Neurocomputing 659 (2026) 131846

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Multi-organ medical image segmentation via adaptive disentangled domain generalization collaborative learning

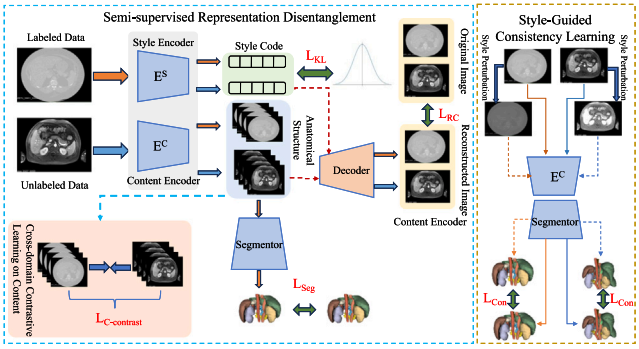
Min Dong [a], Yishuang Liu [a], Ating Yang [a], Ye Zhang [a], Rongchang Zhao [b], Ling Liu [c,*]

[a] School of Electrical and Information Engineering, Zhengzhou University, Zhengzhou, 450001, China
[b] School of Computer Science and Engineering, Central South University, Changsha, 410083, China
[c] School of Stomatology of Zhengzhou University and The First Affiliated Hospital of Zhengzhou University, Zhengzhou, 450052, China

HIGHLIGHTS

- Proposed an adaptive disentangled framework for semi-supervised 3D medical image segmentation.
- Developed an adaptive loss to enhance accuracy for small, clinically critical organs.
- Achieved superior performance on the AMOS dataset with interpretable visualizations.

GRAPHICAL ABSTRACT



ARTICLE INFO

Communicated by V. Conti

Keywords:
Domain generalization
Domain discrepancy
Representation disentanglement
Semi-supervised learning
3D medical image segmentation

ABSTRACT

Semi-supervised 3D medical image segmentation holds great promise for clinical practice but remains challenged by domain shifts caused by imaging protocol variations and by the difficulty of accurately segmenting small organs with sparse annotations. To address these issues, we propose a novel Adaptive Disentangled Domain Generalization Collaborative Learning (AD-DGCL) framework. It integrates two core components: a Semi-Supervised Representation Disentanglement (SSRD) module that separates domain-specific style and anatomical content via dual encoders and cross-domain contrastive learning, and a Style-induced Consistency Training (SCT) module that enhances robustness through synthetic style perturbations and consistency regularization. In addition, an adaptive region-specific loss function is designed to prioritize small organs by dynamically adjusting weights according to pixel frequency. Extensive experiments demonstrate that AD-DGCL substantially improves segmentation accuracy, particularly for small organs, while effectively mitigating domain gaps, thus showing strong potential for robust and generalizable clinical deployment.

1. Introduction

In recent years, semi-supervised 3D medical image segmentation has shown remarkable performance and garnered growing interest. These approaches typically assume that the labeled and unlabeled datasets follow an identical distribution. However, in multi-center clinical practice, supplementary data are frequently obtained from external centers,

* Corresponding author.
Email addresses: iemdong@zzu.edu.cn (M. Dong), yishliu@gs.zzu.edu.cn (Y. Liu), yangating@163.com (A. Yang), ye_zhang@gs.zzu.edu.cn (Y. Zhang), zhaorc@csu.edu.cn (R. Zhao), fccliul21@zzu.edu.cn (L. Liu).

https://doi.org/10.1016/j.neucom.2025.131846
Received 3 September 2025; Received in revised form 4 October 2025; Accepted 15 October 2025
Available online 16 October 2025

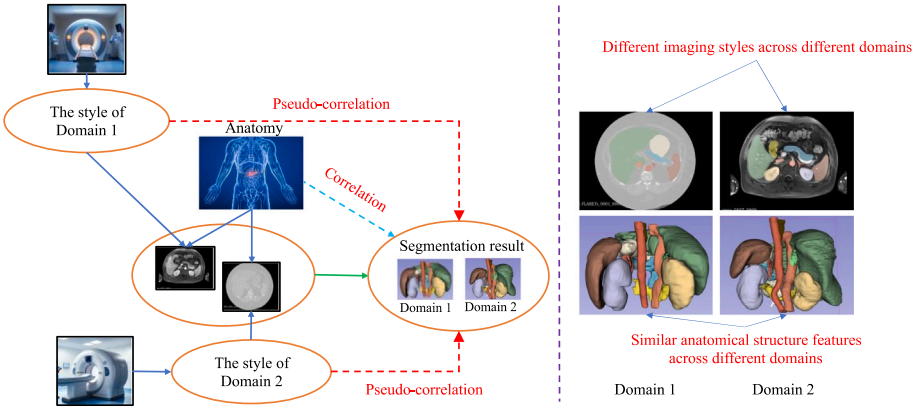
which introduce domain discrepancies between the labeled and unlabeled datasets [1]. Furthermore, in multi-organ segmentation, particularly for small organs, the small size, limited voxel representation, and intricate morphology result in severe class imbalance, thereby posing a substantial challenge to accurate delineation.

Existing research on mitigating domain discrepancies in semi-supervised models has mainly concentrated on aligning the distributions between labeled and unlabeled images. Approaches such as data augmentation (e.g., the copy–paste technique in BCP [2]) often prove ineffective for complex data, as they may distort the original image structure and introduce artificial semantic features. Additionally, methods based on Generative Adversarial Networks (GANs) are capable of generating images that resemble a target domain, but they often lack reliable evaluation metrics to quantify the effectiveness of this alignment. These methods frequently fail to address the root causes of domain shift and may result in substantial knowledge loss.

The accurate segmentation of small organs continues to pose a critical challenge. These organs, including the pancreas, gallbladder, and adrenal glands, occupy fewer than 0.1 % of the total volume [3]. This extreme class imbalance often causes models to overlook these minority classes, biasing learning toward background or larger organs. Moreover, their boundaries are often ambiguous with low contrast and exhibit substantial anatomical variation. Given their indispensable clinical value, enhancing segmentation accuracy is vital for precision medicine.

Furthermore, varying imaging styles tend to form spurious correlations with segmentation predictions, preventing models from capturing genuine anatomical features. Disentangled representation learning, which separates generative factors into independent features, provides a natural solution to this problem [4]. As shown in Fig. 1, despite substantial variations in imaging style, the underlying anatomical structure remains largely consistent. By disentangling style and content, models can focus on content features intrinsically relevant to segmentation while mitigating the influence of confounding style factors. However, existing disentangled representation learning methods [5,6] often fail to achieve effective decoupling under substantial domain discrepancies.

To address these challenges, we propose a novel Adaptive Disentangled Domain Generalization Collaborative Learning (AD-DGCL) framework. The framework enhances generalization and segmentation accuracy in 3D medical image segmentation by leveraging labeled and unlabeled data across domains. As shown in Fig. 2, AD-DGCL first disentangles content and style representations via a Semi-Supervised Representation Decoupling (SSRD) module, enabling the model to capture unified anatomical features. Additionally, cross-domain content contrastive learning is incorporated to strengthen domain-invariant representations. Second, Style-induced Consistency Training (SCT) applies style perturbations to improve robustness and generalization. Crucially, to address small-organ segmentation, we design an adaptive region-specific loss that dynamically adjusts weights according to organ voxel counts.



(a) Mechanism diagram of medical image imaging and segmentation prediction generation.

(b) Schematic diagrams of medical images and anatomical structures under different domains.

Fig. 1. (a) There is a spurious correlation between the style in the input image and the segmentation result. (b) The same organ from different domains has similar anatomical structure features despite the differences in imaging styles.

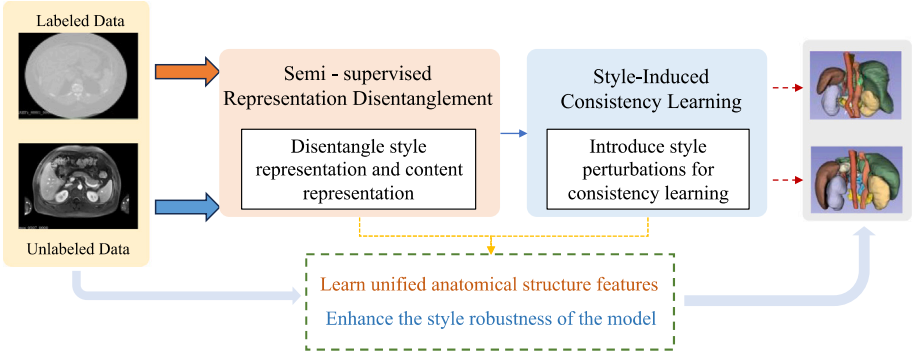


Fig. 2. The proposed AD-DGCL disentangles style and content representations to capture domain-invariant anatomical structures and applies style perturbations for consistency training, thereby enhancing robustness against inter-domain variations.

