




A Semi-Supervised Framework for Cardiac MRI Segmentation via Multi-Constraint Collaborative Self-Training

Rui Chen¹, Zhi Huang², Xisheng Yan¹, Fen Zhu¹, Chaokun Guan¹, Xuan Ke¹, Chunyang Zhang¹, Jie Zhao¹ 
Dongsheng Li^{1*}

¹ Department of Cardiovascular Medicine, Wuhan Third Hospital & Tongren Hospital of Wuhan University & Jiangnan University, Wuhan 430074, China

² Department of Anesthesiology, Wuhan Fourth Hospital, Wuhan 430000, China

Corresponding Author Email: lidongsheng693@sina.com

Copyright: ©2025 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ts.420343>

ABSTRACT

Received: 20 December 2024

Revised: 9 June 2025

Accepted: 16 June 2025

Available online: 30 June 2025

Keywords:

cardiac magnetic resonance imaging, medical image segmentation, pseudo-label quality control, consistency constraints, adversarial learning

To address the challenges of limited annotated data and the adverse impact of noisy pseudo-labels on model generalization in cardiac magnetic resonance imaging (MRI) segmentation, a novel semi-supervised segmentation framework based on multi-constraint collaborative self-training was developed. The proposed approach integrates uncertainty-guided pseudo-label quality control, prototype-driven inter- and intra-class consistency constraints, and a multi-scale adversarial learning mechanism. During each iteration of self-training, a dynamic selection of pseudo-labels generated by unlabeled samples was conducted using a method that fuses confidence and entropy-based uncertainty quantification, thereby enhancing the reliability of the pseudo-supervision signals. Simultaneously, class-specific prototype vectors were dynamically maintained to enforce explicit constraints that encourage intra-class feature aggregation and inter-class feature separability, improving the discriminative capacity of the feature space. In addition, both global and local discriminators were introduced to impose dual-level quality constraints on the global morphology and local structural details of the segmentation outputs, resulting in refined boundary delineation and enhanced structural consistency. Extensive experiments conducted on the publicly available ACDC dataset demonstrated that the proposed method achieved an average Dice Similarity Coefficient (DSC) of 0.730 with only 5% of the annotations, outperforming existing methods such as Deep Co-Training (DCT) (0.711) and Mean Teacher (MT) (0.654) and approaching the performance of full supervision (0.891). When the annotation ratio was increased to 10%, the average DSC further improved to 0.801, consistently surpassing all comparative methods. Ablation studies confirmed the effectiveness of each key module. This study provides an efficient and robust solution for automatic segmentation of medical images in low-resource scenarios, offering promising potential for real-world clinical applications.

1. INTRODUCTION

Cardiovascular diseases remain among the leading causes of mortality worldwide, underscoring the critical importance of accurate diagnosis and treatment assessment for improving patient outcomes [1]. Cardiac MRI has been established as the gold standard for evaluating cardiac function and structure, owing to its high-resolution soft tissue contrast and capacity to provide multi-dimensional anatomical information [2]. In cardiac MRI analysis, precise segmentation of the left ventricle, right ventricle, and myocardium serves as the foundation for calculating key cardiac functional parameters, including ejection fraction, ventricular volume, and myocardial mass. These parameters are essential for clinical diagnosis and the formulation of therapeutic strategies [3, 4]. However, due to the structural complexity of the heart, morphological variability across individuals, and variations in image quality, accurate segmentation remains a technically

challenging task [5].

Conventional cardiac MRI segmentation methods primarily rely on techniques such as thresholding, region growing, active contour models, and graph cuts [6, 7]. Early deformable model-based approaches utilized statistical shape modeling to delineate cardiac structures, yet these methods were sensitive to initialization and struggled to accommodate complex morphological changes [8]. Multi-atlas registration methods, while capable of incorporating prior anatomical knowledge, typically involve high computational complexity and stringent requirements for registration accuracy [9]. Overall, these traditional approaches have demonstrated limitations in terms of sensitivity to image quality, poor generalization capacity, and dependence on extensive manual parameter tuning—factors that hinder their applicability in automated and robust clinical workflows [10].

Recent advances in deep learning have introduced transformative progress in medical image segmentation [11].

Convolutional neural networks (CNNs), particularly U-Net and its variants, have shown remarkable performance in cardiac MRI segmentation by automatically learning image feature representations and enabling end-to-end segmentation [12, 13]. As network architectures have evolved—from the foundational fully convolutional networks (FCNs) [14], to U-Net [15], and further to attention mechanisms [16], residual connections [17], and Transformer-based architectures [18]—notable improvements have been achieved in both segmentation accuracy and robustness. Supervised learning approaches, when trained on sufficiently annotated datasets, have been shown to reach expert-level segmentation performance, with their efficacy validated across multiple cardiac MRI datasets [19, 20].

Despite the success of fully supervised deep learning methods, their performance is critically dependent on the availability of large volumes of high-quality annotated data [21]. In cardiac MRI segmentation tasks, pixel-level annotations must be meticulously delineated by expert radiologists, often requiring several hours to manually label a complete cardiac MRI sequence layer by layer [22]. Moreover, inter-expert variability, subjectivity in interpreting complex anatomical structures, and the difficulty of ensuring annotation quality further increase the cost and complexity of acquiring reliable training data [23]. This dependence on extensive annotations significantly limits the scalability of deep learning models in resource-constrained environments and impedes their rapid deployment in novel domains and datasets.

To mitigate the reliance on manual annotations, semi-supervised learning has emerged as a prominent research direction in medical image segmentation [24]. Existing semi-supervised segmentation strategies generally fall into three categories: (a) consistency regularization, which enforces consistency under perturbations of the same input to encourage the learning of robust feature representations [25]; (b) generative adversarial networks (GANs), which employ discriminators to distinguish between ground-truth labels and predicted results, thereby enhancing segmentation quality through adversarial training [26]; and (c) self-training with pseudo-labels, in which high-confidence predictions generated by the model itself are used as pseudo-labels to expand the training dataset [27]. Among these, self-training has gained considerable attention due to its conceptual simplicity and ease of implementation. However, several critical limitations remain in current self-training approaches. First, the quality control of pseudo-labels remains inadequate. Most methods adopt static thresholds or simple confidence-based filtering strategies, which are prone to introducing noisy labels that degrade performance through the accumulation of errors [28]. Second, effective consistency constraints are often lacking. Existing approaches have not fully exploited the intrinsic structure of data or the inter-class relationships, resulting in inadequate enforcement of feature space consistency [29]. Third, instability during training poses a persistent challenge. Confirmation bias often emerges during the self-training process, where the model tends to reinforce its own erroneous predictions in the absence of effective error correction mechanisms [30].

