablation study on main components of the proposed framework on ACDC and Synapse dataset. PCL: Plain contrastive learning. DA: Density-aware anchor and key sampling at a single scale and in min-batch. MS: Multi-scale density estimation. Bank: Density is estimated with class memory, and positive keys are extended. Positiveness  $w_i$ : Correlation between low-density features and cluster center.

1.51%; ASD: +0.17mm, +0.21mm). This further demonstrates the effectiveness of our proposed framework, especially in the challenging 5% labeled data setting where our model significantly outperforms the contrastive learning-based CAML method (Dice: +1.62%).

**Synapse Dataset.** Table 1 lists the comparison results on the more challenging Synapse dataset with 14 categories. The data features complex anatomical contrasts, convoluted boundaries, and heterogeneous textures. Under different settings, the improvements of DACL over the state-of-the-art methods are +2.47% and +1.89% in Dice demonstrating the outstanding robustness of DACL. The performance improvement of the model is attributed to DACL’s effective utilization of geometric structures in the feature space, which enhances intra-class compactness and robustness, particularly in scenarios involving complex anatomical contrasts and heterogeneous textures.

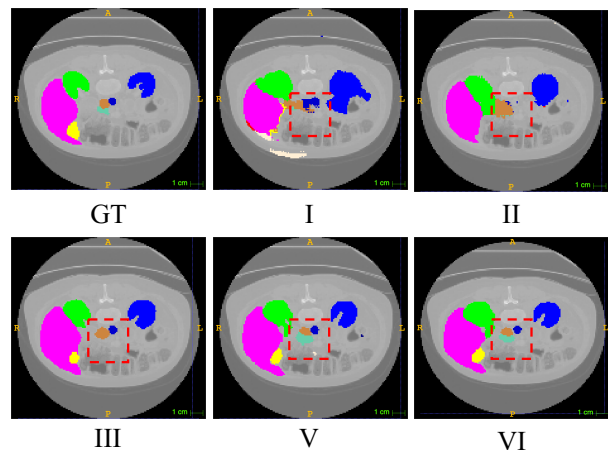


Figure 4: Qualitative results on Synapse dataset using different components, corresponding to experiments I, II, III, V, and VI in the table 2.

**Visual Comparison.** As shown in Fig. 3, the baseline method performs poorly in predicting minority classes and blurry boundaries. The segmentation results of DACL consistently give better boundary accuracy compared to other methods, aligning more closely with the ground truths. Similar to ACTION++, DACL can segment tiny objects with complex boundaries (*e.g.*, gallbladder) and large objects with fine structures (*e.g.*, stomach), as indicated by the orange arrows. This demonstrates DACL’s ability to capture semantics and object structures more accurately.

	Methods	ACDC	Synapse
	Baseline	85.71±0.9	18.07±1.2
Class-wise	+ ED	87.50±0.8	22.88±0.7
	+ IFVD	88.60±0.7	25.00±0.6
	+ PLCL	89.30±0.6	26.50±0.5
	+ Ours	90.50±0.5	27.50±0.5
Pixel-wise	+ InfoNCE	87.00±0.8	23.00±0.7
	+ PLCT	88.20±0.7	25.50±0.6
	+ SCM	89.10±0.6	26.70±0.5
	+ Ours	<b>90.91±0.5</b>	<b>29.88±0.7</b>

Table 3: Comparison of different Class-wise and Pixel-wise losses on the ACDC and Synapse dataset.

## Ablation Study

We conduct an ablation study on the ACDC and Synapse datasets using 10% labeled data, evaluating each component’s contribution with the Dice score.

**Density-Aware Contrastive Learning.** In the ablation study, we analyze the contributions of each component, conducting ablation studies on the Synapse dataset. The quantitative results are given in Table 2, with the CPS (Chen et al. 2021) model as the baseline. We then introduce plain contrastive learning with a random sampling strategy (Experiment II), which only brings limited gains. In Experiment III, we introduce single-scale density information to guide training (mini-batch density estimation and positive keys sampling), significantly improving the performance by 1.13% in Dice. In Experiment IV, we introduce multi-scale density-aware nearest neighbor graphs, improving the performance by 0.85% in Dice. In Experiment V, when estimating density in memory and extending positive samples through memory features, the performance is further improved by 0.82% in Dice, indicating the importance of leveraging the memory bank in our framework. In Experiment VI, considering the hardness of instances in contrastive learning, we introduce adaptive weights  $w_i$  to adjust the low-density features and cluster centers (unlike uniform weight for different features), further improving the performance to achieve the highest Dice score of  $29.88\pm0.7$ .