Overall, our proposed framework has three practical contributions:

- A novel adaptive disentangled domain generalization co-learning framework for semi-supervised 3D medical image segmentation that systematically addresses domain discrepancies via an SSRD module and enhances robustness with an SCT module.
- An innovative adaptive region-specific loss function that dynamically assigns higher weights to rare, small organ categories based on their pixel frequency, effectively mitigating class imbalance and improving the segmentation accuracy of small organs.
- Extensive experiments with superior performance on the AMOS abdominal multi-organ segmentation dataset, validating that the proposed AD-DGCL framework achieves state-of-the-art results in handling domain discrepancies and precisely segmenting small organs.

2. Related work

2.1. Semi-supervised 3D medical image segmentation

Early medical image segmentation relied on traditional feature-based methods such as thresholding, region-based, and clustering approaches [7–10], but their accuracy was often inadequate for modern clinical requirements. Deep learning methods, including fully convolutional networks (FCNs) [11] and U-Net [12], achieved significant progress, yet remain constrained by the cost and difficulty of acquiring large-scale labeled data. Semi-supervised learning has emerged as an effective solution, leveraging limited labeled data alongside abundant unlabeled data to enhance generalization and accuracy. Major approaches include:

- Pseudo-labeling-based methods: These generate pseudo-labels for unlabeled data by leveraging high-confidence predictions from models trained on limited labeled datasets [13]. However, the scarcity of initial labeled data often results in noisy and erroneous pseudo-labels, adversely affecting performance. Although various strategies have been proposed to refine these labels [14–17], they remain vulnerable to noise, particularly in complex regions. Furthermore, pronounced domain differences and the inherent imprecision of pseudo-labels for small organs further exacerbate these challenges [18].
- Consistency Regularization-based methods: These methods assume that model outputs remain consistent under minor perturbations. They leverage unlabeled data by enforcing prediction consistency under perturbations, including data perturbations [19], feature perturbations [20], or model perturbations [21,22]. A key challenge lies in balancing perturbation intensity: weak perturbations fail to fully exploit unlabeled data, whereas strong perturbations may distort image semantics [23]. Crucially, these methods often cannot distinguish content from style variations under domain differences, limiting generalization and potentially leading to the learning of irrelevant features.

2.2. Domain generalization and disentangled representation learning

Recent advances in semi-supervised 3D medical image segmentation have achieved notable performance by leveraging both labeled and unlabeled data, typically under the assumption of a shared distribution. Unsupervised domain adaptation (UDA) further mitigates distribution gaps by transferring models across modalities [24], yet remains dependent on target-domain data.

In clinical practice, however, domain shifts arising from variations in medical imaging equipment and acquisition protocols severely degrade model generalization when target-domain data is unavailable. Domain Generalization (DG) addresses this challenge by learning domain-invariant representations from single or multiple source domains. Existing DG approaches primarily focus on minimizing inter-domain statistical divergence, with recent theoretical work providing a rigorous foundation for adversarial training based on Jensen–Shannon divergence [25]. Alternative formulations instead define domain discrepancy through the performance gap, where a model trained on one domain is expected to generalize effectively to another [26].

Beyond distribution alignment, disentangled representation learning has emerged as a promising strategy. It explicitly separates domain-invariant 'content' from domain-specific 'style', a principle also explored in other areas. For example, in Graph Domain Adaptation, Fang et al. highlight that node attribute shifts dominate over topological differences and proposes an attribute-driven alignment method that jointly aligns attribute-graphs and topology-graphs [27]. Another work addresses discrepancies in graph homophily by applying mixed filters to separately align homophilic and heterophilic signals [28]. In a different modality, EEG-based emotion recognition, Yang et al. proposed spectral–spatial attention alignment, which emphasizes abstract cognitive attention patterns rather than raw features, thereby achieving robust multi-source domain adaptation [29].

However, in 3D medical imaging, the high dimensionality and strong coupling between style (e.g., scanner artifacts) and content (e.g., anatomical structures) make effective disentanglement particularly challenging. Without supervision tailored to clinical semantics, general-purpose DG methods often result in suboptimal separation or even negative transfer, making the development of domain-generalized strategies for medical imaging essential. Therefore, we propose a specialized framework that explicitly leverages the style–content structure of medical images for robust multi-organ segmentation.

2.3. Imbalanced data handling

Class imbalance remains a pervasive challenge in medical image segmentation, especially for tasks involving small organs or lesions. As the voxel count in target regions is substantially lower than in the background, models tend to bias toward the majority class, resulting in poor segmentation of minority classes such as small organs. Moreover, ambiguous boundaries and pronounced anatomical variations in small organs complicate manual annotation, further aggravating data sparsity.

To address this challenge, researchers have proposed various methods. Resampling is a common strategy that balances the dataset by oversampling the minority class or undersampling the majority class [30]. However, this risks overfitting or the loss of crucial information. Loss function weighting assigns different weights to the loss terms for different classes, giving higher weights to minority classes [31]. While methods like Dice Loss and Focal Loss [32] are helpful, they often require manual tuning and are limited in extremely imbalanced scenarios. Furthermore, data augmentation creates more diverse samples for minority classes, but it may not fundamentally change the pixel ratio imbalance and can introduce unrealistic features.

Therefore, existing approaches remain limited in offering a universal and optimal solution for extremely imbalanced small organs. The adaptive region-specific loss function proposed in this study extends loss function weighting, uniquely adjusting weights dynamically according to the actual pixel count of each region, thereby more precisely addressing class imbalance in small organ segmentation.

3. Methodology

The proposed AD-DGCL framework, illustrated in Fig. 3, is designed to address the challenges of semi-supervised, cross-domain 3D medical image segmentation. Its architecture integrates three synergistic components. The **Style-Semantic Representation Disentanglement (SSRD)** module decomposes image representations into a domain-invariant content code and a domain-specific style code through a dual-encoder architecture and cross-domain contrastive learning (Section 3.1). The **Style-induced Consistency Training (SCT)** module exploits unlabeled data by enforcing prediction consistency between original images and their style-perturbed counterparts, thereby enhancing generalization capability (Section 3.2). The **Adaptive Region-Specific Loss** dynamically

adjusts voxel weights according to organ size to alleviate class imbalance and improve segmentation accuracy for small, clinically critical organs (Section 3.3). These modules are jointly optimized within a collaborative learning framework (Section 3.4).

3.1. Semi-supervised representation decoupling module

The SSRD module is a core component of AD-DGCL, responsible for decoupling style and learning domain-invariant content features. It explicitly decomposes medical image variations into task-relevant content and task-irrelevant style components (e.g., arising from different acquisition devices or protocols). As shown in Fig. 4, SSRD employs a dual-encoder architecture comprising a content encoder and a style encoder. Domain-specific normalization layers are incorporated to prevent style information from contaminating the content encoder, ensuring the extraction of pure content features. Additionally, SSRD applies cross-domain content contrastive learning to recognize similar anatomical structures, facilitating the model in acquiring unified, domain-agnostic anatomical representations.