To address the limitations of existing self-training-based semi-supervised segmentation approaches, a framework for semi-supervised cardiac MRI segmentation was proposed based on multi-constraint collaborative self-training. This framework enhances the effectiveness of semi-supervised learning through the coordinated operation of three key

components. First, uncertainty-guided pseudo-label quality control was implemented by integrating prediction confidence and entropy-based uncertainty quantification. A dynamic thresholding strategy was adopted to perform pixel-level refinement of pseudo-label selection, effectively mitigating the accumulation of errors caused by low-quality pseudo-labels. Second, by dynamically maintaining feature prototype vectors for each anatomical category, prototype-guided inter- and intra-class consistency constraints impose explicit constraints in the feature space to encourage intra-class feature aggregation and inter-class feature separability. This enables the establishment of multi-level consistency constraints between labeled and unlabeled samples, thereby improving the discriminative power of feature representations. Third, a global discriminator and a region-specific discriminator designed by the segmentation quality enhancement strategy based on adversarial learning impose quality constraints on the segmentation output, targeting both overall shape integrity and local structural details, thereby encouraging the segmentation network to generate more accurate and realistic masks.

The principal contributions of this study can be summarized as follows:

(a) An uncertainty-guided pseudo-label quality control mechanism was proposed, combining confidence and entropy with a dynamic thresholding scheme to refine selection and enhance the reliability of pseudo-supervision signals.

(b) A prototype-guided inter- and intra-class consistency constraint framework was developed, in which dynamically updated category prototypes guide feature learning to improve intra-class aggregation and inter-class separability.

(c) A multi-scale adversarial learning strategy was constructed, employing both global and region-specific discriminators to jointly enhance the precision of segmentation boundaries and the anatomical consistency of structures.

(d) A unified multi-constraint collaborative training framework was established by integrating uncertainty quantification, prototype learning, and adversarial training, wherein all components interact synergistically to improve overall model performance.

2. MULTI-CONSTRAINT COLLABORATIVE SELF-TRAINING FRAMEWORK FOR SEMI-SUPERVISED CARDIAC MRI SEGMENTATION

2.1 Overall architecture of the multi-constraint collaborative self-training framework

To address the challenge of limited annotated data in cardiac MRI segmentation, a semi-supervised segmentation framework based on a three-stage iterative strategy was proposed, as illustrated in Figure 1. In this framework, high-quality pseudo-labels were progressively incorporated to alleviate the error accumulation commonly associated with conventional self-training methods.

Step 1: Fully supervised training under limited annotations

The segmentation network was initially trained using a small set of annotated cardiac MRI data to establish a foundational capacity for cardiac structural segmentation. This phase provides a reliable model base for subsequent pseudo-label generation.

Step 2: Pseudo-label generation for unlabeled samples

Inference prediction was performed on a large volume of

unlabeled data using the initially trained model. A dual uncertainty quantification mechanism—combining prediction confidence and entropy—was employed to assess pixel-wise prediction reliability. An adaptive selection process was implemented using a dynamic thresholding strategy. During the early training process, strict criteria were applied to ensure quality, which were progressively relaxed as the model performance improved, thereby maximizing the utility of unlabeled data.

Step 3: Multi-constraint collaborative self-training

Labeled data and filtered pseudo-labeled data were jointly utilized for collaborative training through a multi-constraint scheme. The framework includes several mechanisms: a supervised loss ensures the accuracy; prototype-guided consistency constraints maintain class-specific prototype vectors to enforce intra-class feature aggregation and inter-class feature separation; and both global and region-specific discriminators designed by adversarial learning impose multi-granularity quality constraints on the segmentation outputs.

Steps 2 and 3 form a closed-loop iterative mechanism. High-quality pseudo-labels were leveraged for multi-constraint training, and the improved model subsequently generated more accurate pseudo-labels. This progressive learning paradigm mitigates error accumulation and facilitates iterative enhancement of both pseudo-label quality and model performance until convergence is achieved. Through the proposed three-stage iterative architecture, large-scale unlabeled data can be effectively utilized under highly constrained annotation conditions, providing an efficient and robust solution for semi-supervised cardiac MRI segmentation.

2.2 Uncertainty-guided pseudo-label self-training quality control

Conventional self-training methods are highly susceptible

to the influence of low-quality pseudo-labels, often resulting in cumulative errors. This issue is particularly pronounced in cardiac MRI segmentation due to the complexity of anatomical structures and the ambiguity of boundaries. To mitigate this challenge, a pixel-wise pseudo-label quality control mechanism was proposed, integrating prediction confidence and entropy to derive a reliability score. A dynamic thresholding strategy was then applied to perform fine-grained filtering, thereby ensuring the provision of high-quality pseudo-supervision signals during training.

To reduce computational overhead caused by multiple inferences, prediction uncertainty was quantified using a computationally efficient single forward propagation strategy. For each pixel location (i, j) in the output of the segmentation network, a softmax activation function was applied to obtain a categorical probability distribution $\{p_{i,j}^c\}_{c \in \mathcal{C}}$, where $\mathcal{C} = \{BG, LV, RV, MYO\}$ represents the four segmentation classes: background, left ventricle, right ventricle, and myocardium.

Prediction confidence is defined as the probability of the model predicting the most likely category:

$$C(x_{i,j}) = \max_{c \in \mathcal{C}} p_{i,j}^c \quad (1)$$

Confidence reflects the model's certainty in its prediction results. A higher confidence value indicates a stronger belief in classifying the pixel. Prediction entropy was employed to measure the uncertainty of the probability distribution:

$$H(x_{i,j}) = - \sum_{c \in \mathcal{C}} p_{i,j}^c \log p_{i,j}^c \quad (2)$$

Lower entropy values indicate a more concentrated prediction distribution, whereas higher entropy signifies greater uncertainty in the classification of that pixel.

