A More Details of Experimental Setup

A.1 Datasets

LASeg Dataset The Atrial Segmentation Challenge (LASeg) dataset provides 100 3D gadolinium-enhanced MR imaging scans with left atrium segmentation masks for training and validation. Following previous work [1, 2], we split the 100 scans into 80 for training and 20 for evaluation. We conduct experiments under two semi-supervised settings with distinct label ratios: 5% (4 labeled / 76 unlabeled volumes) and 10% (8 labeled / 72 unlabeled volumes). We use the processed datasets from [1], where all the scans were cropped to center on the heart region. This adjustment improves the comparison of segmentation performance across different methods. The datasets were also normalized to have a zero mean and unit variance.

MMWHS Dataset The Multi-Modality Whole Heart Segmentation (MMWHS) dataset comprises 20 sets of unregistered cardiac MR and CT volumes, enabling the evaluation of bidirectional segmentation transfer from MR to CT and vice versa. Following the previous work [3], we choose four classes of cardiac structures. They are the ascending aorta (AA), the left atrium blood cavity (LAC), the left ventricle blood cavity (LVC), and the myocardium of the left ventricle (MYO). For the pre-processing, follow [3], (1) all the scans were cropped centering at the heart region, with four cardiac substructures selected for segmentation; (2) for each 3D cropped image top 2% of its intensity histogram was cut off to alleviate artifacts; (3) each 3D image was then normalized to [0, 1]. To make a fair comparison, we keep the test set the same with prior arts [4, 3].

A.2 Baseline

We benchmark PreTriNet against the following methods that were widely adopted on the LASeg dataset. **V-Net** is a classic fully convolutional 3D architecture. **UA-MT** is an uncertainty-aware mean-teacher framework that weights consistency targets by voxel-wise uncertainty. **SASSNet** is a shape-aware strategy that jointly predicts signed distance maps to impose geometric constraints. **DTC** enforces dual-task consistency between segmentation and level-set representations. **URPC** rectifies multi-scale prediction consistency with uncertainty-guided region selection. **MC-Net** employs two decoders whose discrepancies generate cyclic pseudo-labels for challenging regions. **SS-Net** combines pixel-level smoothness under adversarial perturbations with class-level separation. **BCP** uses bidirectional copy-paste in a mean-teacher setup to bridge the distribution gap. **A&D** aggregates distribution-invariant features via a diffusion encoder and decouples the learning of labeled and unlabeled data. **CS-MT** introduces two alternating teacher models with feature correlation and adaptive CutMix. **LMISA-3D** achieves modality-invariant segmentation through gradient normalization, dual-task learning, and adversarial shape constraints. **SimCVD** performs contrastive distillation on signed distance maps with dropout augmentations. **MLRPL** learns from reliable pseudo-labels selected by confidence and intra-class similarity. **Kumari *et al.*** combines wavelet-based cross-component exchange, a re-weighted diffusion VNet, and cross-decoder pseudo training to handle noise and domain shift.

A&D, **SS-Net**, **BCP**, and **Kumari *et al.*** have already been described above; we therefore only detail the baselines unique to the MMWHS dataset below. **PnP-AdaNet** inserts a plug-and-play adversarial branch that replaces early encoder layers to align multi-scale features across domains. **AdaOutput** performs multi-level adversarial learning directly in the structured output space. **CycleGAN** translates images between modalities via cycle-consistent image-to-image adversarial training. **CyCADA** combines pixel- and feature-level cycle-consistent adversarial adaptation with task supervision. **SIFA** jointly aligns domains in both image and feature spaces with a shared encoder. **DSFN** builds dual source-to-target and target-to-source information flows and fuses their predictions for robust inference. **DSAN** employs a symmetric bidirectional alignment scheme with a shared encoder and private decoders to mitigate large domain shifts.

A.3 Implementation Details

We implement our proposed framework with PyTorch, using a single NVIDIA V100 GPU. The following sections provide a comprehensive overview of the training protocols and the hyperparameter settings used to ensure the reproducibility of our results. The network parameters are optimized using stochastic gradient descent (SGD) with Nesterov momentum of 0.9 and a weight decay of 3e-5. We employ a polynomial learning

rate decay schedule, with an initial learning rate of 0.001 and a power of 0.9. To enhance model generalization and prevent overfitting, we applied a comprehensive suite of online data augmentation techniques during training, including elastic deformation, rotation, scaling, and intensity transformations. The core components of our model, the adaptive confidence-driven fusion and the progressive curriculum learning, are governed by specific weighting coefficients. In the adaptive confidence-driven fusion, we set $\lambda_1 = 0.1$ and $\lambda_2 = 1.0$, following fine grid search. In the progressive curriculum learning, we set $\mu_1 = 1.0$ and apply a Gaussian ramp-up function for μ_2, ensuring a smooth transition from coarse semantic initialization to fine-grained structural refinement. A comprehensive hyperparameter configuration for PriTriNet is provided in Tables 1 and 2.

