Beyond Random Sampling: Distribution-Aware Alignment for Semi-Supervised Medical Image Segmentation

Weihao Yan[1], Yeqiang Qian[1*], Yi Dong[2], and Ming Yang[1]

[1] School of Automation and Intelligent Sensing, Shanghai Jiao Tong University, Shanghai, China
qianyeqiang@sjtu.edu.cn

[2] Department of Ultrasound, Xinhua Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai, China

Abstract. Precise medical image segmentation is crucial for clinical diagnosis and treatment planning, yet relies heavily on expensive expert annotations. Semi-supervised medical image segmentation (SSMIS) offers a cost-effective solution but typically operates under the assumption of independent and identically distributed (i.i.d.) data, defaulting to random sampling. While statistically valid at scale, this strategy suffers from severe representation bias in low-data regimes, failing to capture the heterogeneous medical data manifold. To address this, we propose a highly data-efficient framework driven by distribution alignment. First, we introduce an offline Distribution-Aware Sample Selection strategy. By leveraging Vision Foundation Models (VFMs) and our designed Density-K-Center algorithm, we explicitly identify representative structural anchors, establishing a more representative labeled domain. Second, to bridge the remaining distribution gap, we propose the Memory-guided Copy-Paste (MCP) module. Tailored for the inherent class imbalance in medical scans, MCP leverages a semantic memory mechanism to retrieve historically consistent priors for cross-domain alignment, encouraging semantic consistency. Coupled with an easy-to-hard progressive schedule, this framework effectively mitigates early-stage pseudo-label noise. Extensive experiments on six diverse 2D and 3D datasets demonstrate strong segmentation performance, particularly in extremely low-labeled scenarios (*e.g.*, 1/16 ratio).

Keywords: Semi-Supervised Medical Image Segmentation · Distribution Alignment · Sample Selection · Vision Foundation Models

1 Introduction

Medical image segmentation is a fundamental task for modern healthcare, enabling precise quantification of anatomical structures for diagnosis and therapy planning [36, 37]. However, training deep segmentation models requires massive

* Corresponding author.

pixel-level annotations, necessitating expensive and time-consuming expert labor. Semi-supervised medical image segmentation (SSMIS) has emerged as a pivotal solution, aiming to leverage limited labeled data alongside abundant unlabeled scans using various strategies such as consistency regularization, self-training, and co-training [4, 7, 8, 21, 22, 42, 43].

Despite notable advances, most existing SSMIS frameworks assume labeled and unlabeled data are independent and identically distributed (i.i.d.) [14,42,43], relying on random sampling for initialization. As illustrated in Fig. 1, while random sampling provides adequate coverage with ample annotations (high ratio), it exhibits severe instability in low-data regimes. In constrained scenarios, randomly selected samples clump into specific sub-clusters, failing to cover the global data manifold and causing representation bias. Conversely, our method anchors broader regions of the manifold and explicitly aligns features across domains. This enables our framework to outperform standard methods [4, 8, 16, 21, 22, 43], approaching fully supervised performance (Fig. 1, right).

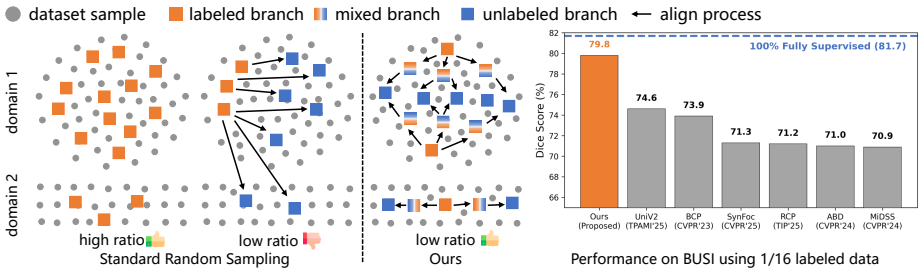


Fig. 1: Motivation and Performance Overview. Left: Standard random sampling covers the manifold at a high labeled ratio but suffers from severe representation bias at a low ratio. Middle: Our method anchors globally representative samples and aligns domains via a mixed branch in progressive schedule. Right: Dice Score comparison on the BUSI dataset (1/16 ratio). Our approach outperforms recent SOTA methods and narrows the gap to the fully supervised setting.

Drawing inspiration from domain alignment principles [5, 10, 12, 35], we argue that effective SSMIS in low-data regimes requires explicitly addressing the distribution shift between the sparse labeled set (source) and the diverse unlabeled set (target). Rather than treating them as a single uniform domain, our objective is twofold: first, to establish a representative source domain that anchors the global manifold, and second, to robustly bridge the cross-domain gap during the learning phase. To this end, we propose a highly data-efficient framework driven by two core innovations. First, we introduce an offline strategy named **Distribution-Aware Sample Selection**. Instead of blind selection, we leverage the semantic priors of Vision Foundation Models (VFMs) such as DINOv2 [24, 31] to map clinical scans into a high-dimensional feature manifold. We then employ the designed Density-K-Center algorithm to select structural anchors. By systematically balancing global diversity with local typicality, this strategy ensures the initial labeled set serves as a less biased and reliable source domain, even with minimal annotations.

Second, we propose the **Memory-guided Copy-Paste (MCP)** module to facilitate robust feature alignment. Conventional mixing strategies like CutMix [4, 8, 16, 22, 46] often fail on medical data due to extreme class imbalance—small anatomical targets (foreground) are easily overwhelmed by large background regions. MCP addresses this by maintaining category-specific Semantic Memory Banks (explicitly separating foreground and background queues). This design allows us to actively retrieve and paste historically consistent foreground patches onto unlabeled images, enhancing the model's sensitivity to small anatomical structures while ensuring semantic consistency. Coupled with an easy-to-hard progressive activation schedule, MCP effectively prevents the model from overfitting to noisy pseudo-labels in early training stages.

The main contributions of this work are summarized as follows:

- We propose the Distribution-Aware Sample Selection strategy to mitigate the representation bias of random sampling in low-data regimes. By leveraging VFM embeddings and Density-K-Center clustering, it identifies representative structural anchors to establish a less biased labeled domain.
- We introduce the Memory-guided Copy-Paste (MCP) module to facilitate robust cross-domain alignment. It employs category-specific memory banks to address class imbalance and integrates an easy-to-hard progressive schedule to suppress early-stage pseudo-label noise.
- Extensive experiments across six diverse 2D and 3D medical datasets (Ultrasound, X-ray, Dermoscopy, and MRI) demonstrate that our method consistently sets new state-of-the-art. Notably, our sampling module functions as a plug-and-play performance multiplier for existing SSMIS methods.

2 Related Work

2.1 Semi-Supervised Medical Image Segmentation

SSMIS reduces reliance on expert annotations by combining limited labeled data with abundant unlabeled scans. Existing methods largely follow three paradigms: (i) *consistency regularization* [8,32,42,43] enforce invariance; (ii) *self-training* [13, 29, 44] refine pseudo-labels; and (iii) *co-training* [7, 15, 21, 28] exploit consensus. However, most frameworks assume i.i.d. distributions, ignoring shifts from heterogeneous clinical acquisition [5, 10]. Under such heterogeneity, standard random sampling [3,21,22,39] causes severe representation bias in low-data regimes. While valid at scale, it fails to anchor the global manifold with scarce labels, weakening initial supervision and degrading performance.

2.2 Domain Adaptation in Medical Imaging

UDA addresses distribution mismatches between labeled source and unlabeled target domains. Representative medical UDA techniques include image-level transfer (*e.g.*, style translation) [40, 45, 50], adversarial feature alignment [23, 33, 34], and self-training-based adaptation [38,41,48,49]. While UDA explicitly models domain gaps, standard SSMIS pipelines rarely treat the labeled/unlabeled

split as a domain shift problem, despite gaps naturally emerging from heterogeneous acquisition [22]. This mismatch between data reality and the i.i.d. assumption motivates revisiting SSMIS from a UDA-inspired perspective.