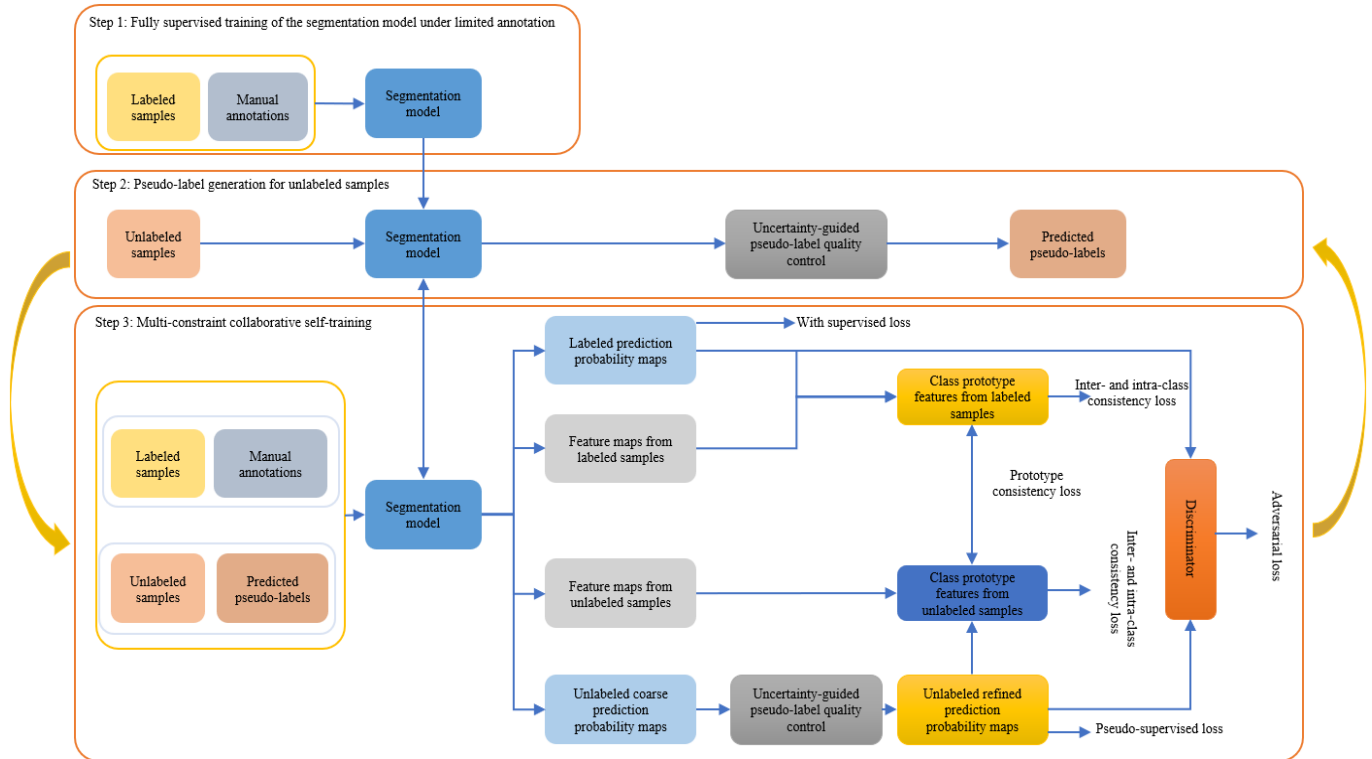


Figure 1. Overall workflow of the multi-constraint collaborative self-training framework

To further exploit the complementary characteristics of prediction confidence and entropy, a fused scoring mechanism was introduced to assess the pixel-level prediction reliability. The computation of the pixel-wise reliability score is defined as follows:

$$R(x_{i,j}) = C(x_{i,j}) \cdot \exp(-H(x_{i,j})) \quad (3)$$

The scoring function integrates multiplicatively confidence and entropy to ensure that only predictions exhibiting both high confidence and low entropy receive high reliability scores. The exponential function introduces a nonlinear modulation of the entropy component, allowing for finer discrimination between high- and low-quality predictions.

Given the progressive improvement of the model during training, a dynamic thresholding strategy was adopted to adaptively adjust the selection criteria:

$$\tau_t = \tau_0 \cdot \left(1 - \frac{t}{T}\right)^{0.5} \quad (4)$$

where, $\tau_0 = 0.7$, t denotes the current training iteration, and T is the total number of training iterations. This strategy enables the model to implement strict selection criteria in the early training stages. The criteria progressively relax as model capacity improves over time, thereby enhancing the utilization of unlabeled data.

Based on the computed reliability score and dynamic threshold, a pixel-wise selection mechanism is defined as:

$$M_{i,j} = I[R(x_{i,j}) > \tau_t] \quad (5)$$

where, $I[\cdot]$ denotes the indicator function. When $M_{i,j} = 1$, it indicates that the pseudo-label of pixel (i, j) has passed the quality check, and the selected pixel region participates in the subsequent model training.

2.3 Prototype-guided inter- and intra-class consistency constraint

In semi-supervised learning, reliance solely on pixel-wise classification loss has been found insufficient for learning highly discriminative feature representations. Due to morphological and textural similarities among cardiac anatomical structures, feature confusion may arise, often resulting in blurred segmentation boundaries. To mitigate this issue, a prototype-guided inter- and intra-class consistency constraint mechanism was introduced. This mechanism dynamically maintains and updates class-specific prototype vectors to explicitly constrain intra-class feature aggregation and inter-class feature separability in the feature space, thereby promoting more discriminative representation learning.

(a) Inter- and intra-class consistency constraint for labeled samples

For labeled samples, reliable class-specific prototypes were constructed based on ground truth annotations. The feature extractor g_ϕ maps the input image into the feature space. For the class $c \in \{0, 1, 2, 3\}$ (corresponding to background, left ventricle, right ventricle, and myocardium), the prototype vector is defined as:

$$p_c^l = \frac{1}{|F_c^l|} \sum_{(i,j) \in F_c^l} g_\phi(x^l)_{i,j} \quad (6)$$

where, $F_c^l = \{(i, j) \mid y_{i,j}^l = c\}$ denotes the set of pixel positions belonging to class c , and $g_\phi(x^l)_{i,j}$ represents the feature vector at location (i, j) .

Based on these class-specific prototypes, the intra-class consistency loss for labeled samples encourages feature vectors of the same class to aggregate toward their corresponding prototype centers, thus enhancing intra-class feature compactness. The computation is formulated as:

$$L_{intra}^l = \frac{1}{C} \sum_{c=0}^{C-1} \frac{1}{|F_c^l|} \sum_{(i,j) \in F_c^l} |g_\phi(x^l)_{i,j} - p_c^l|_2^2 \quad (7)$$

To further improve the separability between different classes, an inter-class separation loss was introduced by maximizing the distance between class prototypes, promoting the separation of inter-class features:

$$L_{inter}^l = -\frac{2}{C(C-1)} \sum_{c=0}^{C-1} \sum_{k=c+1}^{C-1} |p_c^l - p_k^l|_2^2 \quad (8)$$

(b) Inter- and intra-class consistency constraint for unlabeled samples

For unlabeled samples, the absence of ground truth annotations necessitates the construction of prototypes based on model predictions and the outcomes of pseudo-label quality screening. To account for the unreliability of pseudo-labels, only high-reliability pixels that satisfy the quality control criteria were used to update the prototypes:

$$p_c^u = \frac{1}{|F_c^u|} \sum_{(i,j) \in F_c^u} g_\phi(x^u)_{i,j} \quad (9)$$

where, $F_c^u = \{(i, j) \mid y_{i,j}^u = c\}$ denotes the set of pixel positions in the unlabeled samples that are predicted as class c and have passed the quality filtering process. The corresponding consistency constraint for unlabeled samples is then defined as:

$$L_{intra}^u = \frac{1}{C} \sum_{c=0}^{C-1} \frac{1}{|F_c^u|} \sum_{(i,j) \in F_c^u} |g_\phi(x^u)_{i,j} - p_c^u|_2^2 \quad (10)$$

$$L_{inter}^u = -\frac{2}{C(C-1)} \sum_{c=0}^{C-1} \sum_{k=c+1}^{C-1} |p_c^u - p_k^u|_2^2 \quad (11)$$