**Visualization of Results.** As shown in Fig. 4, the baseline combines density-guided contrastive learning (III), multi-scale density estimation (IV), and adaptive weights (VI), further improving segmentation accuracy. The model trained with our DACL strategy can eliminate small false positive regions and accurately detect the details of most organs, especially near adherent edges or in low-contrast regions.

**Comparison of Contrastive Losses.** To verify the capability of density constraints for class reasoning, we replace them with several standard class-wise and pixel-wise losses, including ED (Li et al. 2022), IFVD (Wang et al. 2020), PLCL (Alonso et al. 2021), SCM (Miao et al. 2023), and PLCT (Chaitanya et al. 2023). As given in Table 3, our proposed pixel-wise contrastive learning module significantly outperforms other these methods, demonstrating that DACL enhances semantic discriminative ability in the feature space, whereas other methods lack semantic distinction.

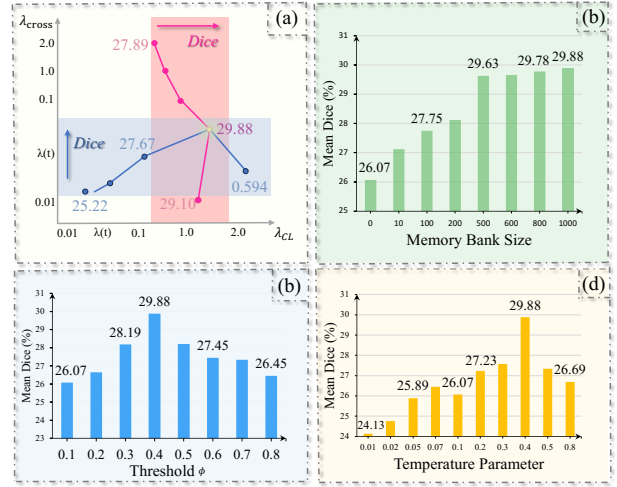


Figure 5: Mean Dice performances on the Synapse dataset with different hyperparameters. Performances with various (a)  $\lambda_{cross}$  and  $\lambda_{CL}$  in Eq. 1, (b) Memory bank size, (c) threshold  $\phi$  for generating the 0-1 seed mask in Eq. 2, and (d) temperature hyperparameters  $\tau$  in Eq. 10.

**Effectiveness of positiveness metrics.** We evaluate the effects of  $\lambda_{cross}$  and  $\lambda_{CL}$ , as well as the Gaussian warming-up function  $\lambda(t)$ . As shown in Fig. 5 (a), we first evaluate the effect of  $\lambda_{CL}$  when  $\lambda_{cross} = 1.0$  (in pink). When the value of  $\lambda_{CL}$  is low (i.e.,  $\lambda_{CL} = 0.5$ ), our density-guided contrastive learning positively impacts optimization. However, we find higher values to be detrimental, possibly because the overall loss is overwhelmed by the high uncertainty of the unlabeled data, hindering the training process. The best performance is achieved when choosing  $\lambda(t)$  as  $\lambda_{CL}$ , as it can adaptively increase the weight according to training iterations. Next, we fixed  $\lambda_{CL} = \lambda(t)$  and used different values of  $\lambda_{cross}$  (in blue). Overall,  $\lambda_{cross} = 1$  achieves the best results. As shown in Fig. 5 (b), compared to not using the Memory bank, Mean Dice increases by at least 1.05%. Also, we observe a positive correlation between Memory bank size and performance, especially at length 1000. As shown in Fig. 5 (c), demonstrates stable performance across different 0-1 seed mask thresholds  $\phi$ , indicating insensitivity to this hyperparameter. Fig. 5 (d) indicates an optimal temperature  $\tau$  of 0.4, consistently outperforming the baseline.

## Conclusion

In this paper, we propose a novel framework named DACL for semi-supervised multi-organ segmentation. Our key idea is to mine effective supervision from the geometry of clusters. To this end, two novel modules: Feature Density-Aware module (FDA) and Soft Density-guided Contrastive Learning (SDCL) are proposed to mine effective neighborhood information by tackling sparse regions inside the feature space. Extensive experimental results on two public multi-organ segmentation datasets demonstrate that we outperform previous state-of-the-art results by a large margin without extra computational load required.

## Acknowledgments

This research was supported by an Australian Government Research Training Program (RTP) Scholarship awarded to Zhongxing Xu.

