STRUCTURED SUBSET-RELATIVE DECISION FIELDS FOR MEDICAL IMAGE SEGMENTATION FROM SCARCE SCRIBBLES

Anonymous authors
Paper under double-blind review

ABSTRACT

Scribble annotations reduce the labeling cost of medical image segmentation, yet when only a few samples are scribble-annotated, models still struggle to learn reliable anatomical semantics from such scarce supervision. Task-specific priors can provide additional inductive guidance beyond direct label supervision, while they often encode higher-order structures, such as region-level organization, semantic relations, and spatial and geometric regularities, which are not readily accommodated by conventional pixel-level supervision. To address this issue, we propose SRDF, an anatomy- and geometry-aware structured representation framework. It formulates multiclass predictions as subset-relative decision fields with locally calibrated distances, yielding differentiable structured objects that jointly encode region semantics and local boundary geometry. Building on SRDF, we further design anatomical-structure and boundary-geometry constraints, allowing anatomical priors and local image evidence to constrain region organization and boundary localization, respectively, at semantic scopes aligned with the corresponding priors, thereby providing structural and geometric guidance for learning under scarce scribble supervision. Experiments on ACDC and Prostate show that the proposed method outperforms existing methods under few-shot scribble supervision.

1 INTRODUCTION

Medical image segmentation (Li et al., 2024a; Lin et al., 2025) plays a central role in disease diagnosis, treatment assessment, and preoperative planning, yet relies on labor-intensive pixel-wise annotations. Scribble supervision (Ji et al., 2019; Li et al., 2023) reduces this burden by assigning class labels only at a few representative locations. In clinical practice, collecting cases and obtaining expert annotations remain difficult, leaving even scribble-annotated data scarce. We therefore consider a weakly semi-supervised setting in which only a few cases carry scribble annotations and all remaining cases are unlabeled.

Existing scribble-supervised methods commonly employ label propagation (Lin et al., 2016), pseudo-label generation (Luo et al., 2022), or consistency learning (Tarvainen & Valpola, 2017) to involve additional unlabeled pixels in training. Although these strategies broaden the coverage of supervision, the added signals still arise primarily from model predictions or propagation assumptions. With only a few annotated cases, a model has limited evidence from which to infer reliable and discriminative anatomical semantics, allowing prediction errors to accumulate and reinforce themselves during pseudo-supervised training. Pretrained vision foundation models (Caron et al., 2021) provide external visual knowledge that may help alleviate this issue, but the gap between generic visual knowledge and medical anatomy limits its direct use for learning discriminative, task-specific representations (Matsoukas et al., 2022; Zoetmulder et al., 2022). Especially in the few-shot scribble-supervised setting, the extremely limited supervision further provides insufficient guidance to adapt such knowledge to task-specific anatomical semantics.

Under such scarce supervision, task-specific priors can provide valuable complementary knowledge for learning reliable segmentation semantics. These priors often encode higher-order semantics and structures beyond isolated pixel labels, such as region-level organization, relations among semantic entities, and spatial or geometric regularities. However, existing scribble-supervised methods

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

mainly expand the coverage of pixel-level supervision, but do not elevate its level of abstraction. Consequently, although these higher-order structures are ultimately determined by pixel-wise predictions, they are not explicitly organized into learnable objects whose semantic scope matches the corresponding priors. This motivates a structured representation that bridges pixel-level predictions and higher-order task knowledge while remaining consistent with the final multiclass segmentation.

To bridge this abstraction gap, we propose SRDF, an anatomy- and geometry-aware structured representation framework that reorganizes multiclass pixel predictions into explicit, differentiable constraint objects aligned with higher-order task knowledge. Specifically, SRDF constructs subset-relative decision fields to represent both individual tissues and composite anatomical regions spanning multiple classes, while preserving consistency with the underlying multiclass decisions. Local geometric calibration further equips these fields with boundary-aware spatial information, allowing region structure and boundary geometry to be described within a common representation. SRDF thereby provides a unified and differentiable structured interface between pixel-wise predictions and higher-order task knowledge. Building on this representation, we further introduce complementary anatomical-structure and boundary-geometry constraints, which respectively use anatomical priors and scribble anchors to regulate the spatial organization of target regions, and transform local image appearance into boundary localization cues to compensate for the missing boundary supervision. In this way, higher-order task knowledge can directly guide learning through structured objects of matching semantic scope, providing structural and geometric evidence for learning reliable segmentation semantics under scarce scribble supervision.

Our contributions are summarized as follows:

1. **Structured Representation Framework:** We propose SRDF, an anatomy- and geometry-aware structured representation framework that combines subset-relative decision fields with local distance calibration to form differentiable constraint objects encoding region semantics and local boundary geometry.
2. **Task-Specific Prior Constraints:** Based on SRDF, we develop anatomical-structure and boundary-geometry constraints tailored to few-shot scribble supervision. These constraints leverage class adjacency, scribble-anchored connectivity reconstruction, and local image evidence to provide structural and geometric guidance under scarce supervision.
3. **Experimental Validation:** Extensive experiments on ACDC and Prostate demonstrate the effectiveness of the proposed method under few-shot scribble supervision, with controlled comparisons and sensitivity analyses showing consistent performance across settings.

2 Related Work

Scribble annotations provide class information at only a small number of pixels, substantially reducing annotation costs for medical image segmentation but leaving object extent and boundaries unspecified. Early methods supervise labeled pixels with partial cross-entropy (pCE) (Can et al., 2018) or extend supervision through label propagation (Lin et al., 2016). Subsequent work exploits unlabeled data through pseudo-labeling and consistency learning (Luo et al., 2022; Wang et al., 2024) or mixing-based augmentation (Zhang & Zhuang, 2022). Vision–class association (Li et al., 2023), superpixel-based region guidance (Zhou et al., 2023; Su et al., 2026), boundary enhancement (Nguyen-Tat et al., 2025), foreground feature discrimination (Liu et al., 2025), and hybrid architectures (Li et al., 2024b) further improve segmentation learning. Yet when only a few cases carry scribble annotations, limited supervision still makes reliable anatomical semantic learning difficult, and prediction-derived pseudo-labels may reinforce early errors.

In the broader field of medical image segmentation, anatomical knowledge has been incorporated through multi-region spatial relations (Delong & Boykov, 2009), inter-class topological interactions (Gupta et al., 2022), persistent-homology constraints (Byrne et al., 2023), and signed-distance supervision (Xue et al., 2020). These studies show that regional relations and boundary geometry can provide structural guidance beyond pixel labels. Under few-shot scribble supervision, however, annotations delineate neither complete regions nor contours, while relevant priors may concern anatomical objects spanning multiple classes. This motivates a structured representation that constructs decision-consistent fields for individual tissues and composite regions from multiclass