2.3 Copy-Paste for Medical Image Segmentation

Copy-Paste (CP) is a standard augmentation in SSMIS. Intuitively, it cuts areas from a source image and pastes them onto a target to synthesize diverse training samples. BCP [4] pioneered bi-directional CP, while later works added semantic constraints [8,16] or curriculum learning [27]. However, existing approaches face two medical-specific challenges. First, networks are optimized using small image mini-batches. Due to extreme class imbalance, within-batch mixing often fails to retrieve sparse foregrounds—as a single batch might contain no lesions—leading to severe background dominance. Second, aggressive mixing with noisy pseudo-labels early in training causes overfitting to incorrect supervision.

3 Methodology

As illustrated in Fig. 2, we propose a data-efficient SSMIS framework that operates in two stages to address representation bias and mitigate pseudo-label noise. **Stage 1 (Offline)** establishes a globally representative labeled set via **Distribution-Aware Sample Selection**, utilizing frozen VFM priors to anchor the data manifold. **Stage 2 (Online)** aligns cross-domain features through the **Memory-guided Copy-Paste (MCP)** module. By integrating a semantic memory bank with an easy-to-hard progressive schedule, this stage robustly bridges the inherent distribution gap.

3.1 Preliminaries and Notations

Let the dataset be $\mathcal{D} = \mathcal{D}_L \cup \mathcal{D}_U$, where $\mathcal{D}_L = \{(x_i, y_i)\}_{i=1}^{N_L}$ represents the labeled set and $\mathcal{D}_U = \{x_j\}_{j=1}^{N_U}$ the unlabeled set. SSMIS aims to learn a robust segmentation model f_θ that accurately predicts dense masks across all data. Our UDA-inspired framework explicitly formulates $\mathcal{D}_L$ as the "source domain" and $\mathcal{D}_U$ as the "target domain." We aim to bridge their inherent distribution gap through strategic sample selection prior to training, followed by effective cross-domain alignment during the semi-supervised learning phase.

3.2 Distribution-Aware Sample Selection

Using a pre-trained Vision Foundation Model (VFM), we project the raw clinical data into a high-dimensional feature space $\mathcal{M}$, where anatomical similarities can be easily measured by feature distances.

Hierarchical Feature Encoding For a 2D medical image $x \in \mathbb{R}^{H\times W\times 3}$, we extract multi-level semantic features using a frozen VFM encoder f_{vfm}. To capture both fine-grained local textures and broader global anatomical context, we select a subset of L intermediate layers $\mathcal{L} = \{l_1, l_2, \ldots, l_L\}$.

Let $\mathbf{T}_l \in \mathbb{R}^{N_p\times D_l}$ be the patch tokens extracted from layer $l \in \mathcal{L}$, where N_p is the number of spatial patches and D_l is the feature dimension. The encoding

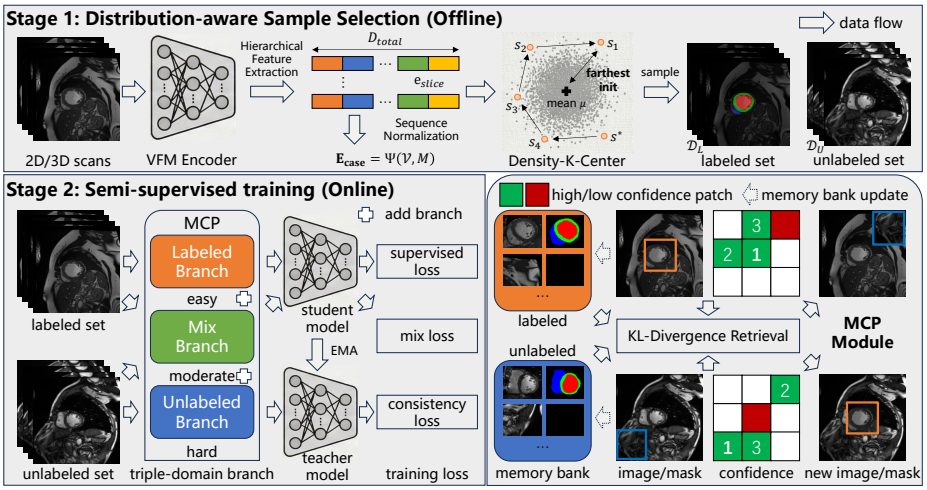


Fig. 2: Overview of our framework. **Top (Stage 1):** Offline Distribution-Aware Sample Selection. Hierarchical VFM features and Density-K-Center clustering identify representative anchors covering the global manifold. **Bottom Left (Stage 2):** Online training pipeline. A teacher-student architecture optimizes labeled, mixed, and unlabeled branches. **Bottom Right:** The MCP module. A category-specific Semantic Memory Bank retrieves historical patches via KL-Divergence for cross-domain alignment.

process aggregates these patch tokens spatially, concatenates them channel-wise, and applies L_2 normalization to obtain the final slice embedding $\mathbf{e}_{slice}$:

$$\mathbf{z}_l = \text{GAP}(\mathbf{T}_l), \quad \mathbf{e}_{raw} = \text{Concat}(\mathbf{z}_{l_1}, \ldots, \mathbf{z}_{l_L}), \quad \mathbf{e}_{slice} = \frac{\mathbf{e}_{raw}}{\|\mathbf{e}_{raw}\|_2} \tag{1}$$

where $\text{GAP}(\cdot)$ denotes Global Average Pooling, and $\mathbf{e}_{raw} \in \mathbb{R}^{D_{total}}$ with $D_{total} = \sum_{i=1}^{L} D_{l_i}$. This normalization projects all embeddings onto a unit hypersphere, making the feature space suitable for cosine-distance-based similarity comparison. For instance, using DINOv2-Small with $L = 4$ layers and $D_l = 384$ yields a unified descriptor $\mathbf{e}_{slice} \in \mathbb{R}^{1536}$.

Volumetric Sequence Normalization In clinical practice, 3D medical modalities (*e.g.*, MRI and CT) provide volumetric data consisting of multiple sequential 2D slices. To treat each patient case as a coherent entity, we propose a Case-Sequence Feature Normalization strategy.

Let a variable-length case be defined as $\mathcal{V} = \{\mathbf{e}_n\}_{n=1}^{N}$. Directly aggregating these slices (*e.g.*, via mean pooling) risks mixing critical pathological details into a diluted single vector. Instead, we map each variable-length sequence into a standardized feature tensor $\mathbf{E}_{case} \in \mathbb{R}^{M \times D_{total}}$, where M is a fixed reference sequence length:

$$\mathbf{E}_{case} = \Psi(\mathcal{V}, M) \tag{2}$$

To implement the normalization function Ψ, we exclusively employ a Structural Centering strategy [17]. Specifically, we extract the M central slices of the vol-

ume. This simple yet effective design aligns with clinical practice: it focuses on the core anatomical regions where lesions or target organs are most frequently located, while ignoring uninformative background slices at the ends. The resulting tensor $\mathbf{E}_{case}$ successfully preserves the patient's structural information, allowing our selection algorithm to evaluate the scan as a whole.

Density-Weighted K-Center Sampling To construct the labeled set $\mathcal{D}_L$, we introduce the Density-K-Center algorithm to balance feature diversity and typicality. We first estimate a local density weight w_i for each sample i based on its k-nearest neighbors (KNN). In our implementation, we set $k = 20$. The local density is computed as:

$$\rho_i = \frac{1}{k} \sum_{j \in \mathrm{KNN}(i)} \|\mathbf{e}_i - \mathbf{e}_j\|_2^2, \quad w_i = \sigma\left(\frac{1}{\rho_i + \epsilon}\right) \tag{3}$$

where $\|\mathbf{e}_i - \mathbf{e}_j\|_2^2$ is the squared Euclidean distance, ϵ is a small constant (*e.g.*, 10^{-6}) to prevent zero division, and $\sigma(\cdot)$ scales the weights into a stable range $[w_{\min}, 1.0]$. We empirically set $w_{\min} = 0.3$ to ensure the weight acts as a soft penalty rather than a hard threshold, preventing extreme bias toward uninformative outlier samples.

Initialization via Farthest-from-Mean To start the selection process, the first sample is chosen as the one with the maximum distance from the global geometric centroid of the dataset:

$$s_1 = \arg\max_i \left(\mathrm{dist}(\mathbf{e}_i, \mu)\right), \quad \mu = \frac{1}{N} \sum_{i=1}^{N} \mathbf{e}_i \tag{4}$$

where μ is the mean feature vector of the entire dataset. This ensures the sampling begins from a distinct boundary of the data distribution.