Table 1: Dataset-Specific Hyperparameters.

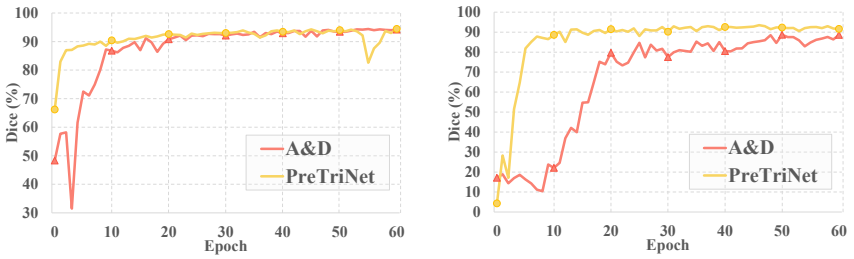
Dataset	Patch Size (D×H×W)	Learning Rate	Batch Size	Feature Size (F)
LASeg	$112 \times 112 \times 80$	3×10^{-2}	4	32
MMWHS	$128 \times 128 \times 128$	5×10^{-3}	4	32

Table 2: Shared Training Hyperparameters.

Parameter	Value
Optimizer	SGD (momentum=0.9, nesterov=True)
Weight Decay	3×10^{-5}
Learning Rate Scheduler	Polynomial Decay (Power=0.9)
Total Epochs	1000
Fusion Weight (λ_1)	0.1
Fusion Weight (λ_2)	1.0
Curriculum Weight (μ_1)	1.0 (fixed)
Curriculum Weight (μ_2)	Gaussian ramp-up (over 40 epochs)
Data Augmentation	Rotation, Scaling, Elastic Deformation, Intensity Shift

B More Analyses of Convergence Rate

Figure 1 illustrates the Dice score curves of PreTriNet and the baseline generative model A&D over the first 60 training epochs. On the LASeg dataset with 5% labeled data (Figure 1 (a)), PreTriNet quickly surpasses a Dice score of 80% within the first three epochs and achieves 90% by epoch 10, indicating rapid convergence. In contrast, A&D exhibits slower convergence, achieving comparable accuracy only after approximately 20 epochs. Similarly, for the MR to CT task on the MMWHS dataset (Figure 1 (b)), PreTriNet rapidly approaches 90% Dice within the first 10 epochs, while A&D shows slower progress, maintaining a Dice score under 30% during the same period.



(a) 5% labeled data (b) MR to CT

Figure 1: Convergence rate of generative (A&D) and our (PriTriNet) models on LASeg (a) and MMWHS (b) datasets.

C More Analyses of Ablation Experiments

To evaluate the contribution of each branch (GSG, LFR, UAB), we systematically remove each branch and analyze the impact on the final segmentation accuracy and training dynamics using the ASD metric, where lower values indicate better performance. Table 3 indicates that omitting any branch increases ASD values, suggesting a degradation in segmentation performance. On the LASeg dataset with 5% labeled data, eliminating the GSG branch leads to a significant increase in ASD, with a 50% higher value, while removing the UAB branch results in a 40% increase. Removing the LFR branch has the smallest impact, causing a 10% increase in ASD. A similar trend is observed for the 10% labeled data, with smaller variations in ASD values. On the MMWHS dataset, the impact of branch removal differs depending on the task. In the MR to CT task, the UAB branch removal causes the largest increase in ASD, with a 70% higher value, indicating its importance in boundary refinement. In the CT to MR task, removing the GSG branch leads to the greatest increase in ASD, with an 80% rise, highlighting its critical role in providing anatomical guidance, followed by the UAB branch.

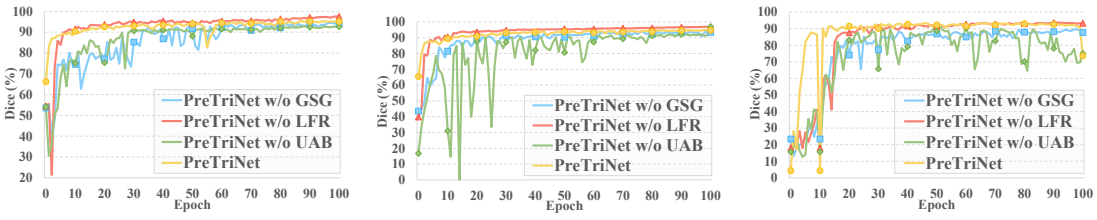
As shown in Figure 2 (a), the results indicate that on the LASeg dataset with 5% labeled data, PreTriNet achieves a Dice score of approximately 90% within 20 epochs. In contrast, the model without the GSG branch takes over 40 epochs to reach a similar performance level and exhibits significant fluctuations in its learning curve. Additionally, the absence of the UAB branch also delays convergence, while removing the LFR branch has a minimal impact on the initial training phase.