(c) Prototype-guided inter- and intra-class consistency constraint across labeled and unlabeled samples

To further enhance the consistency of feature representations, a prototype alignment constraint was introduced between labeled and unlabeled samples. This constraint was enforced by minimizing the discrepancy between the prototypes derived from labeled and unlabeled samples for the same class, thereby promoting the uniformity of feature distributions in both data domains:

$$L_{align} = \frac{1}{C} \sum_{c=0}^{C-1} |p_c^l - p_c^u|_2^2 \quad (12)$$

This constraint ensures consistency in feature

representations learned across the labeled and unlabeled data, thereby mitigating domain shift. To ensure stable prototype updates, an exponential moving average (EMA) strategy was adopted for smooth updates:

$$p_c^{(t+1)} = \mu p_c^{(t)} + (1 - \mu) p_c^{new} \quad (13)$$

where, $\mu \in [0.9, 0.99]$ is the momentum parameter, and p_c^{new} denotes the prototype vector computed from the current batch.

The final prototype-guided consistency loss is defined as:

$$L_{proto} = \lambda_1 (L_{intra}^l + L_{intra}^u) + \lambda_2 (L_{inter}^l + L_{inter}^u) + \lambda_3 L_{align} \quad (14)$$

where, λ_1 , λ_2 , and λ_3 are the balancing weights, all empirically set to 1 in this study.

2.4 Adversarial learning-based segmentation quality enhancement strategy

Conventional semi-supervised segmentation methods often lack high-level semantic constraints that ensure both global quality and region-specific accuracy in segmentation results. Although pixel-level accuracy may be achieved, limitations persist in maintaining global morphological consistency and achieving fine-grained delineation of anatomical structures. To address these issues, an adversarial learning-based segmentation quality enhancement strategy was introduced. This strategy employs both a global discriminator and a region-specific discriminator to assess and constrain segmentation outputs at multiple granularity levels, thereby encouraging the generation of more realistic and accurate segmentation masks.

(a) Global discriminator loss

The global discriminator D_g is designed to evaluate the quality of segmentation outputs from a holistic perspective. It discriminates between ground truth annotation masks and the masks predicted by the model, constraining the segmentation network to generate more realistic segmentation results. Let $y \in \mathbb{R}^{H \times W \times C}$ denote the ground truth segmentation mask and $\hat{y} \in \mathbb{R}^{H \times W \times C}$ the predicted mask, with $C = 4$ representing the total number of classes.

The training objective of the global discriminator is defined as:

$$L_{D_{global}} = -E_{y \sim p_{real}} [\log D_g(y)] - E_{\hat{y} \sim p_{pred}} [\log (1 - D_g(\hat{y}))] \quad (15)$$

Correspondingly, the global adversarial loss for the segmentation network is defined as:

$$L_{adv_{global}} = -E_{\hat{y} \sim p_{pred}} [\log D_g(\hat{y})] \quad (16)$$

This loss encourages the segmentation network to generate high-quality masks that can deceive the discriminator, thereby improving the realism and coherence of segmentation outputs at a global level.

(b) Region-specific discriminator loss

To account for the morphological differences among various cardiac anatomical structures, a region-specific discriminator D_r was further designed to individually evaluate the quality of the left ventricle (LV), right ventricle (RV), and

myocardium (MYO).

For each anatomical region $r \in \{LV, RV, MYO\}$, the corresponding segmented result was first extracted using the mask:

$$M_r = I[\hat{y} = r] \odot \hat{y} \quad (17)$$

where, $I[\cdot]$ denotes the indicator function, and \odot represents element-wise multiplication.

The objective function for the region-specific discriminator is formulated as:

$$L_{D_r} = -E_{M_r^{real}} [\log D_r(M_r^{real})] - E_{M_r^{pred}} [\log (1 - D_r(M_r^{pred}))] \quad (18)$$

The corresponding regional adversarial loss is defined as:

$$L_{adv_r} = -E_{M_r^{pred}} [\log D_r(M_r^{pred})] \quad (19)$$

The final region-specific adversarial loss is defined as:

$$L_{adv_{region}} = \sum_{r \in \{LV, RV, MYO\}} L_{adv_r} \quad (20)$$

Finally, the total adversarial loss integrates both the global and region-specific constraints:

$$L_{adv} = \lambda_g L_{adv_{global}} + \lambda_r L_{adv_{region}} \quad (21)$$

where, λ_g and λ_r are the balancing parameters for the global and region-specific adversarial losses, respectively. In this study, both coefficients were empirically set to 0.5.

2.5 Overall loss function

The total loss function of the proposed framework consists of multiple constraint components, including the supervised loss, pseudo-supervised loss, prototype-guided consistency loss, and adversarial loss:

$$L_{total} = L_{sup} + \alpha L_{pseudo} + \beta L_{proto} + \gamma L_{adv} \quad (22)$$

where, L_{sup} denotes the supervised loss of labeled samples, and L_{pseudo} represents the pseudo-supervised loss of unlabeled samples. Both were calculated using a combination of the standard cross-entropy loss function and the Dice loss function. The coefficients α , β , and γ are the balancing parameters used to control the relative contributions of the pseudo-supervised loss of unlabeled samples, prototype-guided consistency loss, and adversarial loss, respectively.

3. EXPERIMENTS AND RESULT ANALYSIS

3.1 Dataset

All experiments and comparative analyses were conducted using the publicly available benchmark dataset ACDC [31]. The ACDC dataset comprises 200 annotated short-axis cine cardiac MRI sequences from 100 patients. Segmentation masks are provided for the left ventricle, myocardium, and right ventricle, supporting both clinical and algorithmic

research. Each patient contributes two annotated sequences corresponding to end-diastole (ED) and end-systole (ES) phases. Each sequence originally consists of approximately 10 images, from which the five central slices were selected in this study, yielding a total of 1,000 images for analysis.

The dataset was partitioned based on patient IDs. A total of 70 patients (140 sequences, 700 images, comprising 350 ED and 350 ES images) were randomly selected for the training set, while the remaining 30 patients (60 sequences, 300 images, comprising 150 ED and 150 ES images) were assigned to the validation set.

To simulate semi-supervised learning conditions, subsets of the training data were designated as labeled and unlabeled. Specifically, 3 patients (6 sequences, 30 images) and 7 patients (14 sequences, 70 images) were randomly selected to form labeled datasets corresponding to approximately 5% and 10% of the training set, respectively. The remaining 67 patients (134 sequences, 670 images) and 63 patients (126 sequences, 630 images) were treated as unlabeled data. This setting reflects the practical clinical scenario in which annotated medical data are typically scarce.