Greedy Selection Process At each subsequent step, we select the next sample s^* that maximizes the density-weighted distance to the previously selected samples:

$$s^* = \arg\max_{i \notin \mathcal{D}_L} \left(w_i \cdot \min_{s \in \mathcal{D}_L} \mathrm{dist}(\mathbf{e}_i, \mathbf{e}_s) \right) \tag{5}$$

where $\mathbf{e}_i$ and $\mathbf{e}_s$ denote the feature embeddings of a candidate sample from the unselected pool and an already selected anchor, respectively. This iterative process continues until the annotation budget N_L is met. The algorithm encourages samples from diverse regions to be selected, while the density weight w_i guides the model to prefer denser regions where the clinical data is most concentrated.

3.3 Memory-guided Copy-Paste with Progressive Alignment

To bridge the distribution gap between labeled and unlabeled domains, we propose the Memory-guided Copy-Paste (MCP) module. MCP aligns features by combining spatial augmentation with a foreground-aware Memory Bank and a progressive training schedule.

Triple-Domain Branch Design Unlike existing methods focusing on single-domain or bidirectional augmentation [4, 8, 16], our MCP operates across three branches tailored for medical images:

- **Labeled Domain Branch**: Performs augmentation strictly within $\mathcal{D}_L$ to establish a strong supervised baseline.
- **Unlabeled Domain Branch**: Conducts augmentation within $\mathcal{D}_U$ using pseudo-labels that generated from the teacher model, encouraging learning from ambiguous tissue regions.
- **Mixed Domain Branch**: Executes cross-domain Copy-Paste ($\mathcal{D}_L \leftrightarrow \mathcal{D}_U$), transferring reliable anatomical supervision between domains.

An image $x \in \mathbb{R}^{H\times W\times 3}$ is uniformly divided into a 7×7 patch grid. For each active branch, the loss combines Dice and Cross-Entropy (CE) losses:

$$\mathcal{L}_{branch} = \frac{1}{2}\left(\mathcal{L}_{dice}(\hat{y}, y, \omega) + \mathcal{L}_{ce}(\hat{y}, y, \omega)\right) \tag{6}$$

where $\hat{y}$ is the prediction and y the ground truth ($\mathcal{D}_L$) or pseudo-label ($\mathcal{D}_U$). The confidence weight ω is 1.0 for ground truth regions and 0.5 for pseudo-labeled regions to mitigate noise propagation.

Semantic Memory Bank with Foreground-Background Separation The medical images exhibit severe class imbalance, with sparse regions of interest against expansive backgrounds. To overcome this and limited intra-batch diversity, we decouple our Patch Memory Bank into foreground and background subsets: $\mathcal{B} = \{\mathcal{B}^{fg}, \mathcal{B}^{bg}\}$. Each entry $(\mathbf{p}, \mathbf{y}, \mathbf{d}, c, id)$ stores the patch pixels, mask, class probability, mean confidence, and source ID. Target-class patches route to $\mathcal{B}^{fg}$; healthy context patches route to $\mathcal{B}^{bg}$.

Hybrid Eviction Strategy When subsets reach memory bank capacity C_{bank} (1024), we independently apply a hybrid eviction policy: 50% First-In-First-Out (FIFO) for freshness, and 50% removing entries with the lowest confidence c. This class-separated updating prevents majority background patches from flushing out rare foreground anomalies.

Semantic Retrieval During augmentation, MCP selectively retrieves patches by foreground/background identity with probability P_{bank}, while the rest are randomly sampled from the current mini-batch. To synthesize anomalies, a foreground lesion from $\mathcal{B}^{fg}$ can replace a background region. Given a target patch class distribution $\mathbf{d}_t$, we retrieve the most semantically similar source patch $\mathbf{p}_s^*$ by minimizing the Kullback-Leibler (KL) divergence:

$$\mathbf{p}_s^* = \arg\min_{\mathbf{p}\in\mathcal{B}, id\neq id_t} D_{KL}(\mathbf{d}_s \| \mathbf{d}_t), \quad D_{KL}(\mathbf{d}_s \| \mathbf{d}_t) = \sum_i d_{s,i} \log\left(\frac{d_{s,i}}{d_{t,i}}\right) \tag{7}$$

This foreground-aware matching ensures retrieved patches naturally fit the target's tissue context, preventing unnatural semantic mixing.

Progressive Branch Activation To ensure training stability, we use an easy-to-hard progressive activation schedule over a ramp-up period $E_{prog} = R_{prog} \times E_{total}$, where R_{prog} is the epoch ratio and E_{total} is the total number of epochs:

- **Phase 1** ($0 \leq e < \frac{1}{2}E_{prog}$): Only the Labeled Domain Branch is active for a stable baseline.
- **Phase 2** ($\frac{1}{2}E_{prog} \leq e < E_{prog}$): The Mixed Domain Branch is introduced for cross-domain alignment.
- **Phase 3** ($e \geq E_{prog}$): All branches are fully activated.

The overall training objective at epoch e is:

$$\mathcal{L}_{MCP} = \mathcal{L}_{labeled} + \mathbb{I}\left(e \geq \frac{1}{2}E_{prog}\right)\mathcal{L}_{mixed} + \mathbb{I}(e \geq E_{prog})\mathcal{L}_{unlabeled} \quad (8)$$

where $\mathbb{I}(\cdot)$ is an indicator function. This prevents early overfitting to noisy pseudo-labels. The algorithm process is detailed in supplementary material.

4 Experiments

4.1 Datasets and Evaluation Metrics

We evaluate our framework on six public diverse datasets.

2D Datasets. (1) **BUSI [1]**: 780 breast ultrasound images (623/157 split) with benign, malignant, and normal cases. (2) **PMTCXR [47]**: 2,669 chest X-ray images (2,379/290 split) for pneumothorax segmentation. (3) **ISIC [9]**: 2,594 dermoscopy images (1,815/779 split) containing complex skin lesions. (4) **CAMUS [18]**: Cardiac ultrasound sequences from 500 patients (2,800/800 images) annotated for LV and LA structures.

3D Datasets. (1) **PROMISE [20]**: Prostate MRI volumes from 50 patients (split 35/5/10). (2) **ACDC [6]**: Cardiac cine-MRI from 100 patients (70/10/20 split), focusing on RV, LV, and Myocardium segmentation. For both 3D datasets, volumes are converted into 2D slices for training convenience [3], but case-level evaluation is used by concatenating the slice predictions back to 3D volumes.

Evaluation Metrics We employ the Dice Similarity Coefficient (DSC) and Intersection over Union (IoU) to measure regional overlap (higher is better), alongside the 95% Hausdorff Distance (HD95) and Average Surface Distance (ASD) to evaluate boundary localization accuracy (lower is better).

4.2 Implementation Details

Following UniV2 [43], we adopt the DPT decoder [25] with a DINOv2-Small backbone [24]. Training inputs are resized to 518×518. Model is optimized using AdamW (weight decay 0.01) with poly-decayed learning rates (5×10^{-6} for

encoder, 2×10^{-4} for decoder). Models are trained on two RTX 4090 GPUs with balanced labeled/unlabeled mini-batches, using OHEM, cross-entropy, and Dice losses. For offline selection, we apply Density-K-Center ($k = 20$) on L_2-normalized tokens from four intermediate DINOv2 layers. The MCP module employs a 7×7 grid, memory capacity $C_{bank} = 1024$, and retrieval probability $P_{bank} = 0.2$, matching optimal patches via KL-divergence. After a 10% epoch ramp-up ($R_{prog} = 0.1$), all branches share equal loss weights. For fair comparisons, predictions are resized to 256×256 (ACDC) or 224×224 (others) before metric calculation. Full source code will be available at `https://github.com/ywher/DAA4SSMIS`.

4.3 Comparison with State-of-the-Art Methods

Results on 3D Benchmarks (PROMISE, ACDC) Tab. 1 details the results on 3D volumetric data. On **PROMISE**, where random sampling causes severe bias at the 1/16 ratio (only 2 cases), our Distribution-Aware Selection identifies representative volumetric anchors, achieving 87.3% DSC and outperforming UniV2 by 5.3%. On **ACDC**, despite training on 2D slices, our method preserves strong volumetric consistency. At the 1/20 ratio (3 cases), we surpass the Labeled-Only baseline by 8.1% DSC. Even as performance saturates at the 1/5 ratio, our method further improves performance, reaching 91.5% DSC with low boundary error (ASD 0.3).