3.1.1. Dual-encoder architecture and domain-specific batch normalization

The SSRD module in the AD-DGCL model employs a dual-encoder architecture to disentangle content and style. The style encoder utilizes a Variational Autoencoder (VAE) to generate a style code representing the image's intensity distribution. The VAE constructs a smooth, low-dimensional latent space for intensity distributions, ensuring the style code aligns with a prior isotropic Gaussian distribution g(z) = N(0, 1) [33]. During training, the style encoder samples the style code using a reparameterization trick, with optimization guided by the KL divergence between the unit Gaussian and the estimated Gaussian distribution, as shown in Eq. (1).

$$L_{KL} = \frac{1}{N_L + N_U} \sum_{i \in L \cup U} D_{kl}(q(z \mid \mu_i, \sigma_i) \| g(z)) \tag{1}$$

Where $D_{kl}(q(z \mid \mu_i, \sigma_i) \parallel g(z))$ represents the KL divergence between the estimated Gaussian distribution and the unit Gaussian distribution. In addition to the style encoder, the image is processed by the content

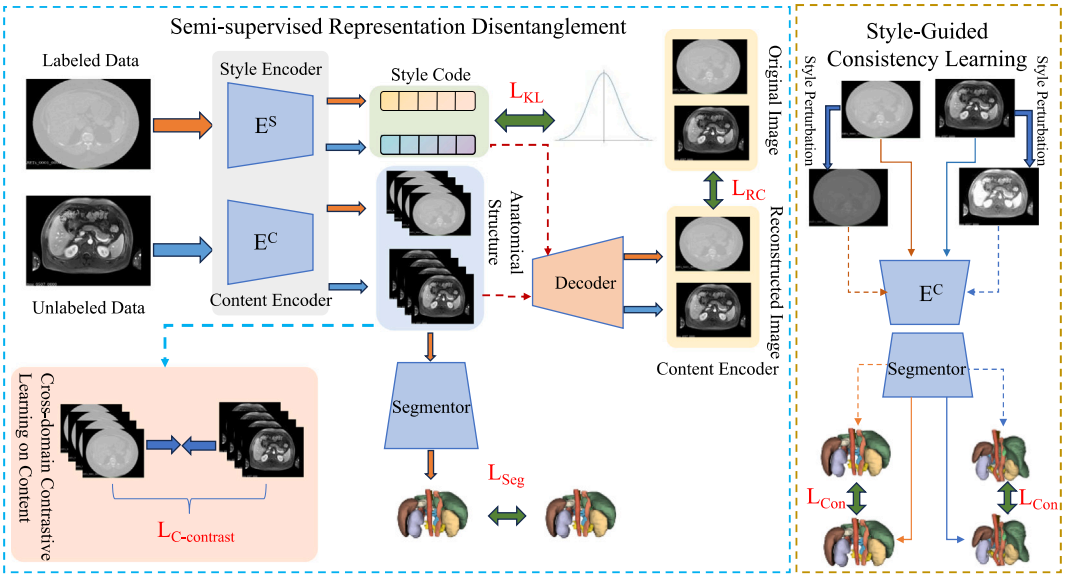


Fig. 3. The proposed AD-DGCL includes: (1) a disentangled dual-encoder to learn domain-invariant content via contrastive learning; (2) a style-induced consistency module to enhance robustness under appearance shifts; and (3) domain-aware losses to adaptively supervise multi-source data.

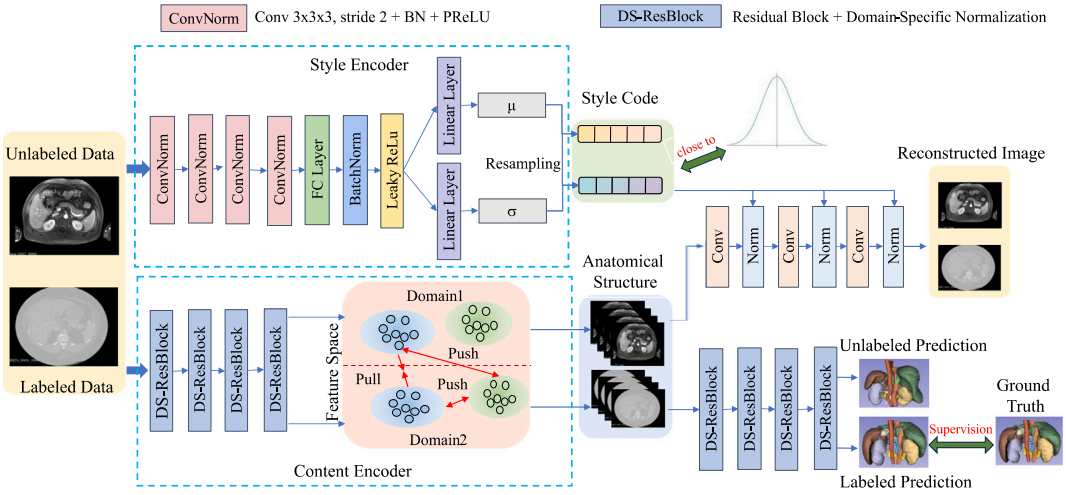


Fig. 4. The SSRD leverages a dual-encoder structure with domain-specific normalization to decouple content and style features, and employs cross-domain contrastive learning to align anatomical structures across domains.

encoder to capture anatomical structures. This encoder, based on V-Net, maps the image to a multi-channel spatial content representation. Subsequently, a decoder integrates the content and style representations to reconstruct the original image. To ensure reconstruction accuracy, the discrepancy between the reconstructed and original images is quantified using the mean squared error (MSE) loss, as shown in Eq. (2).

$$L_{RC} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{x}_i)^2 \tag{2}$$

Furthermore, to enhance the content encoder's independence from style and boost its style-agnostic capabilities, SSRD incorporates a Domain-Specific Batch Normalization layer. This layer can adaptively adjust to different image domains by learning each domain's statistical information separately. Specifically, it computes the mean and variance of feature maps from different domains and then normalizes the features based on these statistics. For a feature map x from source domain k, the DSBN operation is:

$$DSBN(x) = \gamma_k \frac{x - \mu_k}{\sqrt{{\sigma_k}^2 + \varepsilon}} + \beta_k \tag{3}$$

Where μ_k and ${\sigma_k}^2$ are the mean and variance of features from source domain k, γ_k and β_k are learnable scaling and translation parameters. By normalizing features based on each domain's statistics, DSBN ensures that features from images with different styles have similar distributions. This enables the content encoder to focus on the anatomical content of the image, reducing interference from style factors.

3.1.2. Cross-domain content contrastive learning

To make the content encoder more robust and effective, the SSRD module incorporates cross-domain content contrastive learning. This method reinforces domain-invariant information by constructing positive and negative sample pairs. The core idea is to maximize the similarity between feature representations of identical anatomical structures across different domains, while minimizing the similarity between different anatomical structures. This encourages the model to identify unified anatomical structures regardless of style and location.

Specifically, for an anchor image x_i (content features c_i), a positive sample x_{j+} (content features c_{j+}) is selected from a different domain, sharing the same anatomical structure. Concurrently, a negative sample x_{j-} (content features c_{j-}) is chosen, containing a different anatomical structure irrespective of its domain. The cross-domain content contrastive loss is then expressed by Eq. (4):

$$L_{\text{contrast}} = -\log \frac{\exp\left(\frac{\text{sim}(c_i, c_{j+})}{\tau}\right)}{\exp\left(\frac{\text{sim}(c_i, c_{j+})}{\tau}\right) + \sum_{\Lambda^-}\exp\left(\frac{\text{sim}(c_i, c_{j-})}{\tau}\right)} \tag{4}$$

c_i represents the content feature of the anchor image, c_{j+} is the content feature of the positive sample (sharing the same anatomical structure), and c_{j-} denotes the feature representations of negative samples with different anatomical structures across various domains. The term τ is the temperature parameter.

By minimizing this loss function, the model is compelled to learn a feature representation where semantically identical anatomical structures from different domains are clustered closely in the feature space, while distinct structures are pushed farther apart. This explicit contrastive constraint substantially enhances the domain-invariance of content features. Consequently, the model effectively resists interference from imaging style variations and accurately segments anatomical structures even when encountering data from unseen domains.

3.2. Style-induced consistency training

In multi-center medical image segmentation, to ensure that predictions are not affected by changes in imaging style, AD-DGCL introduces SCT. As shown in Fig. 5, SCT applies style perturbations to the input image. Based on the smoothness assumption, it enforces consistency between the model's output before and after these perturbations. This effectively improves segmentation stability across different image styles.