## References

- Alonso, I.; Sabater, A.; Ferstl, D.; Montesano, L.; and Murillo, A. C. 2021. Semi-supervised semantic segmentation with pixel-level contrastive learning from a class-wise memory bank. In *ICCV*.
- Bai, Y.; Chen, D.; Li, Q.; Shen, W.; and Wang, Y. 2023. Bidirectional Copy-Paste for Semi-Supervised Medical Image Segmentation. In *CVPR*.
- Basak, H.; and Yin, Z. 2023. Pseudo-label guided contrastive learning for semi-supervised medical image segmentation. In *CVPR*.
- Bernard, O.; Lalonde, A.; Zotti, C.; Cervenansky, F.; Yang, X.; Heng, P.-A.; Cetin, I.; Lekadir, K.; Camara, O.; Ballester, M. A. G.; et al. 2018. Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: is the problem solved? *TMI*.
- Chaitanya, K.; Erdil, E.; Karani, N.; and Konukoglu, E. 2023. Local contrastive loss with pseudo-label based self-training for semi-supervised medical image segmentation. *MIA*.
- Chen, D.; Bai, Y.; Shen, W.; Li, Q.; Yu, L.; and Wang, Y. 2023a. MagicNet: Semi-Supervised Multi-Organ Segmentation via Magic-Cube Partition and Recovery. In *CVPR*.
- Chen, T.; Kornblith, S.; Norouzi, M.; and Hinton, G. 2020. A simple framework for contrastive learning of visual representations. In *ICML*.
- Chen, X.; Yuan, Y.; Zeng, G.; and Wang, J. 2021. Semi-supervised semantic segmentation with cross pseudo supervision. In *CVPR*.
- Chen, Z.; Pan, Y.; and Xia, Y. 2023. Reconstruction-driven dynamic refinement based unsupervised domain adaptation for joint optic disc and cup segmentation. *JBHI*.
- Chen, Z.; Pan, Y.; Ye, Y.; Cui, H.; and Xia, Y. 2023b. Treasure in distribution: a domain randomization based multi-source domain generalization for 2d medical image segmentation. In *MICCAI*.
- Chen, Z.; Pan, Y.; Ye, Y.; Lu, M.; and Xia, Y. 2024. Each test image deserves a specific prompt: Continual test-time adaptation for 2d medical image segmentation. In *CVPR*.
- Chi, H.; Pang, J.; Zhang, B.; and Liu, W. 2024. Adaptive Bidirectional Displacement for Semi-Supervised Medical Image Segmentation. In *CVPR*.
- Davies, D. L.; and Bouldin, D. W. 1979. A cluster separation measure. *TPAMI*.
- Deng, S.; Feng, Y.; Lin, H.; Fan, Y.; Lee, A. P.-W.; Hu, X.; and Qin, J. 2024. Semi-supervised TEE Segmentation via Interacting with SAM Equipped with Noise-Resilient Prompting. In *AAAI*.
- Dosovitskiy, A.; Beyler, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. 2020. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.
- Gao, S.; Zhang, Z.; Ma, J.; Li, Z.; and Zhang, S. 2023. Correlation-aware mutual learning for semi-supervised medical image segmentation. In *MICCAI*.
- Hu, M.; Xia, P.; Wang, L.; Yan, S.; Tang, F.; Xu, Z.; Luo, Y.; Song, K.; Leitner, J.; Cheng, X.; et al. 2025. Ophnet: A large-scale video benchmark for ophthalmic surgical workflow understanding. In *ECCV*.
- Hu, M.; Yuan, K.; Shen, Y.; Tang, F.; Xu, X.; Zhou, L.; Li, W.; Chen, Y.; Xu, Z.; Peng, Z.; et al. 2024. OphCLIP: Hierarchical Retrieval-Augmented Learning for Ophthalmic Surgical Video-Language Pretraining. *arXiv preprint arXiv:2411.15421*.
- Huang, H.; Huang, Y.; Xie, S.; Lin, L.; Tong, R.; Chen, Y.-W.; Li, Y.; and Zheng, Y. 2024. Combinatorial CNN-Transformer Learning with Manifold Constraints for Semi-supervised Medical Image Segmentation. In *AAAI*.
- Jiang, S.; Wu, H.; Chen, J.; Zhang, Q.; and Qin, J. 2024. PH-Net: Semi-Supervised Breast Lesion Segmentation via Patch-wise Hardness. In *CVPR*.
- Khosla, P.; Teterwak, P.; Wang, C.; Sarna, A.; Tian, Y.; Isola, P.; Maschinot, A.; Liu, C.; and Krishnan, D. 2020. Supervised contrastive learning. *NeurIPS*.
- Landman, B.; Xu, Z.; Gelsias, J. E.; Styner, M.; Langerak, T. R.; and Klein, A. 2015. 2015 miccai multi-atlas labeling beyond the cranial vault workshop and challenge. In *MICCAI Multi-Atlas Labeling Beyond Cranial Vault—Workshop Challenge*.
- Li, J.; Yu, H.; Chen, C.; Ding, M.; and Zha, S. 2022. Category guided attention network for brain tumor segmentation in MRI. *Physics in Medicine & Biology*, 67(8): 085014.
- Liu, X.; Li, W.; and Yuan, Y. 2024. DiffRect: Latent Diffusion Label Rectification for Semi-supervised Medical Image Segmentation. In *MICCAI*.
- Liu, Y.; Zhu, H.; Liu, M.; Yu, H.; Chen, Z.; and Gao, J. 2024. Rolling-Unet: Revitalizing MLP's Ability to Efficiently Extract Long-Distance Dependencies for Medical Image Segmentation. In *AAAI*.
- Liu, Z.; Ma, Q.; Ma, P.; and Wang, L. 2023. Temporal-frequency co-training for time series semi-supervised learning. In *AAAI*.
- Liu, Z.; Wu, S.; Xu, C.; Wang, X.; Zhu, L.; Wu, S.; and Feng, F. 2022. Copy motion from one to another: Fake motion video generation. *arXiv preprint arXiv:2205.01373*.
- Luo, X.; Liao, W.; Chen, J.; Song, T.; Chen, Y.; Zhang, S.; Chen, N.; Wang, G.; and Zhang, S. 2021. Efficient semi-supervised gross target volume of nasopharyngeal carcinoma segmentation via uncertainty rectified pyramid consistency. In *MICCAI*.
- Ma, Q.; Zhang, J.; Qi, L.; Yu, Q.; Shi, Y.; and Gao, Y. 2024. Constructing and Exploring Intermediate Domains in Mixed Domain Semi-supervised Medical Image Segmentation. In *CVPR*.