Table 1: Comparison with SOTA methods on **3D Datasets** (PROMISE, ACDC). Specific labeled ratios (Cases) are defined in the section headers. Labeled Only: the supervised baseline trained solely on the labeled set. Best: **Bold**, Second: Underline.

Methods	Venue	Encoder	Ratio and the number of labeled cases					
			Low Ratio		Medium Ratio		High Ratio	
			IoU/DSC↑	ASD/HD↓	IoU/DSC↑	ASD/HD↓	IoU/DSC↑	ASD/HD↓
Dataset: PROMISE (Ratios: 1/16, 1/10, 1/5 | Cases: 2, 3, 7)								
BCP [4]	CVPR'23	UNet	60.5/73.1	1.2/4.6	61.2/74.9	3.0/11.8	65.6/78.7	7.0/20.6
MiDSS [22]	CVPR'24	UNet	66.2/79.4	1.1/3.4	71.6/83.3	1.0/2.7	70.5/82.5	1.1/2.8
ABD [8]	CVPR'24	SwinUNet	65.7/79.1	1.4/3.6	68.4/80.9	1.3/3.4	68.0/80.7	1.4/3.5
RCP [16]	TIP'25	UNet	68.0/80.9	2.9/9.3	70.2/82.4	1.7/5.1	68.2/80.8	2.9/8.9
AdaMix [27]	MedIA'25	UNet	55.1/70.2	1.7/5.1	66.0/79.4	2.4/4.9	61.8/76.2	2.9/8.7
SynFoc [21]	CVPR'25	MedSAM	75.3/85.7	0.9/2.8	75.9/86.2	1.1/3.6	74.4/85.3	0.9/2.7
UniV2 [43]	TPAMI'25	DINOv2-S	69.8/82.0	0.9/3.2	76.7/86.7	**0.7**/2.4	75.6/86.0	**0.8**/2.6
Labeled Only	-	DINOv2-S	64.0/77.8	1.3/4.3	67.3/79.8	1.1/3.5	67.5/80.2	1.1/3.4
Ours	-	DINOv2-S	**77.7/87.3**	**0.7/2.3**	**78.4/87.9**	**0.7/2.3**	**77.5/87.2**	**0.8/2.3**
Dataset: ACDC (Ratios: 1/20, 1/10, 1/5 | Cases: 3, 7, 14)								
BCP [4]	CVPR'23	UNet	75.6/85.6	1.6/5.3	80.4/88.7	0.7/2.4	83.0/90.4	0.5/1.8
MiDSS [22]	CVPR'24	UNet	80.6/88.9	0.8/3.5	82.0/89.8	0.5/1.5	83.6/90.8	**0.3**/1.2
ABD [8]	CVPR'24	SwinUNet	78.4/87.4	0.6/1.7	82.2/90.0	0.5/**1.3**	83.0/90.5	0.7/2.7
RCP [16]	TIP'25	UNet	79.0/87.9	1.4/5.1	81.5/89.5	1.3/6.2	82.7/90.3	1.0/3.0
AdaMix [27]	MedIA'25	UNet	77.2/86.7	1.2/4.7	78.0/87.2	1.5/5.2	81.1/89.2	0.9/2.0
SynFoc [21]	CVPR'25	MedSAM	73.9/84.3	2.1/6.8	77.9/87.2	0.9/2.5	79.5/88.3	0.8/3.1
UniV2 [43]	TPAMI'25	DINOv2-S	80.2/88.6	0.5/1.6	82.9/90.4	0.5/1.6	83.6/90.9	0.4/1.5
Labeled Only	-	DINOv2-S	70.9/82.3	4.1/15.4	79.3/88.1	1.5/5.3	81.9/89.8	0.6/2.6
Ours	-	DINOv2-S	**82.8/90.4**	**0.3/1.2**	**83.8/91.0**	**0.4/1.3**	**84.6/91.5**	**0.3/1.1**

Results on 2D Benchmarks (BUSI, PMTCXR, ISIC, CAMUS) Tab. 2 summarizes performance across four datasets. On **BUSI**, our framework excels at the scarce 1/16 ratio, achieving 79.8% DSC and reducing HD95 by 12.6 pixels vs. UniV2. On the heterogeneous **PMTCXR**, many SSMIS methods collapse (< 40% DSC) at the 1/16 ratio, whereas our method maintains 52.9% DSC. For **ISIC**, our method outperforms UniV2 by 1.9% DSC at the 1/40 ratio. Finally, on **CAMUS**, we reach 93.2% DSC (1/4 ratio), supporting our method's efficacy on structure-consistent cardiac ultrasounds.

Table 2: Comparison with SOTA methods on **2D Datasets** (BUSI, PMTCXR, ISIC, CAMUS). The specific labeled ratios and counts are indicated in the section headers.