3.2 Dataset preprocessing

Given the relatively large inter-slice spacing of the ACDC dataset, a 2D segmentation approach was adopted instead of a direct 3D segmentation strategy. The preprocessing procedure involved the following steps:

First, all slices were rescaled to a uniform spatial resolution of 256×256 pixels. The intensity values of each slice were normalized to the range $[0, 1]$ to ensure standardized input. To enhance the generalization capability of the model and reduce the risk of overfitting, standard data augmentation techniques were applied to expand the effective size of the training set. These included random cropping of 224×224 patches, random rotations within the range of -10° to 10° , and random horizontal and vertical flipping.

During inference, segmentation predictions were generated in a slice-by-slice manner. The resulting prediction outputs were then stacked to reconstruct the complete 3D volume. To ensure the fairness of experimental comparisons, no post-processing techniques were applied.

3.3 Experimental environment and training settings

The experimental environment was configured as follows: an Intel Core i9-14900KF @ 3.20GHz CPU, Ubuntu 22.04 LTS operating system, a single NVIDIA GeForce RTX 4090 GPU with CUDA 11.7, Python 3.10, and PyTorch 1.7 as the deep learning framework.

Considering the imbalance of various anatomical structure categories in cardiac MRI segmentation, a combined loss function comprising cross-entropy loss and Dice loss was employed to suppress the impact of sample imbalance on model performance. The segmentation network was based on the DeepLabV3+ architecture, with ResNetV1c-50 serving as the backbone network. To accelerate the training speed of the model and enhance its generalization and training stability, the DeepLabV3+ backbone network was initialized with ResNet50 weights pre-trained on the ImageNet dataset, while the parameters of the atrous spatial pyramid pooling (ASPP) decoder and auxiliary head were randomly initialized. Model parameters were optimized using stochastic gradient descent (SGD) with a weight decay of $5e-4$ and a momentum of 0.9.

The batch size was set to 8, comprising an equal number of labeled and unlabeled samples (4 each). The training process was conducted over 100 epochs with an initial learning rate of 0.002, which was scheduled to decay according to a polynomial learning rate decay policy with a power of 0.9.

During semi-supervised training, two annotation ratios were examined: 5% labeled data with 95% unlabeled data, and 10% labeled data with 90% unlabeled data. The pseudo-label weight α was set to 0.5, while the weights for the prototype-guided consistency loss and adversarial loss, denoted as β and γ , were set to 1.0 and 0.1, respectively. The initial confidence threshold τ_0 was set to 0.7 and adjusted dynamically using a decay strategy. Class prototypes were updated using an EMA scheme with a momentum parameter μ of 0.95.

3.4 System evaluation method

Two widely used metrics were employed to quantitatively evaluate the segmentation performance: DSC and Intersection over Union (IoU). DSC measures the overlap between the predicted segmentation and the ground truth annotation. It is defined as:

$$DSC = \frac{2|P \cap G|}{|P| + |G|} \quad (23)$$

where, P and G denote the predicted and ground truth segmentation regions, respectively. DSC ranges from 0 to 1, with higher values indicating greater segmentation accuracy.

IoU assesses the similarity between the predicted segmentation and the ground truth annotation. It is defined as:

$$IOU = \frac{|P \cap G|}{|P \cup G|} \quad (24)$$

Similar to DSC, IoU values range from 0 to 1, with higher scores reflecting better segmentation quality.

3.5 Comparative analysis

To validate the effectiveness of the proposed semi-supervised medical image segmentation strategy based on auxiliary localization and dual-level consistency, comparative experiments were conducted against several state-of-the-art semi-supervised segmentation methods on the polyp dataset. As shown in Table 1, the methods included DCT [32], Entropy Minimization (EM) [33], MT [34], Uncertainty-Aware Mean Teacher (UAMT) [35], and Cross Pseudo Supervision (CPS) [36]. All models were trained and evaluated under identical experimental configurations to ensure fair and reliable comparison.

Under extremely low annotation conditions (5%, corresponding to only 3 patients and 30 images), the proposed method achieved the best performance in segmenting the left ventricle, right ventricle, and myocardium. As presented in Table 2, the DSC for the left ventricle reached 0.851, approaching the fully supervised benchmark of 0.946 and substantially outperforming mainstream methods such as DCT (0.801) and MT (0.787). For the right ventricle, a DSC of 0.663 was achieved, which, although slightly lower than that of DCT (0.679), was accompanied by an IoU of 0.496—surpassing DCT's 0.514—indicating that the model not only maintained high segmentation accuracy of main structures but also exhibited stronger region overlap consistency. For the

myocardium, a DSC of 0.676 was obtained, significantly outperforming DCT (0.653) and MT (0.621), demonstrating superior discrimination in segmenting complex and boundary-ambiguous structures. Overall, the proposed method attained an average DSC of 0.730, which exceeded DCT (0.711) and MT (0.654), and markedly outperformed EM and CPS. These

results confirm the effectiveness and advantage of the proposed tri-fold mechanism—uncertainty-guided filtering, prototype-driven consistency, and adversarial discrimination—under extremely limited annotation scenarios.

Table 1. Comparison of different semi-supervised methods and their core principles

Method	Core Principle
DCT	Enforces inter-model consistency and employs adversarial learning to enhance sample diversity
EM	Combines pixel-level entropy with adversarial loss for semi-supervised training
MT	Trains using consistency between “student” and “teacher” model outputs
UAMT	Enhances MT with uncertainty estimation for optimized update
CPS	Models use each other’s pseudo-labels for learning
Proposed method	Integrates uncertainty quantification, prototype learning, and adversarial training

Table 2. Quantitative comparison of semi-supervised methods under 5% labeled data condition

Method	Left ventricle		Right ventricle		Myocardium		Avg.	
	DSC	IOU	DSC	IOU	DSC	IOU	DSC	IOU
DCT	0.801	0.668	0.679	0.514	0.653	0.485	0.711	0.556
EM	0.728	0.572	0.353	0.215	0.483	0.318	0.521	0.368
MT	0.787	0.649	0.554	0.383	0.621	0.450	0.654	0.494
UAMT	0.784	0.644	0.562	0.391	0.619	0.448	0.655	0.492
CPS	0.611	0.440	0.590	0.418	0.605	0.434	0.602	0.431
Proposed method	0.851	0.741	0.663	0.496	0.676	0.511	0.730	0.582
Fully supervised	0.946	0.897	0.866	0.764	0.861	0.755	0.891	0.805

Table 3. Quantitative comparison of semi-supervised methods under 10% labeled data condition

Method	Left ventricle		Right ventricle		Myocardium		Avg.	
	DSC	IOU	DSC	IOU	DSC	IOU	DSC	IOU
DCT	0.875	0.778	0.760	0.613	0.698	0.536	0.778	0.642
EM	0.857	0.750	0.672	0.506	0.631	0.461	0.720	0.572
MT	0.872	0.772	0.690	0.527	0.667	0.500	0.743	0.600
UAMT	0.871	0.771	0.684	0.520	0.666	0.499	0.740	0.597
CPS	0.746	0.594	0.671	0.505	0.646	0.477	0.688	0.523
Proposed method	0.908	0.832	0.785	0.646	0.710	0.550	0.801	0.676
Fully supervised	0.946	0.897	0.866	0.764	0.861	0.755	0.891	0.805