- Miao, J.; Zhou, S.-P.; Zhou, G.-Q.; Wang, K.-N.; Yang, M.; Zhou, S.; and Chen, Y. 2023. SC-SSL: Self-correcting Collaborative and Contrastive Co-training Model for Semi-Supervised Medical Image Segmentation. *TMI*.
- Mofid, B.; Mosalla, S. M. M.; Goodarzi, M.; and Tavakoli, H. 2024. Deep CNN-based Fully Automated Segmentation of Pelvic Multi-Organ on CT Images for Prostate Cancer Radiotherapy. *Journal of Biomedical Physics and Engineering*.
- Nguyen-Duc, T.; Le, T.; Bammer, R.; Zhao, H.; Cai, J.; and Phung, D. 2023. Cross-adversarial local distribution regularization for semi-supervised medical image segmentation. In *MICCAI*.
- Qu, C.; Zhang, T.; Qiao, H.; Tang, Y.; Yuille, A. L.; Zhou, Z.; et al. 2024. Abdomenatlas-8k: Annotating 8,000 CT volumes for multi-organ segmentation in three weeks. *NeurIPS*.
- Rosenberg, A.; and Hirschberg, J. 2007. V-measure: A conditional entropy-based external cluster evaluation measure. In *EMNLP-CoNLL*.
- Rousseeuw, P. J. 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*.
- Siam, M.; Oreshkin, B. N.; and Jagersand, M. 2019. Amp: Adaptive masked proxies for few-shot segmentation. In *ICCV*.
- Tang, F.; Trinh, M.; Duong, A.; Ly, A.; Stapleton, F.; Chen, Z.; Ge, Z.; and Razzak, I. 2024a. Discriminating retinal microvascular and neuronal differences related to migraines: Deep Learning based Crosssectional Study. *arXiv preprint arXiv:2408.07293*.
- Tang, F.; Xu, Z.; Huang, Q.; Wang, J.; Hou, X.; Su, J.; and Liu, J. 2023. DuAT: Dual-aggregation transformer network for medical image segmentation. In *PRCV*.
- Tang, F.; Xu, Z.; Qu, Z.; Feng, W.; Jiang, X.; and Ge, Z. 2024b. Hunting Attributes: Context Prototype-Aware Learning for Weakly Supervised Semantic Segmentation. In *CVPR*.
- Van der Maaten, L.; and Hinton, G. 2008. Visualizing data using t-SNE. *Journal of machine learning research*.
- Wang, H.; and Li, X. 2023. DhC: Dual-debiased heterogeneous co-training framework for class-imbalanced semi-supervised medical image segmentation. In *MICCAI*.
- Wang, J.; Huang, Q.; Tang, F.; Meng, J.; Su, J.; and Song, S. 2022. Stepwise feature fusion: Local guides global. In *MICCAI*. Springer.
- Wang, X.; Zhang, B.; Yu, L.; and Xiao, J. 2023. Hunting Sparsity: Density-Guided Contrastive Learning for Semi-Supervised Semantic Segmentation. In *CVPR*.
- Wang, Y.; Zhou, W.; Jiang, T.; Bai, X.; and Xu, Y. 2020. Intra-class feature variation distillation for semantic segmentation. In *ECCV*.
- Wong, H. E.; Rakic, M.; Gutttag, J.; and Dalca, A. V. 2024. Scribbleprompt: Fast and flexible interactive segmentation for any medical image. In *ECCV*.
- Wu, L.; Zhuang, J.; and Chen, H. 2024. Voco: A simple-yet-effective volume contrastive learning framework for 3d medical image analysis. In *CVPR*.