Methods	Venue	Encoder	Ratio and the number of labeled images					
			Low Ratio		Medium Ratio		High Ratio	
			IoU/DSC↑	ASD/HD↓	IoU/DSC↑	ASD/HD↓	IoU/DSC↑	ASD/HD↓
Dataset: BUSI (Ratios: 1/16, 1/8, 1/4 | Counts: 39, 78, 156)								
BCP [4]	CVPR'23	UNet	67.2/73.9	33.1/48.2	70.9/76.7	34.7/46.2	71.2/77.5	29.9/41.7
MiDSS [22]	CVPR'24	UNet	64.3/70.9	42.6/56.5	64.2/70.0	39.5/54.0	68.2/74.8	32.7/45.1
ABD [8]	CVPR'24	SwinUNet	63.6/71.0	42.0/57.6	66.3/73.3	39.5/53.5	69.9/76.4	33.1/45.0
RCP [16]	TIP'25	UNet	64.3/71.2	44.2/57.8	64.9/71.9	42.4/55.4	65.1/71.8	38.6/53.5
AdaMix [27]	MedIA'25	UNet	51.6/58.6	65.2/79.3	61.0/67.9	49.7/60.9	63.2/70.0	49.5/58.9
SynFoc [21]	CVPR'25	MedSAM	64.8/71.3	39.2/54.0	69.0/74.9	35.4/47.0	71.8/78.1	27.3/37.3
UniV2 [43]	TPAMI'25	DINOv2-S	67.7/74.6	36.7/48.8	70.8/77.5	35.0/44.7	73.2/79.2	31.0/39.6
Labeled Only	-	DINOv2-S	61.1/67.9	47.5/61.6	68.4/74.5	33.8/45.5	73.7/79.7	27.6/37.1
Ours	-	DINOv2-S	**73.4/79.8**	**27.0/36.2**	**75.0/81.0**	**25.1/34.2**	**75.8/81.9**	**21.8/31.3**
Dataset: PMTCXR (Ratios: 1/16, 1/8, 1/4 | Counts: 149, 297, 595)								
BCP [4]	CVPR'23	UNet	22.8/32.8	43.1/81.9	28.0/39.1	28.8/65.3	30.9/42.5	24.1/59.1
MiDSS [22]	CVPR'24	UNet	21.9/31.7	38.9/76.6	27.7/38.8	29.6/64.1	29.8/40.6	31.7/64.3
ABD [8]	CVPR'24	SwinUNet	24.8/35.9	29.8/69.0	26.8/38.2	26.3/63.3	30.4/42.2	19.9/52.6
RCP [16]	TIP'25	UNet	23.1/33.2	37.4/73.8	24.3/34.1	43.8/77.3	30.4/41.8	25.7/58.5
AdaMix [27]	MedIA'25	UNet	22.4/31.9	24.1/64.7	27.1/37.8	18.5/57.7	31.4/42.9	17.0/53.2
SynFoc [21]	CVPR'25	MedSAM	27.3/38.8	23.1/56.2	31.1/42.9	20.9/53.6	34.1/46.3	19.9/50.6
UniV2 [43]	TPAMI'25	DINOv2-S	36.0/47.6	26.0/52.0	37.3/49.0	25.6/52.5	38.3/50.3	21.2/45.6
Labeled Only	-	DINOv2-S	30.0/41.5	23.7/56.1	34.7/46.8	21.8/50.4	36.6/48.6	20.6/47.7
Ours	-	DINOv2-S	**40.7/52.9**	**18.9/46.0**	**40.8/53.4**	**12.8/38.3**	**41.6/54.5**	**13.8/39.5**
Dataset: ISIC (Ratios: 1/80, 1/40, 1/20 | Counts: 23, 45, 91)								
BCP [4]	CVPR'23	UNet	70.3/79.5	14.5/29.3	71.0/79.6	14.7/28.9	73.3/82.5	11.2/27.4
MiDSS [22]	CVPR'24	UNet	74.2/82.7	10.9/26.2	72.6/81.3	11.4/27.2	76.0/84.5	8.9/21.7
ABD [8]	CVPR'24	SwinUNet	73.7/83.0	10.7/27.2	74.9/83.8	9.6/23.9	76.6/85.1	8.7/21.3
RCP [16]	TIP'25	UNet	74.8/83.0	11.4/25.0	74.5/83.0	9.1/22.5	76.3/84.7	9.1/21.4
AdaMix [27]	MedIA'25	UNet	72.7/81.8	9.7/22.4	74.8/83.4	8.5/20.0	75.9/84.0	7.9/19.5
SynFoc [21]	CVPR'25	MedSAM	78.5/86.2	7.4/18.8	78.8/86.5	6.5/17.1	80.2/87.6	5.6/14.6
UniV2 [43]	TPAMI'25	DINOv2-S	79.4/87.2	6.8/16.7	79.5/87.1	7.0/16.2	80.9/88.4	5.4/13.1
Labeled Only	-	DINOv2-S	75.3/83.9	10.3/26.8	76.8/85.1	8.3/22.1	79.5/87.2	6.2/16.1
Ours	-	DINOv2-S	**81.5/88.7**	**5.6/14.3**	**81.7/89.0**	**5.1/13.4**	**81.8/89.0**	**5.0/12.7**
Dataset: CAMUS (Ratios: 1/16, 1/8, 1/4 | Counts: 175, 350, 700)								
BCP [4]	CVPR'23	UNet	80.3/88.5	2.4/6.5	81.4/89.3	2.6/6.2	82.3/89.9	2.3/5.8
MiDSS [22]	CVPR'24	UNet	81.7/89.5	2.5/5.9	83.2/90.5	1.9/5.1	84.3/91.2	1.8/4.6
ABD [8]	CVPR'24	SwinUNet	80.0/88.3	2.8/6.4	81.9/89.6	2.1/5.3	83.1/90.4	2.0/4.9
RCP [16]	TIP'25	UNet	81.6/89.4	2.2/5.8	83.1/90.4	1.9/5.2	83.8/90.8	1.9/5.0
AdaMix [27]	MedIA'25	UNet	81.2/89.2	2.3/5.8	82.0/89.8	2.1/5.4	82.9/90.3	2.2/5.3
SynFoc [21]	CVPR'25	MedSAM	81.3/89.3	2.3/5.8	82.4/90.0	2.0/5.2	82.7/90.2	2.0/5.1
UniV2 [43]	TPAMI'25	DINOv2-S	83.3/90.5	1.9/5.0	84.8/91.5	**1.7**/4.5	86.4/92.4	1.5/4.0
Labeled Only	-	DINOv2-S	82.4/89.9	2.1/5.4	84.2/91.1	1.8/4.8	86.2/92.3	1.6/4.1
Ours	-	DINOv2-S	**84.1/91.0**	**1.8/4.7**	**85.5/91.8**	**1.7/4.3**	**87.7/93.2**	**1.4/3.6**

4.4 Ablation Study of Core Components

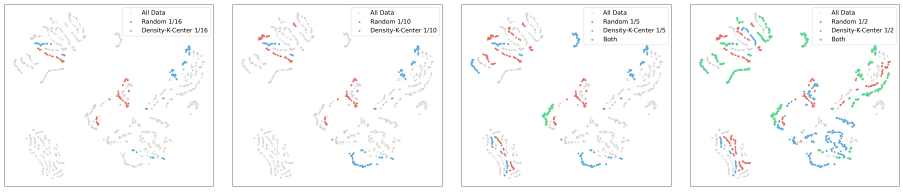
We conduct a step-by-step additive ablation study on the BUSI dataset at the 1/16 and 1/8 labeled ratio (Tab. 3).

Effectiveness of the MCP. Adding basic three-branch Copy-Paste (CP) to the UniV2 baseline moderately improves DSC (74.62% to 75.58%). Integrating the Semantic Memory Bank (MB) for KL-guided patch retrieval further raises DSC to 77.35%. MB also reduces boundary errors (ASD: 32.29 to 29.13 pixels), supporting its role in preserving semantic integrity. Finally, the easy-to-hard Progressive Schedule (Prog.) mitigates early noisy pseudo-labels, improving the fully equipped MCP to 78.06% DSC. To examine why MCP works, we conduct diagnostics on the two targeted designs. MB increases valid foreground copy-paste rate from 86.4% to 100.0% and the retrieved foreground ratio from 60.6% to 82.7%, while Prog. improves early pseudo-label quality at epoch 5 by increasing confidence from 0.88 to 0.91 and reducing entropy from 0.30 to 0.25.

Impact of Distribution-Aware Sample Selection (SS). By selecting anchors that cover broader regions of the data manifold, SS provides a complementary improvement (Tab. 3), reaching 79.80% DSC (1/16). This supports that distribution-aware anchors benefit downstream semi-supervised learning. A further PMTCXR ablation in the supplement supports the complementarity of SS and MCP: MCP improves UniV2 at 1/16 and 1/8 (36.0→38.4 and 37.3→39.4 IoU), and SS+MCP achieves the highest IoU (40.7/40.8).

Table 3: Ablation of core components on BUSI (1/16 and 1/8 ratios). CP: basic three-branch Copy-Paste; MB: Semantic Memory Bank; Prog.: Progressive Activation Schedule; SS: Distribution-Aware Sample Selection.

Baseline	CP	MB	Prog.	SS	1/16 Ratio				1/8 Ratio			
					IoU(%) ↑	DSC(%) ↑	ASD ↓	HD95 ↓	IoU(%) ↑	DSC(%) ↑	ASD ↓	HD95 ↓
✓					67.70	74.62	36.66	48.78	70.84	77.45	34.96	44.69
✓	✓				69.12	75.58	32.29	44.17	71.19	77.35	26.38	38.02
✓	✓	✓			70.96	77.35	29.13	40.75	72.65	78.74	27.29	37.03
✓	✓	✓	✓		71.59	78.06	31.30	40.05	74.05	80.14	25.41	35.06
✓				✓	71.45	78.06	30.55	40.92	74.12	79.82	30.08	38.75
✓	✓	✓	✓	✓	**73.40**	**79.80**	**27.00**	**36.20**	**74.99**	**81.01**	**25.10**	**34.16**



(a) Labeled ratio: 1/16 **(b)** Labeled ratio: 1/10 **(c)** Labeled ratio: 1/5 **(d)** Labeled ratio: 1/2

Fig. 3: t-SNE visualization of the PROMISE feature space. Background (gray) represents unlabeled data. Our Sample Selection (blue) covers broader feature-space clusters, while random sampling (red) shows localized clustering (representation bias).

Fig. 3 illustrates this geometric advantage on PROMISE. Standard random sampling suffers from representation bias, clustering locally. Conversely,

our Density-K-Center strategy scatters anchors across broader regions of the data manifold, improving structural coverage even at a 1/16 ratio to reduce biased supervision. To quantify this effect, we report feature-space coverage and labeled/unlabeled alignment diagnostics in Tab. 4. The 95% radius measures the distance from most unlabeled samples to their nearest anchor; cluster coverage/entropy measure the breadth and evenness of feature clusters; and L/U MMD^2 measures the labeled/unlabeled feature discrepancy after training. Across datasets, DKC improves anchor coverage over random sampling, and the full framework reduces feature discrepancy compared with UniV2.

Table 4: Feature-space coverage and alignment diagnostics. Coverage metrics compare Random avg.→DKC; L/U MMD^2 compares UniV2→Ours.