When the proportion of labeled data was increased to 10% (7 patients, 70 images), performance improvements were observed across all methods, as detailed in Table 3. Nevertheless, the proposed approach consistently achieved superior performance across all metrics. DSC for the left ventricle improved to 0.908, while the right ventricle and myocardium attained DSC values of 0.785 and 0.710, respectively. The average DSC reached 0.801, representing a marked improvement over DCT (0.778) and MT (0.743). Similar advantages were also observed in IoU scores, reflecting enhanced structural awareness and boundary localization capabilities. In comparison with fully supervised segmentation, the proposed method yielded segmentation accuracy that was already highly comparable, despite the limited annotation. This outcome underscores the effectiveness of the proposed framework in leveraging a small quantity of labeled data in conjunction with high-quality pseudo-labels for collaborative training.

Further comparative analysis revealed that the proposed method consistently achieved high segmentation accuracy across all three cardiac anatomical structures—namely, the left ventricle, right ventricle, and myocardium—under both 5% and 10% annotation settings. These structures differ substantially in segmentation difficulty: the left ventricle is morphologically distinct and relatively easier to segment; the

right ventricle exhibits significant anatomical variability and irregular boundaries; and the myocardium presents with narrow, complex geometry. The ability to maintain robust performance across all categories demonstrates the strong generalization capability and robustness of the proposed framework. In general, as the proportion of labeled data increased, not only was a notable improvement in segmentation performance observed, but the performance gain achieved by the proposed method exceeded that of competing approaches. These findings strongly support the unique advantage of the proposed multi-constraint collaborative self-training framework in enhancing medical image segmentation and promoting semi-supervised small sample learning.

Figure 2 presents a visual comparison of segmentation results produced by different semi-supervised methods under the 10% annotation setting (7 patients, 70 images), including DCT, EM, MT, UAMT, CPS, and the proposed method. Segmentation results for two representative cases (Case 1 and Case 2) are shown. Notable differences in segmentation quality across methods and anatomical regions can be clearly observed. In terms of structural completeness and boundary delineation, the proposed method demonstrated superior fidelity in reconstructing the morphology of key anatomical regions, including the left ventricle (blue), right ventricle (red), and myocardium (green). The predicted segmentations were

highly consistent with the ground truth, exhibiting smooth and well-closed contours. In contrast, methods such as DCT and CPS frequently suffered from under-segmentation and discontinuities, particularly in the right ventricle region, as evident in Case 1 where red segments were fragmented or incomplete. Similarly, methods such as MT, EM, and UAMT exhibited blurred or fused boundaries in the myocardium region, impairing their ability to distinguish adjacent tissues. With respect to fine-grained and small-volume anatomical structures, the proposed method demonstrated particularly precise segmentation performance.

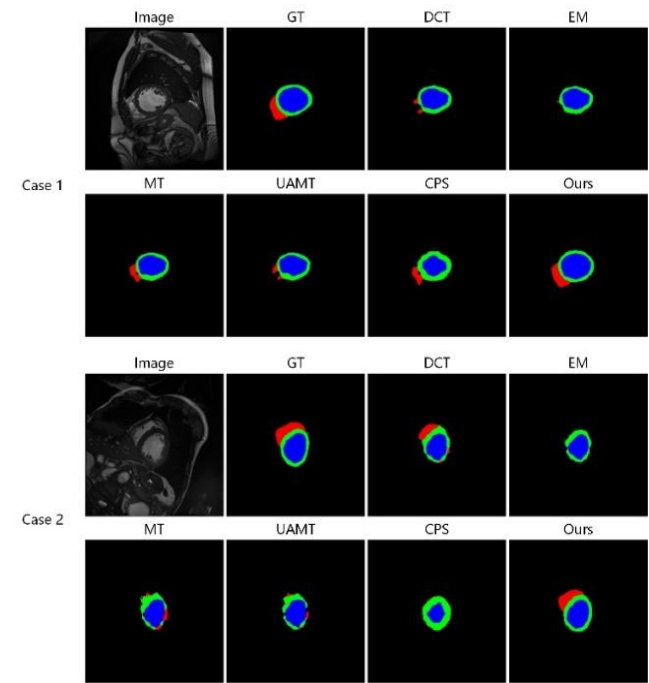


Figure 2. Visual comparison of segmentation results from different semi-supervised methods

For example, in Case 2, the myocardium region (green) was incompletely segmented or underestimated in most competing

methods. In contrast, the proposed method succeeded in accurately reconstructing the myocardial contours and comprehensively covering the target region. Furthermore, the right ventricle boundaries segmented by the proposed method were continuous and intact, substantially outperforming the fragmented or artifact-laden outputs produced by other methods. In terms of cross-case generalization, the proposed method maintained consistently high segmentation quality across both Case 1 and Case 2, demonstrating superior adaptability and robustness to variations in anatomical morphology and structure. Conversely, competing methods showed significant variability in performance across cases—some achieving moderate success in one instance but failing under more complex structural conditions—indicating inferior generalization ability.

Overall, the proposed multi-constraint collaborative self-training method substantially improved the segmentation accuracy and boundary detail capture for all anatomical structures of the heart under limited annotation conditions. The visual results provided strong evidence that the proposed approach not only achieved high overall segmentation precision but also exhibited superior capability in recovering fine-grained structures and complex regions. These findings highlight the practical effectiveness and significant potential for clinical application of the proposed framework.

3.6 Ablation study

To confirm the contribution of each core module to the overall segmentation performance, a series of ablation experiments was conducted under a 10% annotation ratio (corresponding to 7 patients and 70 annotated images). Four configurations were designed for comparison: (a) a conventional self-training strategy without uncertainty quantification, prototype learning, or adversarial training; (b) incorporation of uncertainty quantification alone; (c) integration of prototype learning in addition to uncertainty quantification; and (d) the full model with all three modules included. The quantitative results of the ablation study are summarized in Table 4.