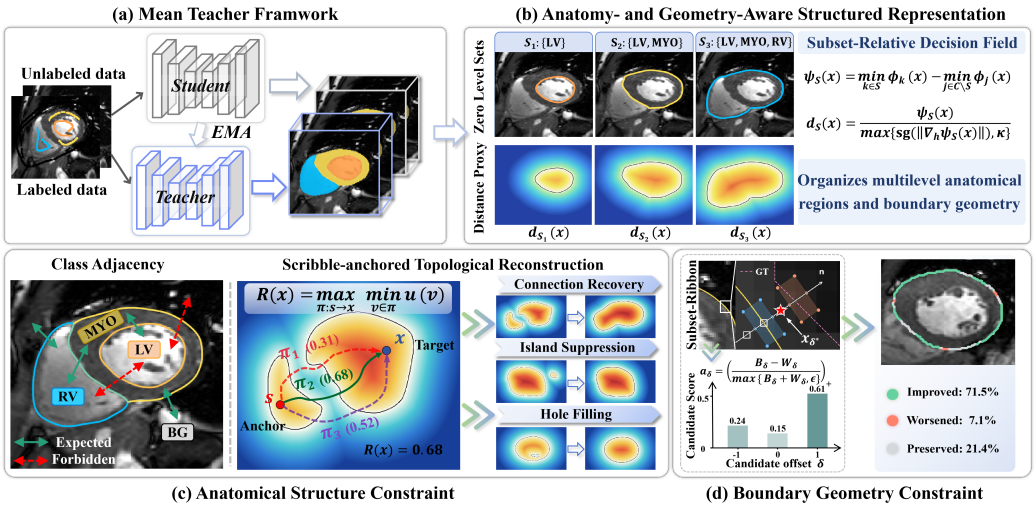


Figure 1: Overview of the proposed method. (a) Mean Teacher training with scribble-annotated and unlabeled data. (b) SRDF, an anatomy- and geometry-aware structured representation framework, jointly encodes individual tissues, composite anatomical regions, and their local boundary geometry. (c) Anatomical structure constraints based on class adjacency and scribble-anchored topological reconstruction. (d) Subset-Ribbon boundary geometry constraint based on local appearance evidence.

predictions and captures their local boundary geometry, providing differentiable objects for priors at different semantic scopes.

3 METHOD

This section introduces the proposed anatomy- and geometry-aware structured representation framework, SRDF, together with the anatomical structure and boundary geometry constraints built upon it. The overall framework is illustrated in Figure 1.

3.1 PROBLEM SETUP AND BASE OBJECTIVE

Let $\mathcal{D}_L$ contain a few scribble-annotated cases and $\mathcal{D}_U$ the remaining unlabeled cases. For a training slice I with pixel domain Ω, the scribbles label locations $A \subset \Omega$ with $Y(x) \in \mathcal{C} = \{0, \dots, C-1\}$. We use partial cross-entropy (pCE) (Luo et al., 2022) on the labeled pixels and Mean Teacher (MT) (Tarvainen & Valpola, 2017) on unlabeled pixels, where an EMA-updated teacher provides pseudo-labels for student supervision. The resulting base losses are denoted by $\mathcal{L}_{\text{pCE}}$ and $\mathcal{L}_{\text{MT}}$; their full definitions are given in Appendix B.

3.2 SRDF: ANATOMY- AND GEOMETRY-AWARE STRUCTURED REPRESENTATION

Task-specific prior knowledge may concern either individual tissues or anatomical regions composed of multiple classes. To impose constraints at the corresponding semantic levels, SRDF constructs subset-relative decision fields consistent with the final multi-class decisions and calibrates local boundary scales, yielding differentiable region and boundary representations aligned with different semantic scopes.

Subset-relative decision field. Let $S \subsetneq \mathcal{C}$ be a nonempty subset of the classes forming a target region. Simply thresholding the sum of probabilities in S need not reproduce the union of its classes under the final multiclass decision. We instead compare the strongest responses inside and outside S. Given model scores $\boldsymbol{z}(x) \in \mathbb{R}^C$, define class potentials $\boldsymbol{\phi}(x) = -\boldsymbol{z}(x)$, so that

$p(x) = \mathrm{softmax}(-\boldsymbol{\phi}(x))$. The subset-relative decision field is defined as

$$\psi_S(x) = \min_{k\in S} \phi_k(x) - \min_{j\in\mathcal{C}\setminus S} \phi_j(x) = \log \frac{\max_{j\notin S} p_j(x)}{\max_{k\in S} p_k(x)}. \tag{1}$$

With a unique winning class, $\psi_S(x) < 0$ if and only if the final multiclass prediction belongs to S, and its zero level set $\{x \in \Omega \mid \psi_S(x) = 0\}$ corresponds to the decision interface between S and its complement. Thus, the negative region of ψ_S coincides with the union of classes in S under the final multiclass segmentation, while interfaces between classes within S are not treated as its outer boundary. The continuous field is piecewise differentiable with respect to the class scores, allowing constraint gradients to propagate back to the class outputs. Moreover, nested subsets yield nested regions, as proved in Appendix A. In this way, different class subsets allow SRDF to flexibly represent regions and boundaries at different semantic levels, ranging from individual tissues to combinations of multiple tissues, as illustrated in Figure 1(b).

Local distance calibration. The magnitude of ψ_S reflects the strength of class competition rather than spatial distance. We apply a boundary-weighted Eikonal loss $\mathcal{L}_{\mathrm{eik}}$ to singleton decision fields to stabilize the gradient scale near their zero level sets, and locally normalize each subset field:

$$d_S(x) = \frac{\psi_S(x)}{\max\left\{\mathrm{sg}\left[\sqrt{\|\nabla_h \psi_S(x)\|_2^2 + \epsilon}\right],\ \kappa\right\}}, \tag{2}$$

where ∇_h is the discrete spatial gradient, sg denotes stop-gradient, $\epsilon > 0$ stabilizes the gradient norm, and $\kappa > 0$ bounds the denominator away from zero. Near a locally smooth interface, d_S approximates signed distance along the interface normal. Because the denominator is strictly positive, it preserves the sign and zero level set of ψ_S. Thus, d_S provides a local geometric scale for each subset field, although it is not an exact global distance transform. The Eikonal objective and related implementation details are given in Appendix B.

3.3 Anatomical Structure Constraints

Inter-tissue adjacency and separation, together with the connectivity and hole structure of target regions, constitute two complementary aspects of anatomical structure. Building on SRDF, we design a class adjacency constraint and a scribble-anchored topology reconstruction constraint, as illustrated in Figure 1(c), to guide the model in learning anatomical structural relations under scarce supervision.

Class adjacency. To express inter-tissue adjacency as a learnable constraint, we measure class contact through the local overlap of soft bands around the SRDF zero level sets. Specifically, for class k in image b, the band response ζ_{bk} measures proximity to its boundary, and the strongest overlap between two bands defines their contact strength:

$$\zeta_{bk}(x) = \frac{1}{1 + 4d_{b,\{k\}}^2(x)}, \qquad c_{b,ij} = \max_{x\in\Omega} \sqrt{\zeta_{bi}(x)\zeta_{bj}(x)}. \tag{3}$$

The response falls to half its boundary value at a distance of half a pixel, matching the pixel-grid boundary scale. Let $\mathcal{E}$ contain the class pairs with specified contact relations, with $a_{ij} = 1$ for required contact and $a_{ij} = 0$ for forbidden contact. The gate $g_{b,ij}$ equals one when both classes appear in the scribbles of image b, and zero otherwise. The adjacency loss is

$$\mathcal{L}_{\mathrm{adj}} = \frac{\sum_{b,(i,j)\in\mathcal{E}} g_{b,ij}(c_{b,ij} - a_{ij})^2}{\sum_{b,(i,j)\in\mathcal{E}} g_{b,ij}}, \tag{4}$$

and is set to zero when no pair is active. It encourages required contact and suppresses forbidden contact without prescribing interface locations.