Dataset	Ratio	95% Rad.↓	Cluster Cov.↑	Cluster Ent.↑	L/U MMD^2 ↓
BUSI	1/8	0.0414→**0.0362**	0.747→**0.860**	0.885→**0.930**	0.0100→**0.0085**
CAMUS	1/8	0.0029→**0.0023**	0.967→**1.000**	0.941→**0.970**	0.0058→**0.0052**
PMTCXR	1/16	0.0161→**0.0150**	0.880→**0.980**	0.930→**0.952**	0.0388→**0.0233**
PROMISE	1/10	0.0681→**0.0423**	0.147→**0.300**	0.445→**0.635**	0.0836→**0.0444**

Tab. 5 quantitatively validates these feature extraction choices. For encoding, utilizing Patch tokens from four intermediate layers consistently outperforms (or performs highly competitively with) the Unlocalized CLS token across all data ratios. This multi-level embedding better preserves the fine-grained spatial semantics crucial for boundary refinement. Furthermore, compared to sampling baselines like K-Means [19], Facility Location [30], and K-Center [26], DKC shows robust scalability. Since all three ratios ($\leq$1/4) represent low-label regimes, the final Patch-Four-DKC configuration is selected by its best average IoU/DSC/ASD/HD performance across them, forming our default pipeline.

Table 5: Ablation study on offline sample selection strategies. We compare different feature encoding methods (CLS, Patch, or both C+P), extracted layers (Last vs. Four intermediate layers), and sampling algorithms. **DKC**: **Density-K-Center**.

Selection Settings			1/16 Ratio		1/8 Ratio		1/4 Ratio	
Encoding	Layers	Sampling	IoU/DSC↑	ASD/HD↓	IoU/DSC↑	ASD/HD↓	IoU/DSC↑	ASD/HD↓
CLS	Four	K-Center	72.0/**79.0**	**26.9**/**37.4**	72.0/78.5	30.6/39.2	75.0/81.0	26.5/34.5
Patch	Four	K-Center	**72.2**/78.7	29.6/39.3	73.3/79.7	**29.0**/38.9	75.4/81.3	24.6/33.5
C+P	Four	K-Center	70.8/77.3	34.9/45.1	71.4/77.7	31.3/41.3	74.9/80.9	26.3/35.5
Patch	Last	K-Center	71.8/78.1	33.3/43.8	73.1/79.2	30.4/39.0	74.3/80.2	26.8/35.4
Patch	Four	K-Means	69.3/75.8	41.4/51.0	70.3/77.0	36.1/46.7	71.6/77.4	36.0/45.2
Patch	Four	Facility	68.6/75.0	38.9/49.6	71.5/77.7	36.0/44.8	71.8/77.8	31.8/40.9
Patch	Four	DKC	71.5/78.1	30.6/40.9	**74.1**/**79.8**	30.1/**38.8**	**76.4**/**82.5**	**17.9**/**28.3**

4.5 Plug-and-Play Capability of Offline Sample Selection

We evaluate the plug-and-play capability of our Distribution-Aware Sample Selection (SS) by integrating it into four diverse SSMIS methods (ABD, RCP,

AdaMix, SynFoc) across varying backbones. As shown in Tab. 6, replacing random sampling with our explicitly selected anchors consistently improves segmentation overlap across all baselines on the BUSI dataset. Improvements are particularly noticeable in the scarce 1/16 regime, yielding DSC gains for AdaMix (+6.5%) and SynFoc (+5.5%), alongside reductions in boundary errors (*e.g.*, an 18.9-pixel HD95 drop for AdaMix). This cross-architecture robustness suggests that capturing the global feature manifold can serve as a versatile performance multiplier for downstream semi-supervised learning.

Stage-1 practicality. DKC ranks samples once using frozen VFM features and requires no query or retraining rounds. This is practical in medical scenarios, where large unlabeled archives are available and can be screened offline before annotation to build a representative labeled set. On BUSI/PMTCXR, feature extraction takes 2.4/8.9 minutes, DKC selection takes 0.5/0.6 seconds, and the Stage-1 cost is 2.9%/2.1% of one SSMIS training run, with no inference overhead.

Active learning comparison. Unlike online active learning, DKC is a one-shot offline selector requiring no warm-up or retraining under identical annotation budgets. On 1/16 BUSI/PMTCXR, it achieves 78.1/48.7 % DSC in just 0.5/0.6s, surpassing both BADGE [2] (78.0/47.8) and BALD [11] (77.4/46.3).

Table 6: Plug-and-play performance of our Sample Selection (SS). The maximum improvement in each column is highlighted in **bold**, and the second-best is underlined. Positive impacts are in green, while negative impacts are in red.

Methods	Venue	Encoder	Ratio and the number of labeled images					
			1/16 (39)		1/8 (78)		1/4 (156)	
			IoU/DSC↑	ASD/HD↓	IoU/DSC↑	ASD/HD↓	IoU/DSC↑	ASD/HD↓
ABD [8]	CVPR'24	SwinUNet	63.6/71.0	42.0/57.6	66.3/73.3	39.5/53.5	69.9/76.4	33.1/45.0
+ our SS			65.8/73.3	39.9/53.5	68.3/75.1	44.0/54.7	71.4/78.1	34.1/44.7
Δ			↑2.2/↑2.3	↓2.1/↓4.1	↑2.0/↑1.8	↑4.5/↑1.2	↑1.5/↑1.7	↑1.0/↓0.3
RCP [16]	TIP'25	UNet	64.3/71.2	44.2/57.8	64.9/71.9	42.4/55.4	65.1/71.8	38.6/53.5
+ our SS			65.5/72.5	42.9/55.2	67.9/75.0	39.9/50.7	68.6/75.2	41.0/51.7
Δ			↑1.2/↑1.3	↓1.3/↓2.6	**↑3.0**/**↑3.1**	↓2.5/↓4.7	**↑3.5**/**↑3.4**	↑2.4/↓1.8
AdaMix [27]	MedIA'25	UNet	51.6/58.6	65.2/79.3	61.0/67.9	49.7/60.9	63.2/70.0	49.5/58.9
+ our SS			58.5/65.1	45.6/60.4	62.6/68.9	50.8/62.6	64.6/71.5	42.7/52.6
Δ			**↑6.9**/**↑6.5**	**↓19.6**/**↓18.9**	↑1.6/↑1.0	↑1.1/↑1.7	↑1.4/↑1.5	↓6.8/↓6.3
SynFoc [21]	CVPR'25	MedSAM	64.8/71.3	39.2/54.0	69.0/74.9	35.4/47.0	71.8/78.1	27.3/37.3
+ our SS			70.5/76.8	40.9/50.3	70.2/76.8	27.5/39.7	72.8/79.4	20.4/30.8
Δ			↑5.7/↑5.5	↑1.7/↓3.7	↑1.2/↑1.9	**↓7.9**/**↓7.3**	↑1.0/↑1.3	**↓6.9**/**↓6.5**

4.6 Comparison with Alternative CP Strategies

Tab. 7 compares Copy-Paste strategies under identical UniV2 [43] baseline and splits. Traditional methods introduce boundary artifacts, causing severe degradation in low-data regimes (*e.g.*, AdaMix's HD95 collapse to 51.07 at 1/16). Conversely, MCP achieves the highest DSC (78.29%) at 1/16 and keeps HD95 below 40 pixels. At 1/4, MCP maintains the best DSC (80.52%). While BCP yields a marginally lower ASD at 1/4, MCP enhances robustness against severe boundary failures (best HD95). This suggests Memory Bank replaces blind spatial mixing with more reliable semantic anchors, preserving anatomical integrity.

Table 7: Comparison of copy-paste strategies. All methods use the same UniV2 baseline and DINOv2-S encoder to ensure a fair evaluation of the mixing strategy itself.