Table 4. Quantitative results of the ablation study on module effectiveness

Uncertainty Quantification	Prototype Learning	Adversarial Training	Left Ventricle		Right Ventricle		Myocardium		Avg.	
			DSC	IOU	DSC	IOU	DSC	IOU	DSC	IOU
-	-	-	0.78	0.64	0.50	0.33	0.21	0.12	0.49	0.36
√	-	-	0.835	0.717	0.695	0.533	0.657	0.489	0.729	0.580
√	√	-	0.893	0.807	0.757	0.609	0.683	0.519	0.778	0.645
√	√	√	0.908	0.832	0.785	0.646	0.710	0.550	0.801	0.676

The conventional self-training strategy yielded the weakest performance. DSCs of only 0.780, 0.500, and 0.210 were observed for the left ventricle, right ventricle, and myocardium, respectively, with a mean DSC of 0.490 and a mean IoU of merely 0.360. These results indicate that, in the absence of any quality control or consistency constraints, noisy pseudo-labels severely compromised model performance, especially in structures such as the right ventricle and myocardium where segmentation results were very limited. Significant improvements were achieved upon the integration of uncertainty quantification. The mean DSC increased to 0.729, and the mean IoU reached 0.580. Specifically, DSC values of 0.835, 0.695, and 0.657 were recorded for the left ventricle, right ventricle, and

myocardium, respectively. These enhancements suggest that the use of dynamic confidence and entropy filtering effectively suppressed low-quality pseudo-labels, thereby improving the reliability of the pseudo-supervision signals and significantly boosting segmentation performance across all structures. Further incorporation of the prototype learning module led to additional gains, with the mean DSC rising to 0.778 and the mean IoU to 0.645. Notably, DSC values for the right ventricle and myocardium increased to 0.757 and 0.683, respectively. These results affirm that prototype learning, through enforcing intra- and inter-class consistency constraints within the feature space, enhanced the discriminative power of features and improved the model’s learning ability for complex morphologies and boundary details. The full configuration, combining uncertainty quantification, prototype learning, and

adversarial training, produced the highest performance across all metrics. The mean DSC reached 0.801 and the mean IoU reached 0.676. The DSC values for the left ventricle, right ventricle, and myocardium improved to 0.908, 0.785, and 0.710, respectively. The inclusion of adversarial training further refined boundary delineation and structural coherence, enabling the model to align closely with ground truth annotations not only at a global level but also in local details.

In summary, the ablation study demonstrates that each individual component contributes substantially to performance enhancement. Uncertainty quantification was found to be foundational in improving pseudo-label quality; prototype learning strengthened feature discriminability; and adversarial training further optimized both global and fine-grained segmentation results. The synergistic integration of these modules enabled the proposed framework to surpass conventional approaches and achieve superior segmentation performance and generalization under limited annotations.

4. CONCLUSION

To address the critical challenge of severely limited annotated data in cardiac MRI segmentation, a multi-constraint collaborative self-training framework was proposed, integrating uncertainty quantification, prototype learning, and adversarial training. Through the synergistic combination of dynamic pseudo-label filtering, prototype-based consistency constraints, and multi-granularity adversarial discrimination, the proposed framework substantially enhanced segmentation performance and generalization under small sample scenarios. Experimental results on the publicly available ACDC dataset demonstrated that, under annotation ratios of 5% and 10%, mean DSCs of 0.730 and 0.801, respectively, were achieved for the left ventricle, right ventricle, and myocardium. These values significantly exceeded those obtained by state-of-the-art semi-supervised methods such as DCT (0.778), EM, MT (0.743), UAMT, and CPS, and approached the performance of fully supervised models (0.891). Ablation studies further confirmed that each module—uncertainty-based selection, prototype learning, and adversarial discrimination—contributed substantially to performance enhancement and that their integration yielded optimal improvements in both segmentation accuracy and robustness. Visual assessments additionally revealed clear advantages in structural completeness and boundary fidelity. Future research will focus on extending the generalizability of the proposed framework to multi-center and multi-modality medical image segmentation tasks. Strategies such as adaptive learning and domain adaptation will be explored to improve adaptation across varying imaging devices, populations, and pathological conditions. Moreover, efforts will be made to apply the proposed multi-constraint self-training framework to additional clinical tasks, including multi-organ segmentation and lesion detection, with the goal of advancing intelligent and efficient computer-aided diagnosis in medical imaging.

REFERENCES

- [1] Roth, G.A., Mensah, G.A., Johnson, C.O., Addolorato, G., et al. (2020). Global burden of cardiovascular diseases and risk factors, 1990–2019: Update from the GBD 2019 study. *Journal of the American College of Cardiology*, 76(25): 2982-3021.
- [2] Petersen, S.E., Aung, N., Sanghvi, M.M., Zemrak, F., et al. (2016). Reference ranges for cardiac structure and function using cardiovascular Magnetic resonance (CMR) in Caucasians from the UK Biobank population cohort. *Journal of Cardiovascular Magnetic Resonance*, 19(1): 18. <https://doi.org/10.1186/s12968-017-0327-9>
- [3] Bai, W., Sinclair, M., Tarroni, G., Oktay, O., et al. (2018). Automated cardiovascular Magnetic resonance image analysis with fully convolutional networks. *Journal of Cardiovascular Magnetic Resonance*, 20(1): 65. <https://doi.org/10.1186/s12968-018-0471-x>
- [4] Chen, C., Qin, C., Qiu, H., Tarroni, G., Duan, J., Bai, W., Rueckert, D. (2020). Deep learning for cardiac image segmentation: A review. *Frontiers in Cardiovascular Medicine*, 7: 25. <https://doi.org/10.3389/fcvm.2020.00025>
- [5] Zhuang, X., Li, L., Payer, C., Štern, D., et al. (2019). Evaluation of algorithms for multi-modality whole heart segmentation: an open-access grand challenge. *Medical Image Analysis*, 58: 101537. <https://doi.org/10.1016/j.media.2019.101537>
- [6] Fritscher, K.D., Pilgram, R., Schubert, R. (2005). Automatic cardiac 4D segmentation using level sets. In *International Workshop on Functional Imaging and Modeling of the Heart*, Barcelona, Spain, pp. 113-122. https://doi.org/10.1007/11494621_12
- [7] Lorenzo-Valdés, M., Sanchez-Ortiz, G.I., Elkington, A.G., Mohiaddin, R.H., Rueckert, D. (2004). Segmentation of 4D cardiac MR images using a probabilistic atlas and the EM algorithm. *Medical Image Analysis*, 8(3): 255-265. <https://doi.org/10.1016/j.media.2004.06.005>
- [8] Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J. (1995). Active shape models-their training and application. *Computer Vision and Image Understanding*, 61(1): 38-59. <https://doi.org/10.1006/cviu.1995.1004>
- [9] Iglesias, J.E., Sabuncu, M.R. (2015). Multi-atlas segmentation of biomedical images: A survey. *Medical Image Analysis*, 24(1): 205-219. <https://doi.org/10.1016/j.media.2015.06.012>
- [10] Peng, P., Lekadir, K., Gooya, A., Shao, L., Petersen, S.E., Frangi, A.F. (2016). A review of heart chamber segmentation for structural and functional analysis using cardiac Magnetic resonance imaging. *Magnetic Resonance Materials in Physics, Biology and Medicine*, 29: 155-195. <https://doi.org/10.1007/s10334-015-0521-4>
- [11] Shamshad, F., Khan, S., Zamir, S.W., Khan, M.H., Hayat, M., Khan, F.S., Fu, H. (2023). Transformers in medical imaging: A survey. *Medical Image Analysis*, 88: 102802. <https://doi.org/10.1016/j.media.2023.102802>
- [12] Jiang, Y., Gong, L.J., Huang, H., Qi, M.M. (2025). MSF-TransUNet: A multi-scale feature fusion transformer-based U-Net for medical image segmentation with uniform attention. *Traitement du Signal*, 42(1): 531-540. <https://doi.org/10.18280/ts.420145>
- [13] Mortazi, A., Karim, R., Rhode, K., Burt, J., Bagci, U. (2017). CardiacNET: Segmentation of left atrium and proximal pulmonary veins from MRI using multi-view CNN. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2017: 20th International Conference*, Quebec City, QC, Canada, pp. 377-385. https://doi.org/10.1007/978-3-319-66185-8_43