Scribble-anchored reconstruction. Adjacency alone cannot resolve breaks, islands, or anomalous holes within a region. A global count of components or holes also leaves open which predicted structures should be connected or suppressed; scribble locations provide spatial anchors with known class membership. We apply reconstruction to anatomically meaningful subsets so that, for example,

a normal cavity enclosed by myocardium need not be treated as a hole of the composite region. For a target subset S, foreground propagation starts from one scribble anchor; reachability at the other anchors supervises their connection. Strongly supported foreground locations that remain unreachable and are not anchors are candidate islands. Background propagation from the image border identifies candidate holes unreachable from the exterior. The two capacities are derived from the SRDF distance proxy:

$$u_S^F = \mathrm{sigmoid}(-2d_S - 1), \qquad u_S^B = \mathrm{sigmoid}(2d_S - 1). \tag{5}$$

Capacity rises with distance inside the corresponding region; its midpoint lies half a pixel from the predicted boundary. Hence, path reachability reflects both connectivity and the distance margin around a connection. The capacity of a path is limited by its weakest location, and the strongest path determines reachability. With capacity u and seed map s, maximum-bottleneck propagation (Vincent, 1993) iterates

$$R^{(0)} = s, \qquad R^{(t+1)}(x) = \max\big(R^{(t)}(x), \min\{u(x), \max_{y \in \mathcal{N}_8^+(x)} R^{(t)}(y)\}\big), \tag{6}$$

where $\mathcal{N}_8^+(x)$ includes x and its eight neighbors. After K iterations, we obtain foreground and background reachability maps R_S^F and R_S^B. Let $\mathcal{M}$, $\mathcal{I}$, and $\mathcal{H}$ denote the active connection anchors, candidate islands, and candidate holes across the batch. Their reconstruction objective can be written compactly as

$$\mathcal{L}_{\mathrm{top}} = \mathbb{E}_{\mathcal{M}}\big[(1 - 2R^F)_+^2\big] + \mathbb{E}_{\mathcal{I}}\big[(2u^F - 1)_+^2\big] + \mathbb{E}_{\mathcal{H}}\big[(2u^B - 1)_+^2\big], \tag{7}$$

where $(v)_+ = \max(v, 0)$ and an empty set contributes zero. The terms raise bottlenecks between anchors, weaken unreachable foreground islands, and fill enclosed background holes, respectively. As a connection is restored, its newly reachable locations leave the island candidates. Set construction and propagation settings are detailed in Appendix B. The anatomical objective is

$$\mathcal{L}_{\mathrm{anat}}(t) = \mathcal{L}_{\mathrm{adj}} + \mathbf{1}[t \bmod q = 0]\mathcal{L}_{\mathrm{top}}, \tag{8}$$

where t is the training step and $\mathbf{1}[\cdot]$ is the indicator function. To reduce iterative propagation cost, topological reconstruction is evaluated every q steps.

3.4 Boundary Geometry Constraint

Anatomical structure constraints provide structural information, but sparse scribbles offer limited evidence for boundary localization. Motivated by the local appearance differences often found across tissue interfaces, we introduce Subset-Ribbon, a boundary geometry constraint. It uses the local distance and orientation encoded by SRDF, taking the teacher's zero level set as a reference and combining cross-interface contrast along the normal with within-side consistency to define a boundary target for the student (Figure 1(d)).

Specifically, for teacher and student distance proxies $\bar{d}_S$ and d_S, the outward teacher normal is $n = \nabla_h \bar{d}_S / \sqrt{\|\nabla_h \bar{d}_S\|_2^2 + \epsilon}$. At a location near the teacher boundary, we form candidate interfaces

$$x_0 = x - \mathrm{clip}(\bar{d}_S(x), -1, 1)n(x), \qquad x_\delta = x_0 + \delta n(x), \quad \delta \in \{-1, 0, 1\}. \tag{9}$$

At each candidate, short ribbons sampled on the two sides of the normal summarize local image intensity and gradient. Let B_δ measure descriptor contrast between the sides and W_δ their average within-side variation. The image-evidence score selects the target interface:

$$a_\delta = \big(\tfrac{B_\delta - W_\delta}{\max\{B_\delta + W_\delta, \epsilon\}}\big)_+, \qquad \delta^* = \arg\max_\delta a_\delta, \qquad x^* = x_{\delta^*}. \tag{10}$$

The gap r between the two highest scores downweights ambiguous choices, while ω selects valid teacher-boundary locations; their construction is given in Appendix B. The student then fits a signed-distance profile centered on x^*:

$$\ell_S(x) = \frac{1}{3} \sum_{\xi \in \{-0.5, 0, 0.5\}} \mathrm{smooth}_{L_1}\big(d_S(x^* + \xi n) - \xi\big), \qquad \mathcal{L}_{\mathrm{rib},S} = \frac{\sum_{b,x} \omega_{b,S}(x) r_{b,S}(x) \ell_{b,S}(x)}{\sum_{b,x} \omega_{b,S}(x)}, \tag{11}$$

where smooth_{L_1} is the smooth-L_1 loss (Girshick, 2015). The center term moves the zero level set, and the two side terms preserve its local distance orientation. Averaging over valid subsets and equally over annotated and unlabeled branches yields $\mathcal{L}_{\mathrm{rib}}$; gradients pass only through the student field.

3.5 Training Objective

Combining the proposed constraints with the base objective yields the full training loss:

$$\mathcal{L} = \mathcal{L}_{\mathrm{pCE}} + r_{\mathrm{MT}}(t)\mathcal{L}_{\mathrm{MT}} + \lambda_{\mathrm{eik}}\mathcal{L}_{\mathrm{eik}} + \lambda_{\mathrm{anat}}\mathcal{L}_{\mathrm{anat}}(t) + \lambda_{\mathrm{rib}}r_{\mathrm{rib}}(t)\mathcal{L}_{\mathrm{rib}}, \tag{12}$$

where r_{MT} and r_{rib} govern the MT warm-up and the delayed ramp-up of the boundary-geometry constraint, respectively. Anatomical constraints use scribble-annotated images; Eikonal and Ribbon also use unlabeled images. The teacher is updated by exponential moving average of the student, and only the student is retained for inference. Loss weights and training settings are given in Section 4.2.

4 Experiments and Results

4.1 Datasets

We evaluate few-shot scribble-supervised segmentation on the ACDC (Bernard et al., 2018) and Prostate (Clark et al., 2013) datasets. The ACDC contains 100 cardiac MRI cases with left-ventricular cavity (LV), myocardium (MYO), and right-ventricular cavity (RV) as targets. The Prostate contains 80 T2-weighted MRI cases with central gland (CG) and peripheral zone (PZ) as targets. We follow the data splits and scribble annotations of BiSCL (Su et al., 2026): the training/validation/test splits contain $70/15/15$ cases for ACDC and $55/10/15$ for Prostate. In each dataset, only five training cases are scribble-annotated; the remaining 65 and 50 training cases, respectively, are treated as unlabeled.