Strategies	Venue	Encoder	1/16 Ratio				1/4 Ratio			
			IoU(%)↑	DSC(%)↑	ASD↓	HD95↓	IoU(%)↑	DSC(%)↑	ASD↓	HD95↓
BCP [4]	CVPR'23		68.62	75.51	27.03	40.36	73.97	80.20	**23.95**	33.85
TP-RAM [22]	CVPR'24		69.43	76.34	28.72	40.98	73.67	79.58	28.68	37.14
ABD [8]	CVPR'24	DINOv2-S	70.51	77.05	30.44	41.92	72.47	78.54	31.55	40.16
RCP [16]	TIP'25		69.41	75.92	30.14	43.31	73.40	79.29	27.10	35.85
AdaMix [27]	MedIA'25		67.92	74.44	39.72	51.07	73.08	78.69	27.44	37.05
MCP (Ours)	-	DINOv2-S	**71.76**	**78.29**	**26.14**	**37.84**	**74.76**	**80.52**	25.07	**33.49**

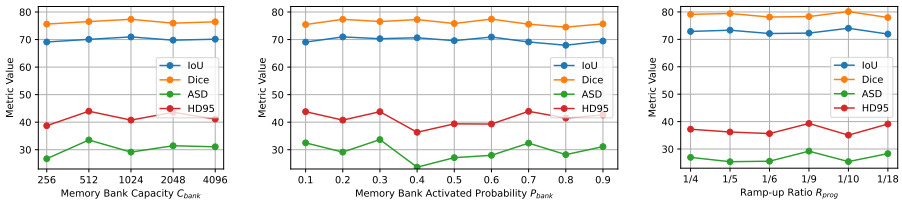
4.7 Hyperparameter Analysis of MCP

Fig. 4 details MCP's hyperparameter sensitivity on BUSI.

Bank Capacity (C_{bank}) at 1/16 ratio. A small capacity (256) restricts diversity, while a large bank ($\geq$ 2048) retains outdated, low-quality patches. Setting $C_{bank} = 1024$ balances diversity and reliability (77.35% DSC).

Retrieval Probability (P_{bank}) at 1/16 ratio. A low probability (0.1) underutilizes memory, whereas high values (≥ 0.8) force over-reliance on historical features, dropping DSC to 74.51%. $P_{bank} = 0.2$ provides the good trade-off.

Progressive Ratio (R_{prog}) at 1/8 ratio. An aggressive schedule (1/18) prematurely introduces complex augmentations, risking overfitting to early noise. Conversely, a prolonged schedule (1/4) delays regularization. The 1/10 ratio safely navigates early stages, peaking at 80.14% DSC and 35.06 HD95.



(a) C_{bank} **(b)** P_{bank} **(c)** R_{prog}

Fig. 4: Performance of MCP module under different hyperparameters on BUSI dataset. (a) Bank Capacity C_{bank}. (b) Activation Probability P_{bank}. (c) Ramp-up Ratio R_{prog}.

4.8 Qualitative Analysis

Fig. 5 shows segmentation results across six medical datasets. Compared to the baseline UniV2, our method demonstrates improved precision. In challenging cases with small structures (PMTCXR, PROMISE, ACDC), UniV2 often under-segments or misses target regions, whereas our approach captures them more reliably. For complex lesions (BUSI, ISIC, CAMUS), we produce smoother, more accurate boundaries aligned with the ground truth.

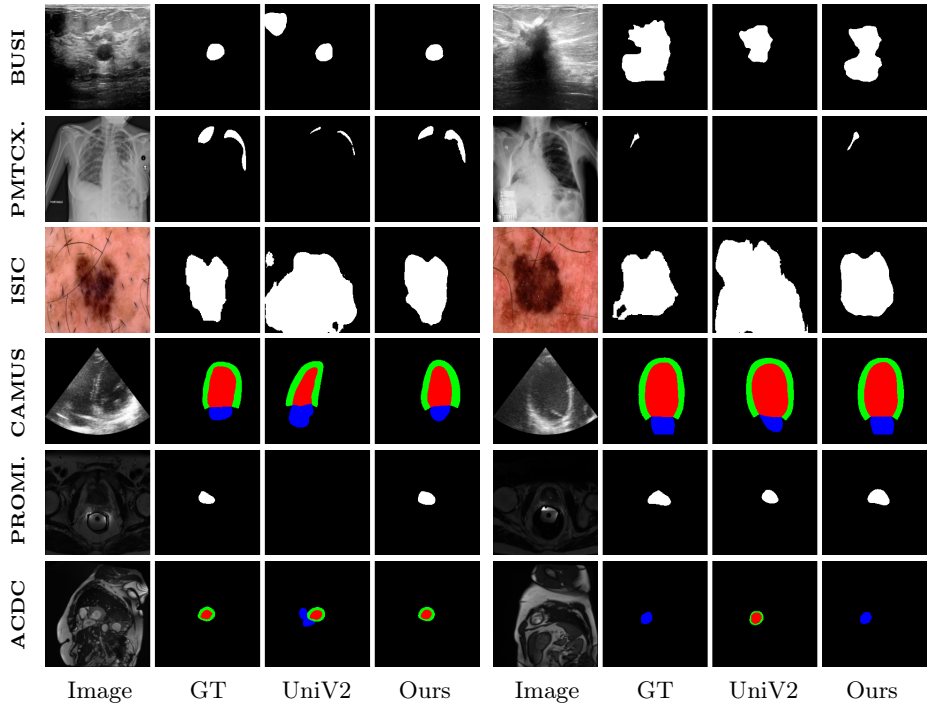


Fig. 5: Qualitative results on BUSI, PMTCXR, ISIC, CAMUS, PROMISE and ACDC (top to bottom). Columns display the Image, Ground Truth (GT), UniV2, and Ours.

5 Conclusion

We propose a data-efficient Semi-Supervised Medical Image Segmentation framework tackling representation bias and pseudo-label noise in extreme low-data regimes. First, Distribution-Aware Sample Selection captures the global anatomical manifold, preventing local minima collapse. Second, the Memory-guided Copy-Paste (MCP) module suppresses noisy pseudo-labels by retrieving historical patches via a KL-guided Semantic Memory Bank. Coupled with progressive training, MCP robustly refines boundaries and preserves structural integrity. Experiments across six datasets show competitive segmentation performance, even at 1/16 labeled ratios. Despite training overhead, inference remains zero-cost. While currently processing 3D data slice-by-slice, future work will extend MCP to direct 3D volume processing and adaptive memory management.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (Grants No. 62473253, U22A20100).

References

1. Al-Dhabyani, W., Gomaa, M., Khaled, H., Fahmy, A.: Dataset of breast ultrasound images. Data in brief **28**, 104863 (2020)
2. Ash, J.T., Zhang, C., Krishnamurthy, A., Langford, J., Agarwal, A.: Deep batch active learning by diverse, uncertain gradient lower bounds. In: International Conference on Learning Representations (2020)
3. Bai, W., Oktay, O., Sinclair, M., Suzuki, H., Rajchl, M., Tarroni, G., Glocker, B., King, A., Matthews, P.M., Rueckert, D.: Semi-supervised learning for network-based cardiac mr image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 253–260. Springer (2017)
4. 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)
5. Ben-David, S., Blitzer, J., Crammer, K., Pereira, F.: Analysis of representations for domain adaptation. Advances in Neural Information Processing Systems **19** (2006)
6. Bernard, O., Lalande, A., Zotti, C., Cervenansky, F., Yang, X., Heng, P.A., Cetin, I., Lekadir, K., Camara, O., Ballester, M.A.G., et al.: Deep learning techniques for automatic mri cardiac multi-structures segmentation and diagnosis: is the problem solved? IEEE Transactions on Medical Imaging **37**(11), 2514–2525 (2018)
7. 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)
8. Chi, H., Pang, J., Zhang, B., Liu, W.: Adaptive bidirectional displacement for semi-supervised medical image segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4070–4080 (2024)
9. Codella, N.C., Gutman, D., Celebi, M.E., Helba, B., Marchetti, M.A., Dusza, S.W., Kalloo, A., Liopyris, K., Mishra, N., Kittler, H., et al.: Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic). In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). pp. 168–172. IEEE (2018)
10. Farahani, A., Voghoei, S., Rasheed, K., Arabnia, H.R.: A brief review of domain adaptation. Advances in Data Science and Information Engineering: Proceedings from ICDATA 2020 and IKE 2020 pp. 877–894 (2021)
11. Gal, Y., Islam, R., Ghahramani, Z.: Deep bayesian active learning with image data. In: International Conference on Machine Learning. pp. 1183–1192. PMLR (2017)
12. Guan, H., Liu, M.: Domain adaptation for medical image analysis: a survey. IEEE Transactions on Biomedical Engineering **69**(3), 1173–1185 (2021)
13. Howlader, P., Das, S., Le, H., Samaras, D.: Beyond pixels: Semi-supervised semantic segmentation with a multi-scale patch-based multi-label classifier. In: European Conference on Computer Vision. pp. 342–360. Springer (2024)
14. Hoyer, L., Tan, D.J., Naeem, M.F., Van Gool, L., Tombari, F.: Semivl: Semi-supervised semantic segmentation with vision-language guidance. In: European Conference on Computer Vision. pp. 257–275. Springer (2024)
15. Huang, K., Zhou, T., Fu, H., Zhang, Y., Zhou, Y., Wu, X.J.: Uncertainty-aware cross-training for semi-supervised medical image segmentation. IEEE Transactions on Image Processing **34**, 5543–5556 (2025)