- [14] Long, J., Shelhamer, E., Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, pp. 3431-3440. <https://doi.org/10.1109/CVPR.2015.7298965>
- [15] Ronneberger, O., Fischer, P., Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, pp. 234-241. https://doi.org/10.1007/978-3-319-24574-4_28
- [16] Ma, N., Jin, S.W. (2025). CCUNet: UNet based on an improved coordinate channel attention mechanism and its applications. *Traitement du Signal*, 42(1): 119-128. <https://doi.org/10.18280/ts.420111>
- [17] Jha, D., Smedsrud, P.H., Riegler, M.A., Johansen, D., De Lange, T., Halvorsen, P., Johansen, H.D. (2019). Resunet++: An advanced architecture for medical image segmentation. In 2019 IEEE International Symposium on Multimedia (ISM), San Diego, CA, USA, pp. 225-2255. <https://doi.org/10.1109/ISM46123.2019.00049>
- [18] Chen, J., Lu, Y., Yu, Q., Luo, X., et al. (2021). Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*. <https://doi.org/10.48550/arXiv.2102.04306>
- [19] Chang, Y., Jung, C. (2020). Automatic cardiac MRI segmentation and permutation-invariant pathology classification using deep neural networks and point clouds. *Neurocomputing*, 418: 270-279. <https://doi.org/10.1016/j.neucom.2020.08.030>
- [20] Campello, V.M., Gkontra, P., Izquierdo, C., Martin-Isla, C., et al. (2021). Multi-centre, multi-vendor and multi-disease cardiac segmentation: The M&Ms challenge. *IEEE Transactions on Medical Imaging*, 40(12): 3543-3554. <https://doi.org/10.1109/TMI.2021.3090082>
- [21] Alsemmeiri, R.A., Iqbal, Z., Bakhsh, S.T., Alghamdi, B.M., Alsulami, A.A., Alturki, B. (2025). Enhancing breast cancer diagnosis: A semi-supervised deep learning model for automated MRI tumor segmentation. *Traitement du Signal*, 42(2): 737-750. <https://doi.org/10.18280/ts.420212>
- [22] Ayoob, M., Nettasinghe, O., Sylvester, V., Bowala, H. Mohideen, H. (2025). Peering into the Heart: A Comprehensive Exploration of Semantic Segmentation and Explainable AI on the MnMs-2 Cardiac MRI Dataset. *Applied Computer Systems*, 30(1): 12-20. <https://doi.org/10.2478/acss-2025-0002>
- [23] Bai, W., Chen, C., Tarroni, G., Duan, J., et al. (2019). Self-supervised learning for cardiac mr image segmentation by anatomical position prediction. In Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, pp. 541-549. https://doi.org/10.1007/978-3-030-32245-8_60
- [24] Wang, K., Zhan, B., Zu, C., Wu, X., Zhou, J., Zhou, L., Wang, Y. (2022). Semi-supervised medical image segmentation via a tripled-uncertainty guided mean teacher model with contrastive learning. *Medical Image Analysis*, 79: 102447. <https://doi.org/10.1016/j.media.2022.102447>
- [25] Huang, Z., Niu, G., Liu, X., Ding, W., Xiao, X., Wu, H., Peng, X. (2021). Learning with noisy correspondence for cross-modal matching. *Advances in Neural Information Processing Systems*, 34: 29406-29419.
- [26] Zhang, P., Zhang, B., Zhang, T., Chen, D., Wang, Y., Wen, F. (2021). Prototypical pseudo label denoising and target structure learning for domain adaptive semantic segmentation. In 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, pp. 12409-12419. <https://doi.org/10.1109/CVPR46437.2021.01223>
- [27] Xie, Q., Luong, M.T., Hovy, E., Le, Q.V. (2020). Self-training with noisy student improves imagenet classification. In 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, pp. 10684-10695. <https://doi.org/10.1109/CVPR42600.2020.01070>
- [28] Yang, L., Zhuo, W., Qi, L., Shi, Y., Gao, Y. (2022). St++: Make self-training work better for semi-supervised semantic segmentation. In 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, pp. 4258-4267. <https://doi.org/10.1109/CVPR52688.2022.00423>
- [29] Hu, H., Wei, F., Hu, H., Ye, Q., Cui, J., Wang, L. (2021). Semi-supervised semantic segmentation via adaptive equalization learning. *Advances in Neural Information Processing Systems*, 34: 22106-22118.
- [30] Xie, Q., Luong, M.T., Hovy, E., Le, Q.V. (2020). Self-training with noisy student improves imagenet classification. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, USA, pp. 10687-10698. <https://doi.org/10.1109/CVPR42600.2020.01070>
- [31] Bernard, O., Lalande, A., Zotti, C., Cervenansky, F. et al. (2018). 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. <https://doi.org/10.1109/TMI.2018.2837502>
- [32] Peng, J., Estrada, G., Pedersoli, M., Desrosiers, C. (2020). Deep co-training for semi-supervised image segmentation. *Pattern Recognition*, 107: 107269. <https://doi.org/10.1016/j.patcog.2020.107269>
- [33] Vu, T.H., Jain, H., Bucher, M., Cord, M., Pérez, P. (2019). Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation. In 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, pp. 2512-2521. <https://doi.org/10.1109/CVPR.2019.00262>
- [34] Tarvainen, A., Valpola, H. (2017). Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, California, USA, pp. 1195-1204.
- [35] Meng, H., Zhao, H., Yu, Z., Li, Q., Niu, J. (2022). Uncertainty-aware mean teacher framework with inception and squeeze-and-excitation block for MICCAI FLARE22 challenge. In MICCAI Challenge on Fast and Low-Resource Semi-supervised Abdominal Organ Segmentation, Singapore, pp. 245-259. https://doi.org/10.1007/978-3-031-23911-3_22
- [36] Chen, X., Yuan, Y., Zeng, G., Wang, J. (2021). Semi-supervised semantic segmentation with cross pseudo supervision. In 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, pp. 2613-2622. <https://doi.org/10.1109/CVPR46437.2021.00264>