4.2 Implementation Details

We implement our method in PyTorch with a natural-image-pretrained DINO ViT-S/8 encoder (Caron et al., 2021) and a lightweight convolutional segmentation head. Input slices are intensity-normalized, resized to 256×256, and augmented with random rotations and flips. Training runs for $10{,}000$ iterations on a single NVIDIA RTX 3090 GPU using AdamW, an initial learning rate of 10^{-4}, weight decay of 5×10^{-4}, and cosine annealing. Each batch contains six scribble-annotated and six unlabeled slices. For both datasets, we set $\lambda_{\mathrm{eik}} = 0.001$, $\lambda_{\mathrm{anat}} = 0.05$, and $\lambda_{\mathrm{rib}} = 0.01$. The teacher EMA coefficient is 0.99, and the pseudo-label confidence threshold is 0.9. The MT weight ramps up over the first $1{,}000$ steps; Ribbon starts at step $1{,}000$ and ramps up over the next $1{,}000$ steps. Reconstruction uses $K = 64$ propagation steps and is evaluated every $q = 4$ training steps. We use $\kappa = 0.1$ and $\epsilon = 10^{-6}$ for distance calibration. Further constraint construction details and computational overhead are reported in Appendices B and C, respectively.

4.3 Comparisons and Evaluation

We compare with the partial cross-entropy baseline PCE (Can et al., 2018), the augmentation-based method CycleMix (Zhang & Zhuang, 2022), and the scribble-supervised methods DMPLS (Luo et al., 2022), SC-Net (Zhou et al., 2023), ScribbleVC (Li et al., 2023), ScribFormer (Li et al., 2024b), QMaxViT-Unet+ (Nguyen-Tat et al., 2025), EFFDNet (Liu et al., 2025), and BiSCL (Su et al., 2026). Baselines follow official implementations where available. For pseudo-label-based methods applicable to both scribble-only and semi-supervised training, we report the better-performing setting. Reported standard deviations are computed across test cases unless stated otherwise.

Tables 1 and 2 present the quantitative comparisons on ACDC and Prostate, respectively, under the setting with five scribble-annotated cases. The proposed method achieves the best average Dice and HD95 on both datasets, while also maintaining leading performance across individual segmentation targets. Notably, on the more challenging Prostate dataset, our method still improves both regional segmentation and boundary localization. Figure 2 presents qualitative comparisons on the same cases, further illustrating the improvements in region integrity and boundary localization. In the first ACDC example, the anatomical structures remain discernible despite low overall contrast, yet most baselines miss substantial foreground regions or produce incomplete segmentations. Our method recovers the structures more completely and delineates their boundaries more accurately. These results further confirm that, in few-shot scribble-supervised segmentation, scarce supervision provides insufficient evidence for the model to infer reliable and discriminative anatomical semantics,

Table 1: Quantitative comparison on ACDC with five scribble-annotated cases. Best and second-best results are marked in red and blue, respectively.

Method	Dice ↑				HD95 (mm) ↓			
	RV	MYO	LV	Mean	RV	MYO	LV	Mean
PCE	0.261±0.160	0.183±0.034	0.311±0.143	0.251±0.084	114.26±10.48	108.96±12.02	115.25±14.28	112.82±11.26
CycleMix	0.516±0.239	0.604±0.141	0.697±0.183	0.606±0.166	45.20±20.47	50.52±18.61	49.32±32.85	48.35±18.26
DMPLS	0.327±0.216	0.254±0.051	0.337±0.190	0.306±0.119	82.76±25.25	106.74±11.38	106.37±18.41	98.62±14.43
SC-Net	0.378±0.228	0.250±0.082	0.265±0.148	0.298±0.116	82.64±25.99	108.22±13.62	110.58±15.30	100.48±15.08
ScribbleVC	0.500±0.244	0.650±0.138	0.736±0.163	0.628±0.159	35.16±31.53	21.72±20.83	34.83±36.22	30.57±23.02
ScribFormer	0.388±0.251	0.598±0.125	0.701±0.188	0.562±0.150	79.38±34.18	57.54±33.92	46.84±38.59	61.25±25.49
QMaxViT-Unet+	0.705±0.177	0.774±0.071	0.851±0.099	0.776±0.100	34.64±38.08	10.16±14.77	21.16±28.13	21.01±20.85
EFFDNet	0.765±0.211	0.794±0.071	0.638±0.186	0.732±0.108	5.73±13.03	14.95±19.37	80.46±22.79	33.71±10.34
BiSCL	0.776±0.132	0.819±0.073	0.889±0.087	0.828±0.082	7.78±15.02	2.09±1.00	6.85±17.30	5.57±7.35
Ours	0.851±0.092	0.848±0.039	0.900±0.080	0.866±0.056	1.50±0.78	1.28±0.30	1.52±1.22	1.44±0.65

Table 2: Quantitative comparison on Prostate with five scribble-annotated cases. Best and second-best results are marked in red and blue, respectively.

Method	Dice ↑			HD95 (mm) ↓		
	PZ	CG	Mean	PZ	CG	Mean
PCE	0.165±0.104	0.356±0.108	0.260±0.068	154.53±14.40	166.68±22.01	160.61±15.21
CycleMix	0.234±0.116	0.541±0.149	0.388±0.111	120.85±17.87	115.62±43.46	118.23±27.24
DMPLS	0.073±0.046	0.275±0.114	0.174±0.070	153.59±22.77	163.53±14.93	158.56±16.52
SC-Net	0.149±0.166	0.305±0.118	0.227±0.133	128.76±35.30	143.10±31.64	135.93±27.70
ScribbleVC	0.369±0.158	0.588±0.116	0.478±0.108	33.57±24.30	30.80±40.37	32.19±29.43
ScribFormer	0.144±0.139	0.349±0.191	0.247±0.075	143.59±33.29	161.23±30.76	152.41±24.23
QMaxViT-Unet+	0.180±0.073	0.557±0.116	0.368±0.065	119.31±17.34	113.55±50.99	116.43±31.66
EFFDNet	0.466±0.190	0.645±0.185	0.555±0.171	56.36±25.32	52.20±38.37	54.28±26.93
BiSCL	0.421±0.230	0.647±0.150	0.534±0.172	25.59±24.16	6.59±3.81	16.09±12.89
Ours	0.593±0.163	0.774±0.051	0.683±0.097	8.37±6.74	5.60±3.73	6.99±4.22

while methods that rely primarily on model-generated pseudo-labels are more prone to error accumulation and self-reinforcement. In contrast, the proposed SRDF organizes multiclass predictions into differentiable structured representations that jointly encode regional semantics and local boundary geometry, allowing anatomical priors and local image evidence to constrain region organization and boundary localization on objects matched to their respective semantic scopes. This provides more effective guidance for learning anatomical semantics under scarce supervision and ultimately improves segmentation performance.

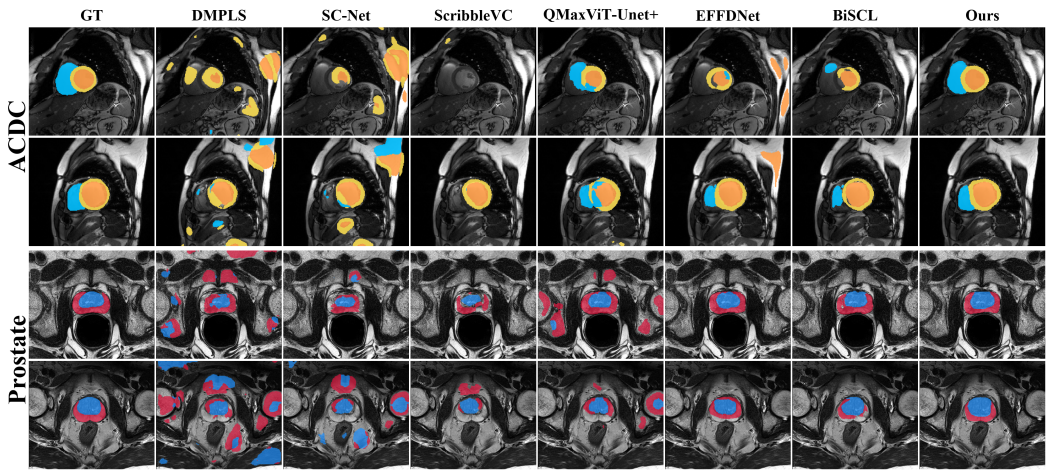


Figure 2: Qualitative comparison on ACDC and Prostate. Each dataset shows two representative slices. Colored overlays denote the predicted anatomical classes, and GT denotes the ground truth.