Specifically, SCT integrates style codes from the semi-supervised representation decoupling module into the input image to create style-reconstructed, perturbed images. For instance, the style code s_B from an MR image is combined with an input CT image x_A, while the style code s_A from a CT image is combined with an MR image x_B, as shown in Eqs. (5) and (6).

$$\hat{x}_A = x_A + s_B \tag{5}$$

$$\hat{x}_B = x_B + s_A \tag{6}$$

The style-reconstructed images are then passed through the content encoder and segmentor. The goal is for them to produce content and style representations that are consistent with the original images, as expressed in Eq. (7). This chapter measures the discrepancy between segmentation predictions before and after style perturbation using Mean Squared Error (MSE) loss:

$$L_{con} = \left\|p(\widehat{x_A}) - p(x_A)\right\|_2^2 + \left\|p(\widehat{x_B}) - p(x_B)\right\|_2^2 \tag{7}$$

$p(x_A)$ and $p(\widehat{x_A})$ are the segmentation predictions for the original CT image and the style-perturbed CT image, respectively. Similarly, $p(x_B)$ and $p(\widehat{x_B})$ correspond to the original and style-perturbed MR images. Through this consistent learning approach, SCT compels the model to learn content representations that remain consistent even after style reconstruction. This mechanism effectively leverages unlabeled data to significantly improve the model's robustness to style variations.

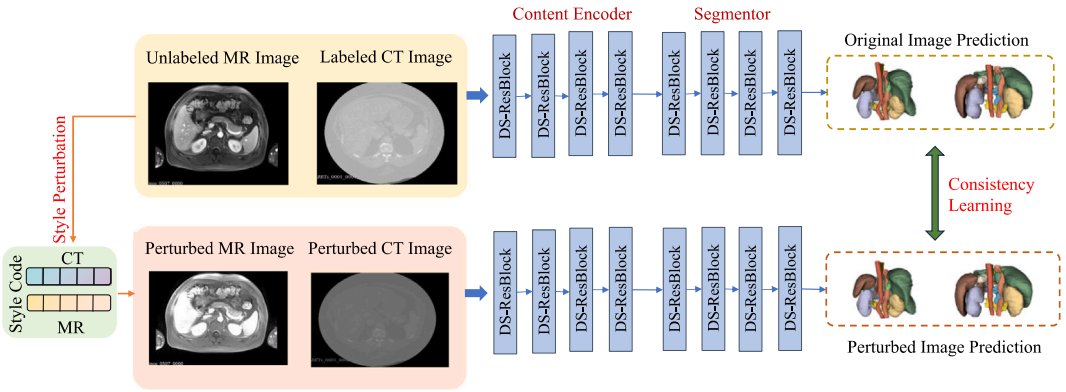


Fig. 5. The SCT enhances segmentation robustness by enforcing consistent outputs between original and style-perturbed inputs, based on the smoothness assumption.

3.3. Adaptive region-specific loss function

In medical image segmentation, the standard Dice loss has limitations. It fails to account for variations in segmentation difficulty across different regions and struggles with data imbalance because target organs are often much smaller than the background. Dice loss also weights false positives and false negatives equally, which is often clinically unacceptable. To address this, we introduce an Adaptive Region-Specific Loss. This function is designed to more precisely capture the unique features of each region, enabling a targeted optimization of each sub-region's segmentation performance.

To better address data imbalance, particularly when target organ volumes are significantly smaller than the background, we leverage the Tversky loss function [34]. Tversky loss introduces two parameters, α and β, which control the penalties for false positives (FP) and false negatives (FN), respectively. The general form of the loss is shown in (8):

$$L_{Tversky} = 1 - \frac{TP + \varepsilon}{TP + \alpha \cdot FP + \beta \cdot FN + \varepsilon} \tag{8}$$

Here, TP, FP, and FN denote true positives, false positives, and false negatives. A larger α value increases the penalty on FPs, while a larger β value increases the penalty on FNs, compelling the model to minimize missed targets. However, manual hyperparameter tuning is a laborious trial-and-error process. To overcome this, our method adaptively adjusts these parameters for each subregion k based on its specific false positive and false negative rates, as defined in Eqs. (9) and (10):

$$\alpha_{Adaptive,k} = A + B \cdot \frac{FP_k + \varepsilon}{FP_k + FN_k + \varepsilon} \tag{9}$$

$$\beta_{Adaptive,k} = A + B \cdot \frac{FN_k + \varepsilon}{FP_k + FN_k + \varepsilon} \tag{10}$$

Where A and B are constants that regulate the range of the hyperparameters, and the terms FP_k and FN_k are computed for each subregion. To handle the varying segmentation difficulty across different image regions, as illustrated in Fig. 6, we first define the adaptive regional loss for each individual subregion k:

$$L^k_{region} = 1 - \frac{TP_k + \varepsilon}{TP_k + \alpha_{Adaptive,k} \cdot FP_k + \beta_{Adaptive,k} \cdot FN_k + \varepsilon} \tag{11}$$

The final segmentation loss, L_{Seg}, is then computed as the sum of the losses from all K subregions. This approach allows for targeted optimization tailored to the unique characteristics of each region, thereby improving overall segmentation accuracy:

$$L_{Seg} = \sum_{k=1}^{K} L^k_{region} \tag{12}$$

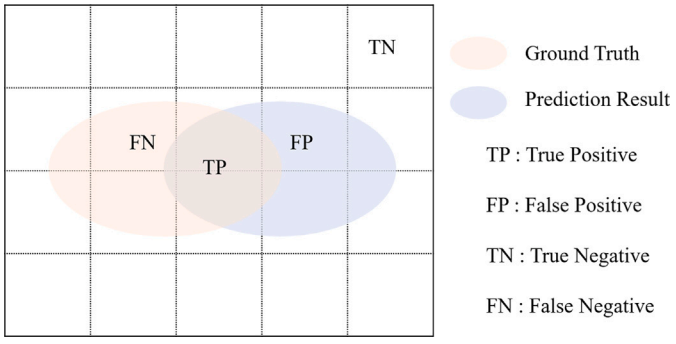


Fig. 6. The adaptive regional loss divides the image into multiple sub-regions (K sub-regions), calculating losses separately to optimize each region's characteristics, thereby enhancing segmentation accuracy by capturing distinct features more precisely.

3.4. Overall loss function

The AD-DGCL framework is optimized by a joint objective that combines multiple sub-loss terms:

$$L_{total} = L_{Seg} + \alpha L_{KL} + \beta L_{RC} + \eta L_{con} + \mu L_{contrast} \tag{13}$$

Here, α, β, η, and μ are hyperparameters that balance the contribution of each term. Following prior work, the KL term is given a small weight ($\alpha = 0.001$) to serve as an auxiliary regularizer, while the reconstruction term is weighted at $\beta = 0.1$ to preserve fidelity without dominating training. The consistency weight η is gradually increased using a ramp-up schedule to stabilize early learning. Finally, we perform a sensitivity analysis on the contrastive weight μto assess robustness, as detailed in our experiments.

This multi-term objective function establishes a collaborative learning process: the adaptive segmentation loss L_{Seg} drives accurate predictions on labeled data, particularly for small organs, while the disentanglement terms (L_{KL}, L_{RC}) and the generalization terms (L_{con}, $L_{contrast}$) leverage both labeled and unlabeled data to promote robust, domain-invariant feature learning.

4. Experiments

This section outlines the experimental setup, evaluation metrics, and performance of the proposed AD-DGCL framework. We validate its effectiveness through a comparison with state-of-the-art methods and detailed ablation studies, focusing on its ability to handle domain shifts and small organ segmentation.

4.1. Dataset and preprocessing

This study utilizes the Abdominal Organ Segmentation (AOS) dataset from the MICCAI FLARE 2024 challenge, a publicly available and de-identified dataset comprising 3D abdominal CT and MR scans collected from multiple medical centers. The dataset provides annotations for major abdominal organs (e.g., liver, spleen, kidneys) as well as challenging small organs (e.g., pancreas, stomach). In our semi-supervised domain generalization setting, the CT modality was used primarily as the labeled source domain, while the MR modality served as the unlabeled target domain.

From the full dataset, 50 CT scans and 100 MR scans were employed. To simulate the semi-supervised domain generalization setting, 40 CT scans and 8 MR scans were designated as labeled data for supervised training, with an 80/20 split for training and validation to tune hyperparameters. The remaining 80 MR scans were treated as unlabeled data during training to promote adaptation from CT to MR. For final evaluation, an independent test set comprising 10 CT and 12 MR scans was used, ensuring no subject overlap across subsets.

As the FLARE 2024 dataset is publicly released in a de-identified format, detailed patient demographic information (e.g., age, sex, scanner distribution) is not available. For preprocessing, all scans underwent intensity normalization to mitigate scanner-related differences, followed by cropping and padding to a uniform size of $160 \times 160 \times 32$ voxels to fit the 3D convolutional network.