As shown in Figure 2 (b), similar trends are observed on the LASeg dataset with 10% labeled data, though the performance gaps are narrower.

On the MMWHS dataset, a key finding is that the importance of our proposed tri-branch is task-dependent. As shown in Figure 2 (c), in CT to MR translation, the GSG branch is the dominant factor. Its removal causes a sharp 18% performance drop and doubles the convergence time from 20 to over 40 epochs. For the MR to CT task, however, the Uncertainty-Aware Boundary (UAB) branch becomes more critical, with its absence leading to a 10% loss in Dice score. In comparison, the LFR branch has a lesser impact. These results confirm the effectiveness of our tri-branch design, with the complete PreTriNet model achieving a 90% Dice score within 20 epochs. In summary, the GSG branch enhances global alignment, the UAB branch stabilizes uncertain regions, and the LFR branch refines details, collectively enabling faster and more robust convergence across all tasks.

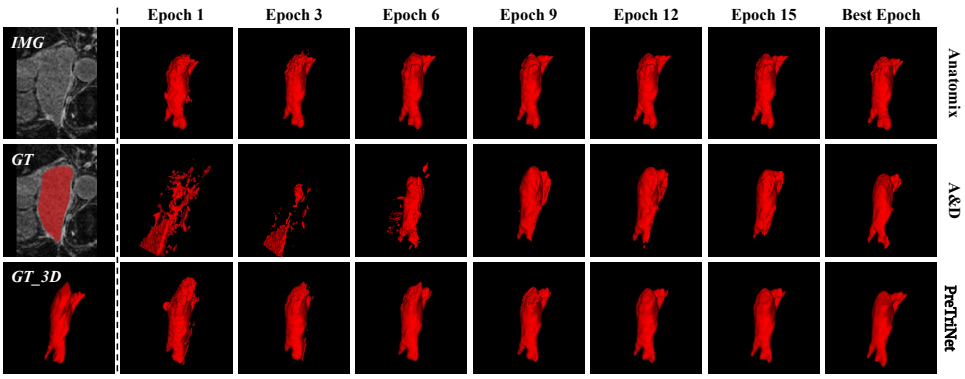
Table 3: Evaluation of ablation results on four tasks across two datasets with the ASD metric.

Methods	PreTriNet	w/o GSG	w/o LFR	w/o UAB
LA (5%)	**1.76**	3.70	1.94	3.59
LA (10%)	**1.60**	2.07	1.68	2.69
CT to MR	**1.30**	2.87	1.32	5.21
MR to CT	**3.00**	16.36	5.42	13.13