4.4 Ablation Studies

Table 3 examines the sequential component configurations under the same DINO backbone. The pCE+MT baseline achieves a Dice score of 0.804 on ACDC but still has substantial boundary error;

Table 3: Constraint ablation with a fixed DINO ViT-S/8 backbone. All configurations include pCE; ✓ indicates an enabled loss term.

ID	$\mathcal{L}_{\mathrm{MT}}$	$\mathcal{L}_{\mathrm{eik}}$	$\mathcal{L}_{\mathrm{adj}}$	$\mathcal{L}_{\mathrm{top}}$	$\mathcal{L}_{\mathrm{rib}}$	ACDC Dice ↑	ACDC HD95 ↓	Prostate Dice ↑	Prostate HD95 ↓
I	✓					0.804±0.093	22.66±18.38	0.345±0.086	137.88±27.31
II	✓	✓				0.827±0.069	6.08±11.24	0.385±0.104	131.56±21.91
III	✓	✓	✓			0.848±0.071	3.64±6.63	0.655±0.131	7.40±4.65
IV	✓	✓	✓	✓		0.852±0.067	2.60±5.49	0.679±0.101	7.42±5.41
V	✓	✓	✓	✓	✓	0.866±0.056	1.44±0.65	0.683±0.097	6.99±4.22

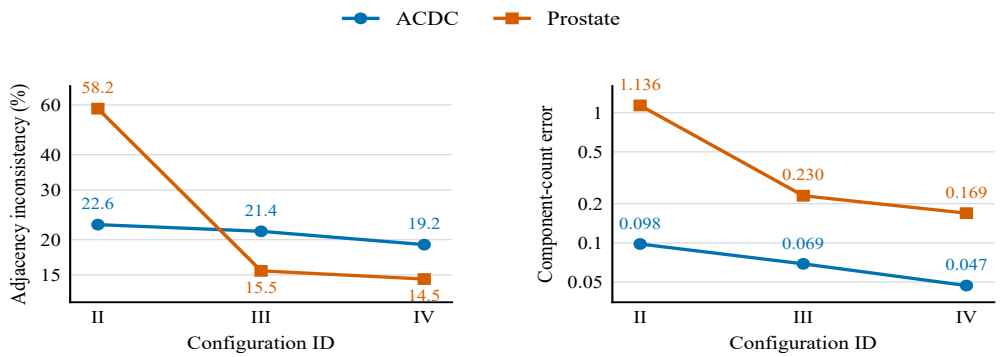


Figure 3: Structural correctness on the ACDC and Prostate test sets for configurations II–IV in Table 3. Left: percentage of slices whose predicted class-contact graph differs from the reference. Right: mean absolute component-count error on slices where the anatomical subset appears in the reference. Both vertical axes are logarithmic; lower is better.

on the more challenging Prostate dataset, its Dice is only 0.345. This indicates that, although pretraining provides general visual priors, such priors cannot ensure the learning of reliable anatomical semantics when supervision is extremely scarce. Adding Eikonal regularization alone yields some improvement, particularly in ACDC HD95, suggesting that stabilizing the gradient scale near zero level sets benefits boundary geometry modeling. Class adjacency further improves performance on both datasets: on Prostate, Dice rises from 0.385 to 0.655, accompanied by a substantial reduction in HD95. This highlights the structural information provided by inter-tissue relations under scarce supervision. Topological reconstruction further improves Dice on both datasets; on Prostate, Dice reaches 0.679 while HD95 remains near 7.4. Finally, Subset-Ribbon further reduces HD95 while improving Dice, demonstrating the value of local image evidence for boundary localization.

With the backbone held fixed, these progressive gains show that the improvement cannot be attributed to pretrained representations alone. Formulating anatomical priors and local image evidence as constraints on region organization and boundary localization through SRDF provides effective structural and geometric guidance for few-shot scribble supervision.

Figure 3 evaluates structural correctness for configurations II–IV on both test sets. Adjacency inconsistency and anatomical-subset component-count error decrease across these configurations, particularly on Prostate after adding class adjacency. These output-level measures complement the Dice and HD95 comparisons by showing more accurate anatomical organization.

4.5 Sensitivity Analysis

We further examine the effect of data partitioning using three random five-shot patient splits. As shown in Table 4, the proposed method consistently achieves the best performance compared with representative methods.

Table 4: Comparison across three random five-shot patient splits (mean±standard deviation).

Method	ACDC Dice ↑	ACDC HD95 ↓	Prostate Dice ↑	Prostate HD95 ↓
ScribbleVC	0.581±0.172	40.25±40.49	0.410±0.079	66.46±39.90
QMaxViT-Unet+	0.774±0.018	23.99±3.01	0.517±0.133	80.98±36.48
BiSCL	0.821±0.017	9.85±4.57	0.614±0.071	15.11±7.20
Ours	**0.849±0.018**	**1.75±0.27**	**0.694±0.008**	**7.47±2.29**

Table 5: Hyperparameter sensitivity on the ACDC dataset.

q	3	4	5	λ_{anat}	0.04	0.05	0.06
Dice ↑	0.859±0.060	0.866±0.056	0.861±0.059	Dice ↑	0.857±0.063	0.866±0.056	0.854±0.057
HD95 ↓	2.33±3.42	1.44±0.65	1.49±0.77	HD95 ↓	1.59±0.85	1.44±0.65	1.50±0.62
λ_{rib}	0.005	0.010	0.015	λ_{eik}	0.0005	0.0010	0.0015
Dice ↑	0.855±0.063	0.866±0.056	0.860±0.060	Dice ↑	0.858±0.060	0.866±0.056	0.857±0.059
HD95 ↓	1.57±0.86	1.44±0.65	1.54±0.80	HD95 ↓	1.53±0.71	1.44±0.65	1.81±1.76

Table 5 examines sensitivity to the topological evaluation interval q and three loss weights on ACDC, varying one parameter at a time while holding the others fixed. Performance remains broadly consistent across the tested ranges, indicating limited sensitivity to moderate variations in these parameters.

To investigate the influence of the pretrained backbone, Table 6 compares the complete method across different backbone architectures. The results show that the proposed method is compatible with both Transformer-based and convolutional backbones. Together with the fixed-backbone ablation, these findings suggest that its effectiveness is not tied to a specific pretrained encoder.

Table 6: Segmentation results with different pretrained backbones.