16. Huang, S., Ge, Y., Liu, D., Hong, M., Zhao, J., Loui, A.C.: Rethinking copy-paste for consistency learning in medical image segmentation. IEEE Transactions on Image Processing **34**, 1060–1074 (2025)
17. Isensee, F., Kickingereder, P., Wick, W., Bendszus, M., Maier-Hein, K.H.: No new-net. In: International MICCAI Brainlesion Workshop. pp. 234–244. Springer (2018)
18. Leclerc, S., Smistad, E., Pedrosa, J., Østvik, A., Cervenansky, F., Espinosa, F., Espeland, T., Berg, E.A.R., Jodoin, P.M., Grenier, T., et al.: Deep learning for segmentation using an open large-scale dataset in 2d echocardiography. IEEE Transactions on Medical Imaging **38**(9), 2198–2210 (2019)
19. Likas, A., Vlassis, N., Verbeek, J.J.: The global k-means clustering algorithm. Pattern Recognition **36**(2), 451–461 (2003)
20. Litjens, G., Toth, R., Van De Ven, W., Hoeks, C., Kerkstra, S., Van Ginneken, B., Vincent, G., Guillard, G., Birbeck, N., Zhang, J., et al.: Evaluation of prostate segmentation algorithms for mri: the promise12 challenge. Medical Image Analysis **18**(2), 359–373 (2014)
21. Ma, Q., Zhang, J., Li, Z., Qi, L., Yu, Q., Shi, Y.: Steady progress beats stagnation: Mutual aid of foundation and conventional models in mixed domain semi-supervised medical image segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5175–5185 (2025)
22. Ma, Q., Zhang, J., Qi, L., Yu, Q., Shi, Y., Gao, Y.: Constructing and exploring intermediate domains in mixed domain semi-supervised medical image segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11642–11651 (2024)
23. Mahmood, F., Chen, R., Durr, N.J.: Unsupervised reverse domain adaptation for synthetic medical images via adversarial training. IEEE Transactions on Medical Imaging **37**(12), 2572–2581 (2018)
24. Oquab, M., Darcet, T., Moutakanni, T., Vo, H.V., Szafraniec, M., Khalidov, V., Fernandez, P., HAZIZA, D., Massa, F., et al.: DINOv2: Learning robust visual features without supervision. Transactions on Machine Learning Research (2024)
25. Ranftl, R., Bochkovskiy, A., Koltun, V.: Vision transformers for dense prediction. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 12179–12188 (2021)
26. Sener, O., Savarese, S.: Active learning for convolutional neural networks: A core-set approach. In: International Conference on Learning Representations (2018)
27. Shen, Z., Cao, P., Su, J., Yang, J., Zaiane, O.R.: Adaptive mix for semi-supervised medical image segmentation. Medical Image Analysis p. 103857 (2025)
28. Shen, Z., Cao, P., Yang, H., Liu, X., Yang, J., Zaiane, O.R.: Co-training with high-confidence pseudo labels for semi-supervised medical image segmentation. In: Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI-23. pp. 4199–4207 (2023)
29. Shi, Y., Zhang, J., Ling, T., Lu, J., Zheng, Y., Yu, Q., Qi, L., Gao, Y.: Inconsistency-aware uncertainty estimation for semi-supervised medical image segmentation. IEEE Transactions on Medical Imaging **41**(3), 608–620 (2021)
30. Shmoys, D.B., Tardos, É., Aardal, K.: Approximation algorithms for facility location problems. In: Proceedings of the Twenty-ninth Annual ACM Symposium on Theory of Computing. pp. 265–274 (1997)
31. Siméoni, O., Vo, H.V., Seitzer, M., Baldassarre, F., Oquab, M., Jose, C., Khalidov, V., Szafraniec, M., Yi, S., Ramamonjisoa, M., et al.: Dinov3. arXiv preprint arXiv:2508.10104 (2025)

32. Sohn, K., Berthelot, D., Carlini, N., Zhang, Z., Zhang, H., Raffel, C.A., Cubuk, E.D., Kurakin, A., Li, C.L.: Fixmatch: Simplifying semi-supervised learning with consistency and confidence. Advances in Neural Information Processing Systems **33**, 596–608 (2020)
33. Tsai, Y.H., Hung, W.C., Schulter, S., Sohn, K., Yang, M.H., Chandraker, M.: Learning to adapt structured output space for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 7472–7481 (2018)
34. Vu, T.H., Jain, H., Bucher, M., Cord, M., Pérez, P.: Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2517–2526 (2019)
35. Wang, Q., Li, W., Gool, L.V.: Semi-supervised learning by augmented distribution alignment. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 1466–1475 (2019)
36. Wang, Y., Zhou, Y., Shen, W., Park, S., Fishman, E.K., Yuille, A.L.: Abdominal multi-organ segmentation with organ-attention networks and statistical fusion. Medical Image Analysis **55**, 88–102 (2019)
37. Wu, H., Wang, Z., Song, Y., Yang, L., Qin, J.: Cross-patch dense contrastive learning for semi-supervised segmentation of cellular nuclei in histopathologic images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11666–11675 (2022)
38. Xie, Q., Li, Y., He, N., Ning, M., Ma, K., Wang, G., Lian, Y., Zheng, Y.: Unsupervised domain adaptation for medical image segmentation by disentanglement learning and self-training. IEEE Transactions on Medical Imaging **43**(1), 4–14 (2022)
39. Yan, W., Qian, Y., Li, Y., Li, T., Wang, C., Yang, M.: Ss-ada: A semi-supervised active domain adaptation framework for semantic segmentation. arXiv preprint arXiv:2407.12788 (2024)
40. Yan, W., Qian, Y., Wang, C., Yang, M.: Threshold-adaptive unsupervised focal loss for domain adaptation of semantic segmentation. IEEE Transactions on Intelligent Transportation Systems **24**(1), 752–763 (2023)
41. Yan, W., Qian, Y., Zhuang, H., Wang, C., Yang, M.: Sam4udass: When sam meets unsupervised domain adaptive semantic segmentation in intelligent vehicles. IEEE Transactions on Intelligent Vehicles **9**(2), 3396–3408 (2024)
42. Yang, L., Qi, L., Feng, L., Zhang, W., Shi, Y.: Revisiting weak-to-strong consistency in semi-supervised semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7236–7246 (2023)
43. Yang, L., Zhao, Z., Zhao, H.: Unimatch v2: Pushing the limit of semi-supervised semantic segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence **47**(4), 3031–3048 (2025)
44. Yang, L., Zhuo, W., Qi, L., Shi, Y., Gao, Y.: St++: Make self-training work better for semi-supervised semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4268–4277 (2022)
45. Yang, Y., Soatto, S.: Fda: Fourier domain adaptation for semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4085–4095 (2020)
46. Yun, S., Han, D., Oh, S.J., Chun, S., Choe, J., Yoo, Y.: Cutmix: Regularization strategy to train strong classifiers with localizable features. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 6023–6032 (2019)

47. Zawacki, A., Wu, C., Shih, G., Elliott, J., Fomitchev, M., Hussain, M., Lakhani, P., Culliton, P., Bao, S.: Siim-acr pneumothorax segmentation 2019 (2019)
48. Zhang, Y., Wei, Y., Wu, Q., Zhao, P., Niu, S., Huang, J., Tan, M.: Collaborative unsupervised domain adaptation for medical image diagnosis. IEEE Transactions on Image Processing **29**, 7834–7844 (2020)
49. Zheng, Z., Yang, Y.: Rectifying pseudo label learning via uncertainty estimation for domain adaptive semantic segmentation. International Journal of Computer Vision **129**(4), 1106–1120 (2021)
50. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 2223–2232 (2017)