4.2. Implementation details

All experiments were conducted on an NVIDIA GeForce RTX 3090 GPU using the PyTorch 2.0.0 framework, with a fixed random seed of 42 for reproducibility. V-Net [35] was adopted as the backbone for the segmentation network. To enhance generalization, we applied both weak augmentations (e.g., random brightness and noise) and strong augmentations (e.g., random cropping and color jittering). The model was trained for 500 epochs with the SGD optimizer and a batch size of 1 (one labeled and one unlabeled sample).

Performance was evaluated using Dice coefficient, 95 % Hausdorff Distance (HD95), and Average Surface Distance (ASD). We further

assessed the computational efficiency of our framework. For this purpose, 49 representative 3D volumes were selected from the training, validation, and test sets. The results indicate an average inference time of 8.16 ± 0.15 s per volume and a peak GPU memory usage of 2745.33 MB, which is well within the 24 GB capacity of the RTX 3090 GPU. These measurements serve as a reproducible reference, since variations in implementation details make direct computational comparisons across methods less straightforward.

4.3. Results and analysis

This section evaluates the performance of the AD-DGCL framework on the AOS dataset. Its effectiveness is validated in four aspects: first, comprehensive performance evaluation under domain shifts demonstrates the model's learning capacity; second, comparisons with state-of-the-art methods highlight its advantages in multi-domain semi-supervised segmentation; third, ablation studies quantify the contribution of each module; and finally, sensitivity analysis of the contrastive loss weight confirms the model's stability and robustness.

4.3.1. AD-DGCL performance on abdominal organ segmentation

As shown in Table 1, AD-DGCL effectively leverages unlabeled data with domain shifts, even with a small amount of labeled data. It achieves an average Dice score of 68.52 % on the multi-domain test set, significantly outperforming the fully supervised V-Net (42.93 %) and Mean-teacher (42.15 %). This result directly validates AD-DGCL's superior performance in handling multi-modal data. Notably, Mean-teacher performed worse than the fully supervised V-Net, suggesting that simple semi-supervised methods may fail in the presence of domain shifts and highlighting the necessity of AD-DGCL in addressing this problem.

Further analysis in Table 2 reveals that with the introduction of unlabeled MR scans, AD-DGCL's performance on the MR domain test set significantly improves, achieving an average Dice score of 67.35 %, higher than Mean-Teacher's 51.98 %.At the same time, its performance on the CT domain test set also increases, reaching an average Dice score of 69.68 %, surpassing both Mean-Teacher (33.92 %) and fully supervised V-Net (66.35 %). Table 2, which integrates MR and CT results, highlights AD-DGCL's substantial gains across both domains, validating its ability to leverage target domain data while maintaining cross-domain performance effectively.

Crucially, after incorporating unlabeled MR data, Mean-Teacher's Dice score on the CT test set sharply decreased from 66.35 % to 33.92 %. This catastrophic decline highlights Mean-Teacher's knowledge-forgetting failure under domain shifts, likely due to the pronounced domain gap. In contrast, AD-DGCL maintains a high Dice score on the CT domain, surpassing fully supervised V-Net. This provides strong evidence for the effectiveness of its domain-invariant learning, indicating that it successfully disregards domain-specific information such as image style to focus on unified anatomical structures.

To provide a clearer understanding of AD-DGCL's domain generalization ability, we employed Grad-CAM visualizations on the AOS dataset. As shown in Figs. 7 and 8, V-Net exhibits unstable and sample-dependent attention, while AD-DGCL consistently highlights anatomically meaningful structures across both CT and MR domains. This advantage stems from the synergistic effects of the SSRD and SCT modules, which guide the model to capture domain-invariant features, thereby ensuring robust segmentation even under domain shifts and in challenging small-organ scenarios.

4.3.2. Comparison with SOTA methods

To further validate the effectiveness of the proposed semi-supervised model, this section compares AD-DGCL with existing methods, including the classic Mean Teacher model [36], the BCP model [2], which uses a copy–paste strategy to mitigate distribution mismatch, and CSCADA [37], a cross-domain model designed for different anatomical structures. All comparison methods employ V-Net as their backbone. Experimental results indicate that AD-DGCL can be effectively applied to multi-domain semi-supervised scenarios, achieving precise segmentation predictions.

Table 3 reports the segmentation results of AD-DGCL and existing methods on the AOS dataset. AD-DGCL achieves the best overall Dice score (68.52 %), outperforming others on the liver, spleen, aorta, stomach, and left kidney. Although its Dice score on the right kidney (67.14 %) is slightly lower than CSCADA's (74.44 %), AD-DGCL substantially improves the left kidney (72.17 % vs. 57.81 %), indicating more balanced performance across paired organs. For small and challenging organs such as the stomach, AD-DGCL achieves a large gain (39.67 % Dice). Regarding boundary accuracy, AD-DGCL obtains the lowest HD95 (42.78 voxels) and ASD (10.11 voxels), confirming that its contours are closer to the ground truth and better preserve organ shapes and details.

Furthermore, Table 4 presents a comparative analysis of segmentation results on different domain test sets of the AOS dataset. When using Mean-teacher, while the segmentation of MR images improved, its performance on CT images suffered a notable decline, primarily because critical knowledge was discarded during the training process. In contrast, AD-DGCL maintained high-level performance across both the MR and CT domains, outperforming all other methods. This demonstrates that AD-DGCL effectively mitigates knowledge loss from cross-modal domain differences and leverages unified content information from labeled and unlabeled data to enable efficient semi-supervised learning. The final results show AD-DGCL achieving Dice scores of 67.35 % and 69.68 %, HD95 values of 23.19 and 62.37 voxels, and ASD values of 5.74 and 14.97 voxels on the MR and CT test sets, respectively, demonstrating robust multi-domain adaptability and high segmentation accuracy.

Table 1
The proposed AD-DGCL achieves robust organ segmentation Dice scores (%) on diverse domain test sets. Note that V-Net is a fully supervised baseline.

Method	Liver	Right kidney	Spleen	Aorta	Stomach	Left kidney	Average
V-Net (fully supervised)	68.22	33.85	36.33	56.10	31.37	31.69	42.93
Mean-teacher [36]	69.52	45.27	54.15	49.84	13.59	25.35	42.15
Proposed method	**85.32**	**67.14**	**67.12**	**79.72**	**39.67**	**72.17**	**68.52**

Table 2
The proposed AD-DGCL achieves robust organ segmentation Dice scores (%) on both MR and CT domains. Note that V-Net is a fully supervised baseline.

Method	Liver		Right kidney		Spleen		Aorta		Stomach		Left kidney		Average	
	MR	CT	MR	CT	MR	CT	MR	CT	MR	CT	MR	CT	MR	CT
V-Net	48.39	85.84	1.42	62.68	11.06	58.80	32.83	76.79	4.54	**55.22**	1.21	58.78	16.57	66.35
Mean-teacher [36]	78.65	60.40	51.99	38.54	61.06	47.24	71.09	28.58	16.18	11.00	32.94	17.75	51.98	33.92
Proposed method	**83.40**	**87.23**	**57.69**	**76.60**	**69.91**	**64.34**	**78.07**	**81.36**	**36.73**	42.61	**78.34**	**65.99**	**67.35**	**69.68**

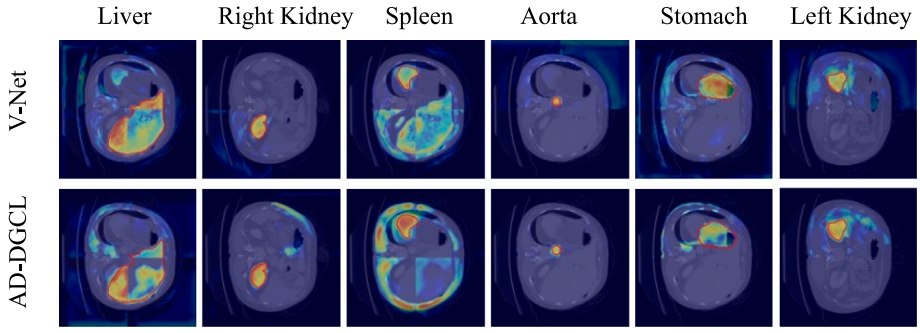


Fig. 7. Grad-CAM results of V-Net (fully supervised) and AD-DGCL on the CT domain of the AOS dataset. Red contours indicate ground-truth labels. Compared with V-Net, AD-DGCL provides more consistent attention on anatomical regions, reflecting better domain-invariant representation.