Pretrained backbone	ACDC		Prostate	
	Mean Dice ↑	Mean HD95 ↓	Mean Dice ↑	Mean HD95 ↓
Swin-Small (Liu et al., 2021)	0.833±0.067	3.89±8.60	0.689±0.097	29.73±37.02
ConvNeXt-Small (Liu et al., 2022)	0.848±0.070	4.61±10.54	0.660±0.104	8.34±4.60
DINO ViT-S/8 (Caron et al., 2021)	0.866±0.056	1.44±0.65	0.683±0.097	6.99±4.22

5 Conclusion

Under few-shot scribble supervision, limited pixel annotations make it difficult to learn reliable anatomical semantics. For segmentation tasks with well-defined anatomical relationships, available task-specific priors can provide additional learning evidence, yet the regional relations and boundary geometry they describe are difficult to translate directly into effective constraints. To address this challenge, we propose SRDF, which combines subset-relative decision fields with local distance calibration to organize multiclass predictions into a differentiable structured representation of region semantics and local boundary geometry. Building on this representation, we develop anatomical-structure constraints and Subset-Ribbon to guide region organization and boundary localization, respectively. Experiments on ACDC and Prostate show that our method outperforms the compared methods with only five scribble-annotated cases per dataset. Same-backbone ablations further show that these constraints improve performance under identical pretraining conditions. These findings indicate that SRDF provides a structured representation for translating task knowledge into learning constraints aligned with its semantic scope. The two constraints developed here demonstrate concrete uses of this representation; future work may explore other constraint mechanisms and assess their applicability to segmentation tasks with well-defined structural priors.

AI USE STATEMENT

Generative AI tools were used to assist with language editing, LaTeX formatting, and the examination of mathematical derivations. All AI-assisted suggestions were independently reviewed and verified by the authors, who take full responsibility for the final content of this work.

REPRODUCIBILITY STATEMENT

The main text specifies the method, dataset splits, network architecture, training settings, and evaluation protocol. Appendix A provides derivations for the properties of SRDF and finite-step reconstruction; Appendix B details the loss functions and constraint implementations; and Appendix C reports computational overhead. Except for the random-split comparison, reported standard deviations are computed across test cases; the random-split comparison reports statistics across three five-shot patient splits.

REFERENCES

Olivier Bernard, Alain Lalande, Clement Zotti, Frederick Cervenansky, Xin Yang, Pheng-Ann Heng, Irem Cetin, Karim Lekadir, Oscar Camara, Miguel Angel Gonzalez Ballester, 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.

N Byrne, J R Clough, I Valverde, G Montana, and A P King. A Persistent Homology-Based Topological Loss for CNN-Based Multiclass Segmentation of CMR. *IEEE Transactions on Medical Imaging*, 42(1):3–14, 2023. doi: 10.1109/TMI.2022.3203309.

Y B Can, K Chaitanya, B Mustafa, L M Koch, E Konukoglu, and C F Baumgartner. Learning to Segment Medical Images with Scribble-Supervision Alone. In *DLMIA*, 2018.

M Caron, H Touvron, I Misra, et al. Emerging Properties in Self-Supervised Vision Transformers. In *ICCV*, pp. 9650–9660, 2021.

Kenneth Clark, Bruce Vendt, Kirk Smith, John Freymann, Justin Kirby, Paul Koppel, Stephen Moore, Stanley Phillips, David Maffitt, Michael Pringle, et al. The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository. *Journal of Digital Imaging*, 26(6):1045–1057, 2013.

A Delong and Y Boykov. Globally Optimal Segmentation of Multi-Region Objects. In *ICCV*, 2009.

R Girshick. Fast R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1440–1448, 2015.

S Gupta, X Hu, J Kaan, et al. Learning Topological Interactions for Multi-Class Medical Image Segmentation. In *ECCV*, 2022.

Zhanghexuan Ji, Yan Shen, Chunwei Ma, and Mingchen Gao. Scribble-Based Hierarchical Weakly Supervised Learning for Brain Tumor Segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 175–183, 2019.

Shijie Li, Yunbin Tu, Qingyuan Xiang, and Zheng Li. MAGIC: Rethinking Dynamic Convolution Design for Medical Image Segmentation. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pp. 9106–9115, 2024a.

Zihan Li, Yuan Zheng, Xiangde Luo, Dandan Shan, and Qingqi Hong. ScribbleVC: Scribble-Supervised Medical Image Segmentation with Vision-Class Embedding. In *Proceedings of the 31st ACM International Conference on Multimedia*, pp. 3384–3393, 2023.

Zihan Li, Yuan Zheng, Dandan Shan, Shuzhou Yang, Qingde Li, Beizhan Wang, Yuanting Zhang, Qingqi Hong, and Dinggang Shen. ScribFormer: Transformer Makes CNN Work Better for Scribble-Based Medical Image Segmentation. *IEEE Transactions on Medical Imaging*, 43(6): 2254–2265, 2024b.

D Lin, J Dai, J Jia, K He, and J Sun. ScribbleSup: Scribble-Supervised Convolutional Networks for Semantic Segmentation. In *CVPR*, pp. 3159–3167, 2016.

Shiying Lin, Rong Hu, Zuoyong Li, Qinghua Lin, Jiawei Wu, and Changqing Zhang. Gradient-Aware Revitalization of Non-Effective Samples in Medical Image Segmentation. In *Proceedings of the 33rd ACM International Conference on Multimedia*, pp. 4728–4737, 2025.

J Liu, S Y Tan, X Yang, Y Xu, and S Y Yeo. EFFDNet: A Scribble-Supervised Medical Image Segmentation Method with Enhanced Foreground Feature Discrimination. In *MICCAI*, pp. 194–204, 2025.

Z Liu, Y Lin, Y Cao, et al. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. In *ICCV*, pp. 10012–10022, 2021.

Z Liu, H Mao, C-Y Wu, et al. A ConvNet for the 2020s. In *CVPR*, pp. 11976–11986, 2022.

Xiangde Luo, Minhao Hu, Wenjun Liao, Shuwei Zhai, Tao Song, Guotai Wang, and Shaoting Zhang. Scribble-Supervised Medical Image Segmentation via Dual-Branch Network and Dynamically Mixed Pseudo Labels Supervision. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 528–538, 2022.

C Matsoukas, J F Haslum, M Sorkhei, M Söderberg, and K Smith. What Makes Transfer Learning Work for Medical Images: Feature Reuse & Other Factors. In *CVPR*, pp. 9225–9234, 2022.

Thien B Nguyen-Tat, Hoang-An Vo, and Phuoc-Sang Dang. QMaxViT-Unet+: A Query-Based MaxViT-Unet with Edge Enhancement for Scribble-Supervised Segmentation of Medical Images. *Computers in Biology and Medicine*, 187:109762, 2025.

X-X Su, Y Ye, Y Zheng, M Gan, and G-Y Chen. Bi-Level Collaborative Learning for Few-Shot Scribble-Supervised Medical Image Segmentation. In *Proceedings of the 34th ACM International Conference on Multimedia*, 2026.

Antti Tarvainen and Harri Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

L Vincent. Morphological grayscale reconstruction in image analysis: applications and efficient algorithms. *IEEE Transactions on Image Processing*, 2(2):176–201, 1993.

Tao Wang, Xinlin Zhang, Zhenxuan Zhang, Yuanbo Zhou, Yuanbin Chen, Longxuan Zhao, Chaohui Xu, Shun Chen, Guang Yang, and Tong Tong. ScribbleVS: Scribble-Supervised Medical Image Segmentation via Dynamic Competitive Pseudo Label Selection. arXiv preprint arXiv:2411.10237, 2024.

Y Xue, H Tang, Z Qiao, et al. Shape-Aware Organ Segmentation by Predicting Signed Distance Maps. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, 2020.