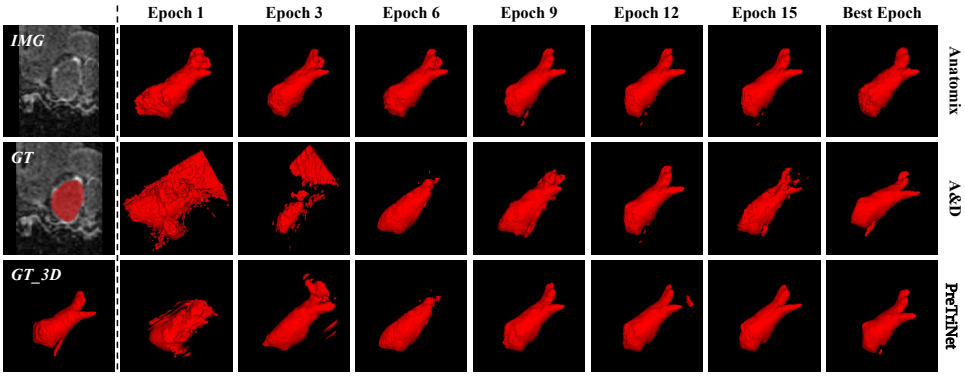


(a) 5% labeled data (b) 10% labeled data (c) MR to CT translation

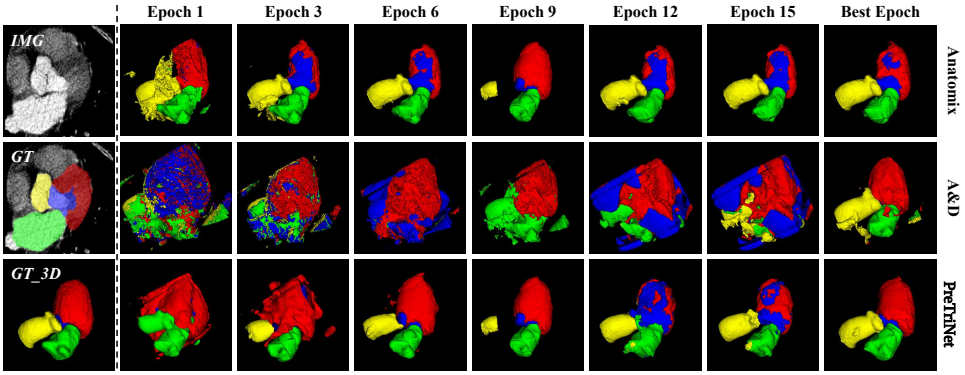
Figure 2: Training convergence curves on LASeg (a-b) and MMWHS (c) datasets.



(a) 5% labeled data



(b) 10% labeled data



(c) MR to CT

Figure 3: Comparative visualization of segmentation predictions for discriminative (Anatomix), generative (A&D), and our (PriTriNet) models on LASeg (a-b) and MMWHS (c) datasets.

D More Qualitative Analysis

Figure 3 (a) depicts the progressive delineation of the left-atrial segmentation over epochs in the semi-supervised setting with 5% labeled data. At Epoch 1, PreTriNet exhibits a stable and coherent boundary delineation, highlighting its robustness to sparse annotations. In contrast, Anatomix captures the gross anatomical shape but falls short in boundary detail and introduces extraneous segmentation boundaries that do not correspond to the ground truth. A&D performs poorly in the initial stages, yielding fragmented and incomplete segmentations. By the Best Epoch, PreTriNet outperforms the others in terms of boundary precision and overall anatomical consistency. The left atrium is outlined with greater fidelity, maintaining clear demarcations from adjacent structures. This visual evidence aligns with the quantitative metrics, confirming that the

synergistic integration of GSG, LFR, and UAB branches enables PreTriNet to achieve clinically reliable and structurally accurate segmentations even under conditions of sparse supervision.

Figure 3 (b) tracks the evolution of the left-atrial segmentation over epochs in the semi-supervised setting with 10% labeled data. PreTriNet matches Anatomix in rapid convergence yet surpasses both in final segmentation smoothness. By Epoch 9, PreTriNet establishes a coherent endocardial outline with distinct ostio-pulmonary junctions, contrasting with the coarse delineation of Anatomix and the fragmented mitral annulus of A&D. At the optimal epoch, PreTriNet achieves sub-millimeter precision, maintaining intact septal and appendage morphology. These observations align with quantitative metrics, validating its enhanced exploitation of limited labels for precise anatomical segmentation.

Figure 3 (c) illustrates the temporal evolution of four cardiac substructures-AA (yellow), LAC (green), LVC (red), and MYO (blue)-during the MR to CT task. by Epoch 9, PreTriNet achieves coherent segmentation of AA and LAC, with a well-defined LAC and a continuous MYO. In contrast, A&D exhibits fragmented myocardial walls and incomplete atrial segmentation, while Anatomix presents irregular boundaries along the interventricular septum. At the optimal epoch, PreTriNet produces anatomically accurate segmentations, with smooth delineation of the MYO without disruptions, clear demarcation of the LVC, and distinct separation of the AA and LAC from adjacent soft tissues. Anatomix shows under-segmentation of the atrial appendage and encroachment into the pericardial space, and A&D oversmooths the LVC, resulting in mislabeling of the ventricular cavity. These visual observations align with quantitative metrics, validating the ability of our proposed tri-branch architecture to deliver reliable and accurate segmentations under semi-supervised conditions.

References

[1] Lequan Yu, Shujun Wang, Xiaomeng Li, Chi-Wing Fu, and Pheng-Ann Heng. Uncertainty-aware self-ensembling model for semi-supervised 3d left atrium segmentation. In *MICCAI*, pages 605–613, 2019.

[2] Yicheng Wu, Zhonghua Wu, Qianyi Wu, Zongyuan Ge, and Jianfei Cai. Exploring smoothness and class-separation for semi-supervised medical image segmentation. In *MICCAI*, pages 34–43, 2022.

[3] Cheng Chen, Qi Dou, Hao Chen, Jing Qin, and Pheng Ann Heng. Unsupervised bidirectional cross-modality adaptation via deeply synergistic image and feature alignment for medical image segmentation. *TMI*, 39(7):2494–2505, 2020.

[4] Qi Dou, Cheng Ouyang, Cheng Chen, Hao Chen, Ben Glocker, Xiahai Zhuang, and Pheng-Ann Heng. Pnp-adanet: Plug-and-play adversarial domain adaptation network at unpaired cross-modality cardiac segmentation. *IEEE Access*, 7:99065–99076, 2019.