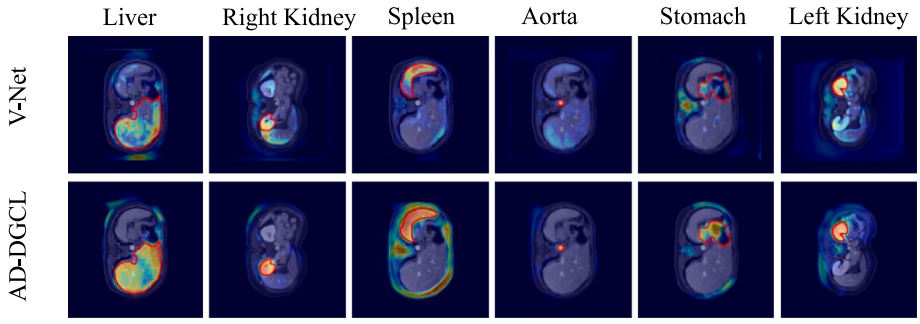


Fig. 8. Grad-CAM results of V-Net (fully supervised) and AD-DGCL on the MR domain of the AOS dataset. Red contours indicate ground-truth labels. AD-DGCL produces more stable and anatomically aligned attention than V-Net, showing improved cross-domain generalization.

Table 3
Comparison of organ-wise Dice (%), HD95, and ASD (voxels) between AD-DGCL and existing methods on the AOS dataset.

Metric	Method	Liver	Right kidney	Spleen	Aorta	Stomach	Left kidney	Average
Dice↑	Mean-teacher [36]	69.52	45.27	54.15	49.84	13.59	25.35	42.15
	BCP [2]	72.48	57.61	41.58	71.92	18.85	31.09	48.92
	CSCADA [37]	80.27	**74.44**	56.74	77.36	20.49	57.81	61.18
	Proposed method	**85.32**	67.14	**67.12**	**79.72**	**39.67**	**72.17**	**68.52**
HD95↓	Mean-teacher[36]	66.43	95.94	91.58	96.62	75.67	112.52	89.79
	BCP [2]	54.15	85.95	95.70	30.49	137.80	88.44	82.09
	CSCADA [37]	26.57	**49.12**	**48.87**	32.00	114.79	**17.87**	48.20
	Proposed method	**22.06**	60.11	84.21	**14.23**	**39.38**	36.73	**42.78**
ASD↓	Mean-teacher[36]	17.30	27.08	22.39	63.66	34.92	72.87	39.70
	BCP [2]	12.87	35.53	29.00	6.12	68.12	36.39	31.34
	CSCADA [37]	5.84	9.43	**10.35**	5.00	85.77	**7.09**	20.58
	Proposed method	**4.03**	**7.88**	15.33	**2.97**	**10.83**	19.64	**10.11**

Table 4
Comparison of domain-wise segmentation performance between AD-DGCL and existing methods on the AOS dataset.

Method	Testing set			Testing set (MR)			Testing set (CT)		
	Dice↑	HD95↓	ASD↓	Dice↑	HD95↓	ASD↓	Dice↑	HD95↓	ASD↓
Mean-teacher [36]	42.95	89.79	39.70	51.98	46.65	16.75	33.92	132.94	62.65
BCP [2]	48.92	82.09	31.34	40.68	82.47	38.83	57.17	81.70	23.84
CSCADA [37]	61.18	48.20	20.58	63.42	27.75	10.45	58.94	68.66	30.71
Proposed method	**68.52**	**42.78**	**10.11**	**67.35**	**23.19**	**5.74**	**69.68**	**62.37**	**14.97**

To visually demonstrate morphological differences in organ predictions between AD-DGCL and established semi-supervised methods, one MR and one CT example from the test set were selected for 3D segmentation visualization. Figs. 9 and 10 illustrate 3D segmentation results from AD-DGCL and other methods on MR and CT test sets, respectively. These visualizations demonstrate that, compared to other methods, AD-DGCL segments more complete organ contours in both domains, with left and right kidney segmentations closely matching the ground truth.

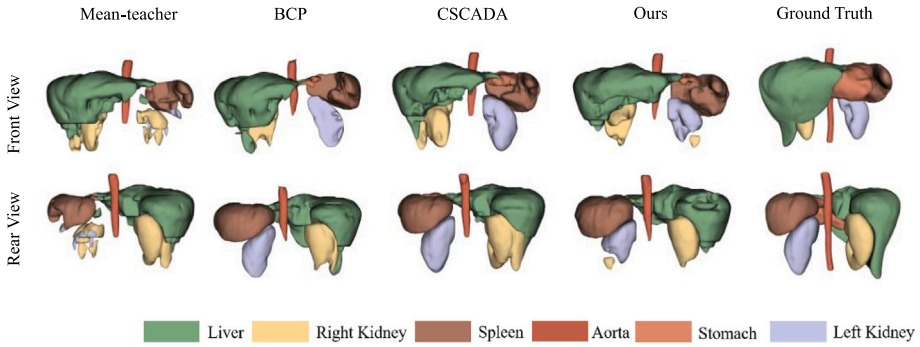


Fig. 9. 3D visualization of segmentation results on the MR test set. To intuitively illustrate the morphological differences between AD-DGCL and existing semi-supervised segmentation methods, a representative MR case from the test set is selected. Compared with other methods, AD-DGCL generates more complete and accurate organ contours, particularly for the left and right kidneys, closely matching the ground truth.

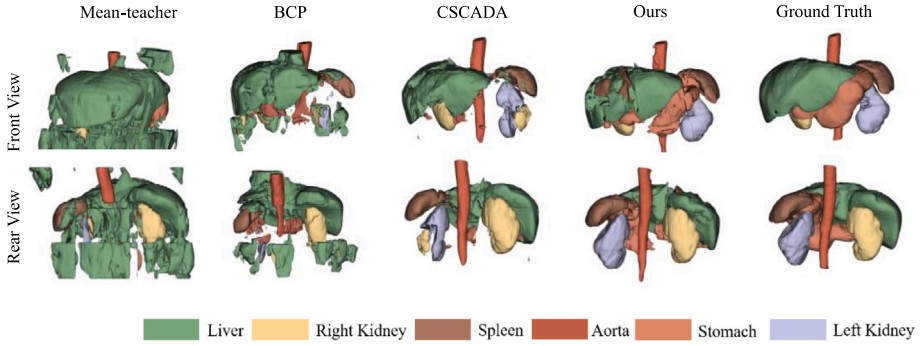


Fig. 10. 3D visualization of segmentation results on the CT test set. A representative CT case from the test set is selected to demonstrate the superiority of AD-DGCL in organ morphology prediction. The results show that AD-DGCL achieves more coherent and anatomically plausible segmentations across different organs compared to other methods.

Table 5
Ablation study on the impact of SSRD and SCT modules. Results show their individual and combined contributions to overall segmentation performance.

SSRD	SCT	Testing set			Testing set (MR)			Testing set (CT)		
		Dice↑	HD95↓	ASD↓	Dice↑	HD95↓	ASD↓	Dice↑	HD95↓	ASD↓
		42.93	100.50	46.52	16.57	113.35	61.08	66.35	89.08	33.59
✓		63.92	**37.14**	13.73	66.32	**15.01**	**4.54**	61.52	**59.28**	22.94
✓	✓	**68.52**	42.78	**10.11**	**69.68**	23.19	5.74	**65.18**	62.37	**14.97**

4.3.3. Ablation experiments

In this section, we conduct a series of ablation experiments to rigorously validate the contributions of the SSRD module and the SCT module within our AD-DGCL model.

Table 5 shows the impact of the SSRD and SCT modules on model performance. Experimental results indicate that introducing the SSRD module (second row) enables the model to learn unified anatomical features from labeled and unlabeled data, achieving strong performance across MR and CT domains. This is reflected by a notable increase in average Dice coefficient and substantial improvement in boundary accuracy (reduced HD95 and ASD values). This confirms that SSRD is essential for learning domain-invariant content features and mitigating the adverse effects of domain differences. With the addition of the SCT module (third row), the model's segmentation Dice further improves. However, HD95 slightly increases, indicating that minor outlier points may be introduced in segmentation predictions. Despite these minor outliers, the overall predicted segmentation shapes remain accurate, preserving high practical utility. This demonstrates that SCT effectively enhances the model's domain generalization by improving robustness to style variations.