Ke Zhang and Xiahai Zhuang. CycleMix: A Holistic Strategy for Medical Image Segmentation from Scribble Supervision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11656–11665, 2022.

Meng Zhou, Zhe Xu, Kang Zhou, and Raymond Kai-yu Tong. Weakly Supervised Medical Image Segmentation via Superpixel-Guided Scribble Walking and Class-Wise Contrastive Regularization. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 137–147, 2023.

R Zoetmulder, E Gavves, M Caan, and H Marquering. Domain- and task-specific transfer learning for medical segmentation tasks. *Computer Methods and Programs in Biomedicine*, 214:106539, 2022.

A PROPERTIES OF SRDF AND RECONSTRUCTION

Subset decision consistency and nesting. Let $a_S = \min_{k \in S} \phi_k$ and $b_S = \min_{j \notin S} \phi_j$, so that $\psi_S = a_S - b_S$. With a unique global winning class, the winner belongs to S if and only if $a_S < b_S$, or equivalently $\psi_S < 0$. Moreover, $p_k = \exp(-\phi_k)/\sum_j \exp(-\phi_j)$ gives $\max_{j \notin S} p_j / \max_{k \in S} p_k = \exp(a_S - b_S)$, yielding the probability form in Eq. equation 1. For $S \subset T \subsetneq \mathcal{C}$, $a_T \leq a_S$ and $b_T \geq b_S$, hence $\psi_T \leq \psi_S$ and the negative regions remain nested as the subset expands. This property preserves consistent region membership without prescribing spatial relations between tissues.

Local distance calibration. The denominator in Eq. equation 2 is strictly positive, so local normalization preserves the sign and zero level set of the relative decision field. When a fixed pair of strongest classes inside and outside the subset jointly dominates a boundary neighborhood, the subset field locally reduces to the difference between their potentials; the singleton Eikonal regularizer then constrains the gradient scale of that interface. At multiclass junctions or switches of the dominant classes, d_S remains a local first-order distance proxy rather than an exact global Euclidean distance.

Finite-step maximum-bottleneck reconstruction. For a path $\pi = (v_0, \ldots, v_\ell = x)$ with at most K edges, Eq. equation 6 can be written as

$$R^{(K)}(x) = \max_{\pi:\, \ell \leq K} \min\{s(v_0), u(v_1), \ldots, u(v_\ell)\}. \tag{13}$$

For $K = 0$, the expression contains only the seed value. Assuming it holds at $K = t$, the neighborhood term in the recurrence enumerates paths of length at most t extended by one edge to x, and the outer maximum retains the existing shorter paths. Induction therefore gives Eq. equation 13. Thus, $R^{(K)}$ is a length-limited continuous surrogate for maximum-bottleneck reachability used to optimize connectivity; it does not assert exact topology of the final discrete segmentation.

Why SRDF for Scarce Scribble Supervision? Under few-shot scribble supervision, annotations identify the classes of only a small number of pixels. They specify neither complete region extents nor reliable boundary locations. The task priors available in this setting, however, concern higher-order objects: class adjacency describes contact between tissues; connectivity and holes concern individual tissues or composite anatomical regions; and local image evidence concerns the boundaries of those regions. To use these priors for training, we need differentiable region and boundary representations whose semantic scopes match the priors while remaining consistent with the final multiclass segmentation. This motivates SRDF.

A direct construction for a target class subset S is to sum its class probabilities, $P_S = \sum_{k \in S} p_k$. Although P_S is the probability of membership in S, the region defined by $P_S > 0.5$ need not equal the union of classes selected by the final multiclass decision. For example, if $p = (0.40, 0.35, 0.25)$ and $S = \{1, 2\}$, then $P_S = 0.60$, while the predicted class is $0 \notin S$. A structural constraint imposed on the former region may therefore act on a different object from the one produced by the final segmentation. SRDF instead compares the strongest class responses inside and outside S. Its sign exactly matches subset membership under the multiclass decision, and nested class subsets yield nested decision regions, as proved below.

Decision consistency alone does not provide the geometry needed for boundary constraints: the magnitude of a subset-relative decision field measures class competition rather than spatial distance. SRDF therefore applies local gradient calibration. Near a smooth interface with a nondegenerate gradient, the calibrated field provides a first-order distance proxy while preserving the sign and zero level set of the original field. The resulting representation supports boundary bands for adjacency, distance-aware capacities for reconstruction, and the boundary locations and normals used by Subset-Ribbon. Conventional supervised signed-distance-map regression typically constructs distance targets from dense masks, which scribbles do not provide. A distance map could instead be derived from a predicted hard mask, but such a target would inherit errors in that prediction and introduce a nondifferentiable mask-construction step. SRDF obtains a local geometric proxy directly from the multiclass output without requiring dense distance targets.

We do not claim that anatomical or boundary priors cannot be imposed on pixelwise probability maps. Rather, SRDF provides a shared interface for the priors used in this work: individual and

composite regions remain consistent with the final multiclass decision, their boundaries acquire a local geometric scale, and each constraint acts at its intended semantic scope. The calibrated field is a local distance proxy, not an exact global signed distance transform; likewise, the proposed structural constraints do not guarantee the exact topology of the final discrete segmentation. Their purpose is to provide optimizable structural and geometric guidance when annotation is scarce.

B Implementation Details

Base supervision. Let $\mathcal{B}_L$ and $\mathcal{B}_U$ be the scribble-annotated and unlabeled image indices in a batch. The teacher parameters $\bar{\theta}$ are updated as an exponential moving average of the student parameters θ, with predictions $p_{\bar{\theta}}$ and p_θ. We use

$$\mathcal{L}_{\mathrm{pCE}} = \frac{1}{N_A} \sum_{b\in\mathcal{B}_L} \sum_{x\in A_b} \mathrm{CE}(p_{\theta,b}(x), Y_b(x)), \quad \mathcal{L}_{\mathrm{MT}} = \frac{1}{2} \sum_{\mathcal{B}\in\{\mathcal{B}_L,\mathcal{B}_U\}} \frac{1}{N_{\mathcal{B}}^{\eta}} \sum_{b\in\mathcal{B}} \sum_{x\in U_b^{\eta}} \mathrm{CE}(p_{\theta,b}(x), \bar{y}_b(x)), \tag{14}$$

where $N_A = \sum_{b\in\mathcal{B}_L} |A_b|$, U_b^{η} comprises unlabeled pixels satisfying $\max_c p_{\bar{\theta},b,c}(x) \geq \eta$, $\bar{y}_b(x) = \arg\max_c p_{\bar{\theta},b,c}(x)$, and $N_{\mathcal{B}}^{\eta} = \sum_{b\in\mathcal{B}} |U_b^{\eta}|$. Empty contributions are zero.

Anatomical relations. Anatomical segmentation often involves tissues with well-established spatial organization. Such knowledge can be obtained from anatomical references or clinical expertise, independently of dense segmentation annotations. Under few-shot scribble supervision, it provides structural evidence that sparse labels alone cannot convey. SRDF expresses these known relations through anatomical subsets, adjacency pairs, and reconstruction targets, turning them into differentiable constraints at the appropriate semantic level. The approach is most applicable when stable anatomical relations are available; morphology-dependent contacts remain unconstrained. The ACDC and Prostate instantiations are given below.