- Wu, S.; Chen, H.; Yin, Y.; Hu, S.; Feng, R.; Jiao, Y.; Yang, Z.; and Liu, Z. 2024a. Joint-Motion Mutual Learning for Pose Estimation in Video. In *ACMMM*.
- Wu, S.; Liu, Z.; Zhang, B.; Zimmermann, R.; Ba, Z.; Zhang, X.; and Ren, K. 2024b. Do as I Do: Pose Guided Human Motion Copy. *IEEE Transactions on Dependable and Secure Computing*.
- Wu, Y.; Ge, Z.; Zhang, D.; Xu, M.; Zhang, L.; Xia, Y.; and Cai, J. 2022a. Mutual consistency learning for semi-supervised medical image segmentation. *MIA*.
- Wu, Y.; Wu, Z.; Wu, Q.; Ge, Z.; and Cai, J. 2022b. Exploring smoothness and class-separation for semi-supervised medical image segmentation. In *MICCAI*.
- Wu, Y.; Xu, M.; Ge, Z.; Cai, J.; and Zhang, L. 2021. Semi-supervised left atrium segmentation with mutual consistency training. In *MICCAI*.
- Xu, C.; Zhang, T.; Zhang, D.; Zhang, D.; and Han, J. 2024a. Deep Generative Adversarial Reinforcement Learning for Semi-Supervised Segmentation of Low-Contrast and Small Objects in Medical Images. *TMI*.
- Xu, Z.; Tang, F.; Chen, Z.; Zhou, Z.; Wu, W.; Yang, Y.; Liang, Y.; Jiang, J.; Cai, X.; and Su, J. 2024b. PolypMamba: Polyp Segmentation with Visual Mamba. In *MICCAI*.
- Yang, Y.; Chen, H.; Liu, Z.; Lyu, Y.; Zhang, B.; Wu, S.; Wang, Z.; and Ren, K. 2023. Action recognition with multi-stream motion modeling and mutual information maximization. *arXiv preprint arXiv:2306.07576*.
- Ye, Y.; Xie, Y.; Zhang, J.; Chen, Z.; Wu, Q.; and Xia, Y. 2024a. Continual self-supervised learning: Towards universal multi-modal medical data representation learning. In *CVPR*.
- Ye, Y.; Zhang, J.; Chen, Z.; and Xia, Y. 2022. Desd: Self-supervised learning with deep self-distillation for 3d medical image segmentation. In *MICCAI*.
- Ye, Y.; Zhang, J.; Chen, Z.; and Xia, Y. 2024b. CADs: A Self-supervised Learner via Cross-modal Alignment and Deep Self-distillation for CT Volume Segmentation. *TMI*.
- You, C.; Dai, W.; Min, Y.; Staib, L.; Sekhon, J.; and Duncan, J. S. 2023. Action++: Improving semi-supervised medical image segmentation with adaptive anatomical contrast. In *MICCAI*.
- Yu, L.; Wang, S.; Li, X.; Fu, C.-W.; and Heng, P.-A. 2019. Uncertainty-aware self-ensembling model for semi-supervised 3D left atrium segmentation. In *MICCAI*.
- Zhao, X.; Tang, F.; Wang, X.; and Xiao, J. 2024a. Sfc: Shared feature calibration in weakly supervised semantic segmentation. In *AAAI*.
- Zhao, Z.; Wang, Z.; Wang, L.; Yuan, Y.; and Zhou, L. 2024b. Alternate Diverse Teaching for Semi-supervised Medical Image Segmentation. In *ECCV*.
- Zheng, Z.; Hayashi, Y.; Oda, M.; Kitasaka, T.; Misawa, K.; and Mori, K. 2024. Federated 3D multi-organ segmentation with partially labeled and unlabeled data. *International Journal of Computer Assisted Radiology and Surgery*.