Table 6 provides a detailed analysis of the contributions of each SSRD module component. Experimental results indicate that dual-branch style normalization (DBSN), dual encoders, and cross-domain content contrastive learning (C2DCL) each contribute variably to model performance. Optimal performance across all metrics is attained when all three components are combined (fourth row), highlighting the critical role of each in effective decoupling and domain-invariant feature learning. The superior performance of the full AD-DGCL model demonstrates the synergistic effects of the SSRD, SCT, and adaptive region-specific loss modules. Together, they offer an effective solution for semi-supervised 3D medical image segmentation, particularly under combined challenges of substantial domain shifts and small-organ segmentation.

4.3.4. Sensitivity analysis on the contrastive weight

As discussed in the overall loss function, the weights for KL divergence ($\alpha = 0.001$), reconstruction ($\beta = 0.1$), and consistency (η) follow established practices to ensure stable training. In contrast, the contrastive weight μplays a more critical role in learning domain-invariant features. To examine its effect, we evaluated $\mu \in \{0.01, 0.1, 0.5, 1.0\}$ with all other hyperparameters fixed. Each model was trained for 100 epochs,

Table 6
Ablation study on the SSRD components. Results show that each component contributes to performance, and their combination yields the best outcome, validating the overall design.

DSBN	Dual-Encoder	C2DCL	Testing set			Testing set (MR)			Testing set (CT)		
			Dice↑	HD95↓	ASD↓	Dice↑	HD95↓	ASD↓	Dice↑	HD95↓	ASD↓
			42.93	100.50	46.52	16.57	113.35	61.08	66.35	89.08	33.59
✓			60.85	62.93	16.07	59.25	40.89	7.59	62.45	66.61	24.55
✓	✓		64.67	63.70	30.67	63.23	51.19	25.97	61.52	59.28	22.93
✓	✓	✓	**68.52**	**42.78**	**10.11**	**67.36**	**23.20**	**5.25**	**69.68**	**62.37**	**14.97**

Table 7
Sensitivity of the contrastive weight μat epoch 100 on the validation set. The best Mean Dice performance is obtained with $\mu = 0.1$.

Contrastive weight μ	Mean dice (↑)
0.01	0.2954 ± 0.0786
0.10	**0.3405 ± 0.1082**
0.50	0.2954 ± 0.0786
1.00	0.2505 ± 0.0868

a reduced schedule sufficient to reveal relative hyperparameter effects while limiting computation. As shown in Table 7, $\mu = 0.1$ achieved the best Mean Dice (0.3405), whereas smaller or larger values led to performance degradation. These results confirm $\mu = 0.1$ as an effective trade-off, adopted in all main experiments.

5. Discussion

The proposed AD-DGCL framework mitigates domain discrepancies through a synergistic design. At the representation level, the SSRD module separates anatomical "content" from imaging "style," as supported by Grad-CAM visualizations that highlight consistent attention to clinical anatomy across domains. At the output level, the SCT module enforces prediction consistency under style shifts, while the adaptive region-specific loss improves sensitivity to small organs. Together, these mechanisms yield stable and clinically meaningful segmentations that remain robust under domain variations.

Compared with general-purpose domain generalization approaches that primarily focus on aligning global feature distributions or adopting causal inference paradigms, AD-DGCL offers a more targeted solution tailored for segmentation. By introducing an adaptive disentangled framework with a loss specifically designed to enhance sensitivity for small, clinically critical organs, it not only preserves fine-grained anatomical structures but also provides interpretable evidence through visualization. This targeted design delivers domain-robust representations critical for clinical practice and may serve as a reference for broader domain generalization problems beyond medical imaging.

Nevertheless, the proposed framework has trade-offs and limitations. A primary concern is computational efficiency, as the dual-encoder and consistency modules increase computational and memory costs, which may hinder large-scale deployment. Another limitation is our validation on a single benchmark, reflecting a broader challenge: the scarcity of suitable multi-modal datasets. This limitation is underscored by the design of the FLARE 2024 challenge—structured as an unsupervised task precisely because large-scale annotated MR data is rarely available—and it currently restricts extensive external validation for methods in this setting. Future work will therefore focus on: (i) validating AD-DGCL on new multi-modal datasets as they emerge; (ii) improving efficiency via lightweight architectures and knowledge distillation; and (iii) enhancing robustness in low-label regimes with uncertainty-aware strategies, ultimately aiming for practical clinical deployment.

6. Conclusion

We propose AD-DGCL, integrating representation disentanglement, style-induced consistency training, and an adaptive region-aware loss to mitigate domain disparities in semi-supervised 3D medical image segmentation. Extensive experiments on MICCAI FLARE2024 Task 3 (AOS dataset) demonstrate that AD-DGCL enhances segmentation accuracy and cross-modal generalization compared with prior methods. Importantly, the framework produces interpretable outputs, as evidenced by Grad-CAM, aligning with clinically relevant anatomical features and enhancing transparency and potential utility in clinical and health-data applications requiring explainability and robustness across imaging protocols.

CRediT authorship contribution statement

Min Dong: Writing – review & editing, Supervision, Resources, Project administration, Methodology, Conceptualization. **Yishuang Liu:** Writing – original draft, Software, Methodology. **Ating Yang:** Formal analysis, Data curation. **Ye Zhang:** Investigation, Formal analysis. **Rongchang Zhao:** Writing – review & editing, Supervision, Methodology. **Ling Liu:** Writing – review & editing, Conceptualization.

Declaration of generative AI/generative AI

During the preparation of this manuscript, the authors used DeepL Write and ChatGPT to improve the language and readability. After using these tools, the authors carefully reviewed and revised the content where necessary and take full responsibility for the final content of the publication.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the Henan Provincial Science and Technology Research Project, China (Grant No. 242102210002), the Zhengzhou's Science and Technology Innovation Guidance Program in Healthcare, China (Grant No. 2024YLZDJH026), and the National Natural Science Foundation of China (Grant No. 62372474).

Data availability

Data will be made available on request.

References

[1] S. Fu, Y. Lu, Y. Wang, Y. Zhou, W. Shen, E. Fishman, A. Yuille, Domain adaptive relational reasoning for 3d multi-organ segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2020, pp. 656–666.
[2] Y. Bai, D. Chen, Q. Li, W. Shen, Y. Wang, Bidirectional copy-paste for semi-supervised medical image segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 11514–11524.
[3] Y. Ji, H. Bai, C. Ge, J. Yang, Y. Zhu, R. Zhang, Z. Li, L. Zhanng, W. Ma, X. Wan, et al., Amos: a large-scale abdominal multi-organ benchmark for versatile medical image segmentation, Adv. Neural Inf. Process. Syst. 35 (2022) 36722–36732.