ACDC. Topological reconstruction operates on $S_1 = \{\mathrm{LV}\}$, $S_2 = \{\mathrm{LV}, \mathrm{MYO}\}$, and $S_3 = \{\mathrm{LV}, \mathrm{MYO}, \mathrm{RV}\}$, and is enabled on each slice according to the class presence confirmed by the scribbles. Foreground and background reconstruction constrain connectivity and anomalous holes, respectively, at each anatomical level. Reconstructing S_2 includes the normal left-ventricular cavity within the union and prevents it from being treated as an anomalous hole inside the myocardium. Subset-Ribbon uses the same hierarchy, with $P(S_1) = S_2$, $P(S_2) = S_3$, and S_3 as the root region. The required-contact and forbidden-contact relations are

$$\begin{aligned} \mathcal{E}^{+}_{\mathrm{ACDC}} &= \{(\mathrm{BG}, \mathrm{RV}), (\mathrm{BG}, \mathrm{MYO}), (\mathrm{RV}, \mathrm{MYO}), (\mathrm{MYO}, \mathrm{LV})\}, \\ \mathcal{E}^{-}_{\mathrm{ACDC}} &= \{(\mathrm{BG}, \mathrm{LV}), (\mathrm{RV}, \mathrm{LV})\}. \end{aligned} \tag{15}$$

Prostate. Topological reconstruction operates on the central gland $S_1 = \{\mathrm{CG}\}$ and the whole gland $S_2 = \{\mathrm{PZ}, \mathrm{CG}\}$, and is enabled on each slice according to the class presence confirmed by the scribbles. This subset hierarchy captures the topology of the central gland and the whole prostate, while Subset-Ribbon likewise treats S_2 as the parent region of S_1. Based on the relative tissue positions, the adjacency constraint defines $(\mathrm{BG}, \mathrm{PZ})$ and $(\mathrm{PZ}, \mathrm{CG})$ as contact relations. Because $(\mathrm{BG}, \mathrm{CG})$ contact varies with PZ morphology, this pair is left unconstrained. All constrained targets lie within the preprocessed field of view, and background reconstruction is initialized from the image border.

Field calibration. We compute ∇_h using central differences and set $\epsilon = 10^{-6}$ and $\kappa = 0.1$. For singleton fields $\psi_{bk} = \psi_{b,\{k\}}$, the boundary-weighted Eikonal objective is

$$\mathcal{L}_{\mathrm{eik}} = \frac{1}{2} \sum_{\mathcal{B}\in\{\mathcal{B}_L,\mathcal{B}_U\}} \frac{1}{Z_{\mathcal{B}}^{\mathrm{eik}}} \sum_{b\in\mathcal{B}} \sum_{k\in\mathcal{C}} \sum_{x\in\Omega} w_{bk}(x) H_1\left(\sqrt{\|\nabla_h \psi_{bk}(x)\|_2^2 + \epsilon} - 1\right), \tag{16}$$

where H_1 is smooth-L_1 with threshold 1, $q_{bk} = \mathrm{sigmoid}(-\psi_{bk})$, $w_{bk} = \mathrm{sg}[4q_{bk}(1-q_{bk})]$, and $Z_{\mathcal{B}}^{\mathrm{eik}} = \sum_{b\in\mathcal{B}} \sum_{k\in\mathcal{C}} \sum_{x\in\Omega} w_{bk}(x)$. The weight concentrates regularization near zero level sets; the two data branches are normalized separately and averaged equally. Stop-gradient in Eq. equation 2 makes subsequent constraints update the relative decision value through the numerator. At multiclass junctions or switches of the strongest classes, d_S remains a local first-order proxy rather than an exact global distance field.

Topological reconstruction. We run reconstruction every $q = 4$ training steps. The distance field is bilinearly mapped to a 64×64 grid without rescaling its amplitude, while adaptive max pooling maps the scribbles to the same grid. The first anchor in raster order initializes foreground reconstruction, and all remaining anchors contribute to the connection supervision; the image border initializes background reconstruction. Both branches run for $K = 64$ iterations, and the candidate sets are

$$\mathcal{I}_S = \{x : u_S^F(x) > \tfrac{1}{2},\ R_S^F(x) < \tfrac{1}{2},\ x \notin M_S\}, \qquad \mathcal{H}_S = \{x : u_S^B(x) > \tfrac{1}{2},\ R_S^B(x) < \tfrac{1}{2}\}. \quad (17)$$

By Eq. equation 5, $u_S^F > 1/2$ and $u_S^B > 1/2$ select locations at least half a pixel inside and outside the predicted boundary, respectively, defining strong propagation support through a local distance margin. Candidate sets are fixed during backpropagation. A term is zero when no corresponding target is active or its candidate set is empty.

Subset-Ribbon. Ribbon retains the 256×256 training resolution. Candidate normal offsets are $\{-1, 0, 1\}$. At each candidate, inner offsets $\{-1.5, -0.5\}$ and outer offsets $\{0.5, 1.5\}$ are paired with tangent offsets $\{-1, 0, 1\}$, producing six samples on each side. The tangent direction is $(-n_y, n_x)$, and all offsets are measured in pixels. The image descriptor is $F = [\hat{I}, \mathrm{mean}_c \sqrt{\|\nabla_h \hat{I}_c\|^2 + \epsilon}]$, where $\hat{I}$ is channel-wise standardized. Let D be the descriptor dimension, and let $\mu_\delta^\pm$ and $V_\delta^\pm$ denote the sample mean and root-mean-square dispersion on the two sides. We compute

$$B_\delta = \sqrt{\|\mu_\delta^- - \mu_\delta^+\|_2^2/D + \epsilon}, \qquad W_\delta = \tfrac{1}{2}(V_\delta^- + V_\delta^+). \quad (18)$$

Candidate scoring uses the normalized difference a_δ in Eq. equation 10, and positional discriminability is the gap between its largest and second-largest values, $r = a_{(1)} - a_{(2)}$. The teacher boundary-band weight is $w_\partial = 4\bar{q}(1 - \bar{q})$, where $\bar{q} = \mathrm{sigmoid}(-\bar{d}_S/\tau)$. For a non-root subset, the parent gate is $h_S = \mathrm{sigmoid}[-\bar{d}_{P(S)}(x_0)/\tau]$; for a root subset, $h_S = 1$. If v indicates that a sampling location is at least 2.5 pixels from the image border, then

$$\omega = w_\partial h_S v, \qquad \omega^{\mathrm{img}} = \omega r, \qquad \tau = 0.5. \quad (19)$$

B_δ and W_δ in Eq. equation 18 measure the between-side mean difference and average within-side dispersion of this descriptor. The student fits the local distance profile at normal offsets $\{-0.5, 0, 0.5\}$ using

$$\mathrm{smooth}_{L_1}(e) = \begin{cases} e^2, & |e| < 1/2, \\ |e| - 1/4, & \text{otherwise}, \end{cases} \quad (20)$$

where e is the residual between the predicted distance and its target. Gradients are stopped through the teacher prediction, image descriptors, candidate selection, scores, and weights; only the student distance field receives the Ribbon gradient.

C Computational Overhead

The student segmentation network contains 23.541 million parameters. On ACDC, training for 10,000 iterations with 256×256 inputs and a batch size of 12 takes approximately 2.29 hours on a single RTX 3090 and uses 12.08 GiB of peak CUDA memory. The EMA teacher and all constraints are used only during training. Inference uses only the student encoder and segmentation head, so the proposed constraints add no inference parameters or computation.