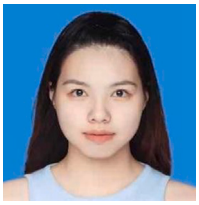
[4] R. Gu, G. Wang, J. Lu, J. Zhang, W. Lei, Y. Chen, W. Liao, S. Zhang, K. Li, D.N. Metaxas, et al., Cddsa: contrastive domain disentanglement and style augmentation for generalizable medical image segmentation, Med. Image Anal. 89 (2023) 102904.
[5] Z. Cai, J. Xin, C. You, P. Shi, S. Dong, N.C. Dvornek, N. Zheng, J.S. Duncan, Style mixup enhanced disentanglement learning for unsupervised domain adaptation in medical image segmentation, Med. Image Anal. 101 (2025) 103440.
[6] H. Basak, Z. Yin, Semi-supervised domain adaptive medical image segmentation through consistency regularized disentangled contrastive learning, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2023, pp. 260–270.
[7] X. Yang, R. Wang, D. Zhao, F. Yu, A.A. Heidari, Z. Xu, H. Chen, A.D. Algarni, H. Elmannai, S. Xu, Multi-level threshold segmentation framework for breast cancer images using enhanced differential evolution, Biomed. Signal Process. Control 80 (2023) 104373.
[8] J. Lian, Z. Yang, W. Sun, Y. Guo, L. Zheng, J. Li, B. Shi, Y. Ma, An image segmentation method of a modified SPCNN based on human visual system in medical images, Neurocomputing 333 (2019) 292–306.
[9] I. Mehidi, D.E.C. Belkhiat, D. Jabri, An improved clustering method based on k-means algorithm for MRI brain tumor segmentation, in: 2019 6th International Conference on Image and Signal Processing and Their Applications (ISPA), IEEE, 2019, pp. 1–6.
[10] Y. Feng, L.S. Chow, N.M. Gowdh, N. Ramli, L.K. Tan, S. Abdullah, S.S. Tiang, Gradient-based edge detection with skeletonization (GES) segmentation for magnetic resonance optic nerve images, Biomed. Signal Process. Control 80 (2023) 104342.
[11] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3431–3440.
[12] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 234–241.
[13] B. Rahmati, S. Shirani, Z. Keshavarz-Motamed, Semi-supervised segmentation of medical images focused on the pixels with unreliable predictions, Neurocomputing 610 (2024) 128532.
[14] J. Ma, Z. Nie, C. Wang, G. Dong, Q. Zhu, J. He, L. Gui, X. Yang, Active contour regularized semi-supervised learning for COVID-19 CT infection segmentation with limited annotations, Phys. Med. Biol. 65 (22) (2020) 225034.
[15] B.H. Thompson, G. Di Caterina, J.P. Voisey, Pseudo-label refinement using superpixels for semi-supervised brain tumour segmentation, in: 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI), IEEE, 2022, pp. 1–5.
[16] L.-L. Zeng, K. Gao, D. Hu, Z. Feng, C. Hou, P. Rong, W. Wang, Ss-TBN: a semi-supervised tri-branch network for COVID-19 screening and lesion segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 45 (8) (2023) 10427–10442.
[17] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A.C. Berg, W.-Y. Lo, et al., Segment anything, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 4015–4026.
[18] N. Li, L. Xiong, W. Qiu, Y. Pan, Y. Luo, Y. Zhang, Segment anything model for semi-supervised medical image segmentation via selecting reliable pseudo-labels, in: International Conference on Neural Information Processing, Springer, 2023, pp. 138–149.
[19] G. Bortsova, F. Dubost, L. Hogeweg, I. Katramados, M. De Bruijne, Semi-supervised medical image segmentation via learning consistency under transformations, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2019, pp. 810–818.
[20] Y. Ouali, C. Hudelot, M. Tami, Semi-supervised semantic segmentation with cross-consistency training, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 12674–12684.
[21] K. Wang, B. Zhan, C. Zu, X. Wu, J. Zhou, L. Zhou, Y. Wang, Semi-supervised medical image segmentation via a tripled-uncertainty guided mean teacher model with contrastive learning, Med. Image Anal. 79 (2022) 102447.
[22] H. Huang, Z. Chen, C. Chen, M. Lu, Y. Zou, Complementary consistency semi-supervised learning for 3d left atrial image segmentation, Comput. Biol. Med. 165 (2023) 107368.
[23] A.L. Maas, A.Y. Hannun, A.Y. Ng, et al., Rectifier nonlinearities improve neural network acoustic models, in: Proc. Icml, Vol.30, Atlanta, GA, 2013, p. 3.
[24] Y. Ge, Z.-M. Chen, G. Zhang, A.A. Heidari, H. Chen, S. Teng, Unsupervised domain adaptation via style adaptation and boundary enhancement for medical semantic segmentation, Neurocomputing 550 (2023) 126469.
[25] C. Shui, Q. Chen, J. Wen, F. Zhou, C. Gagné, B. Wang, A novel domain adaptation theory with Jensen–Shannon divergence, Knowl.-Based Syst. 257 (2022) 109808.
[26] B. Wang, J. Mendez, M. Cai, E. Eaton, Transfer learning via minimizing the performance gap between domains, Adv. Neural Inf. Process. Syst. 32 (2019).
[27] R. Fang, B. Li, Z. Kang, Q. Zeng, N.H. Dashtbayaz, R. Pu, B. Wang, C. Ling, On the benefits of attribute-driven graph domain adaptation, arXiv preprint arXiv:2502.06808 (2025).
[28] R. Fang, B. Li, J. Zhao, R. Pu, Q. Zeng, G. Xu, C. Ling, B. Wang, Homophily enhanced graph domain adaptation, arXiv preprint arXiv:2505.20089 (2025).
[29] Y. Yang, Z. Wang, W. Tao, X. Liu, Z. Jia, B. Wang, F. Wan, Spectral-spatial attention alignment for multi-source domain adaptation in EEG-based emotion recognition, IEEE Trans. Affect. Comput. 15 (4) (2024) 2012–2024.
[30] L. Wu, J. Zhuang, W. Chen, Y. Tang, C. Hou, C. Li, Z. Zhong, S. Luo, Data augmentation based on multiple oversampling fusion for medical image segmentation, PLoS One 17 (10) (2022) e0274522.
[31] Y. Song, J.Y.-C. Teoh, K.-S. Choi, J. Qin, Dynamic loss weighting for multiorgan segmentation in medical images, IEEE Trans. Neural Netw. Learn. Syst. 35 (8) (2023) 10651–10662.
[32] T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object detection, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2980–2988.
[33] D.P. Kingma, M. Welling, Auto-encoding variational bayes, arXiv preprint arXiv:1312.6114 (2013).
[34] S.S.M. Salehi, D. Erdogmus, A. Gholipour, Tversky loss function for image segmentation using 3D fully convolutional deep networks, in: International Workshop on Machine Learning in Medical Imaging, Springer, 2017, pp. 379–387.
[35] F. Milletari, N. Navab, S.-A. Ahmadi, V-net: fully convolutional neural networks for volumetric medical image segmentation, in: 2016 Fourth International Conference on 3D Vision (3DV), IEEE, 2016, pp. 565–571.
[36] A. Tarvainen, H. Valpola, Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results, Adv. Neural Inf. Process. Syst. 30 (2017).
[37] R. Gu, J. Zhang, G. Wang, W. Lei, T. Song, X. Zhang, K. Li, S. Zhang, Contrastive semi-supervised learning for domain adaptive segmentation across similar anatomical structures, IEEE Trans. Med. Imaging 42 (1) (2022) 245–256.

Author biography

Min Dong received the Ph.D. degree in Radio Physics from the School of Information Engineering, Lanzhou University, Lanzhou, China, in 2016. She is currently an Associate Professor with the School of Electrical and Information Engineering, Zhengzhou University, Zhengzhou, China. Her research interests include image processing, artificial intelligence, and defect detection.

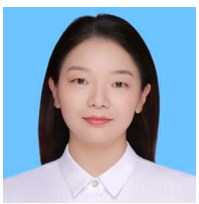
Yishuang Liu received the B.E. degree in Electronic Information Engineering from the School of Physics and Electronics, Shandong Normal University, Jinan, China, in 2024. She is currently pursuing the M.E. degree in the School of Electrical and Information Engineering, Zhengzhou University, Zhengzhou, China. Her research interests include medical image segmentation, semi-supervised learning, and domain generalization.

Ating Yang received the M.E. degree in Communication Engineering from the School of Electrical and Information Engineering, Zhengzhou University, Zhengzhou, China, in 2025. Her research interests include medical image segmentation, semi-supervised learning, and domain generalization.



Ye Zhang received the B.E. degree in Electronic Information Engineering from Liangjiang International College, Chongqing University of Technology, Chongqing, China, in 2023. She is currently pursuing the M.E. degree in the School of Electrical and Information Engineering, Zhengzhou University, Zhengzhou, China. Her research interests include image segmentation, deep learning, and medical image processing.

Rongchang Zhao now is an Associate Professor and Ph.D. Supervisor at the School of Computer Science, at Central South University. Before that, he received his Ph.D. and bachelor's degrees from Department of Electronic Information Science and Technology, Lanzhou University, China, in 2006 and 2011, respectively. Her research interests include Medical Image Analysis, Artificial Intelligence, Deep Learning, Computer Vision.



Ling Liu received the Doctor of Dental Medicine from the School of Dentistry, Centro Escolar University, Manila,the Philippines, in 2014. She is currently a Lecture with the School of Stomatology, Zhengzhou University, Zhengzhou, China. Her research interests include medical image processing, defect detection.