ITC 2/55
Information Technology and Control
Vol. 55 / No. 2/ 2026
pp. 423-446
DOI 10.5755/j01.itc.55.2.42493

Distribution-Constrained Semi-Supervised Learning for Medical Organ Segmentation

Received 2025/08/08 | Accepted after revision 2025/12/02

HOW TO CITE: Liu, L., Huang, S., Yu, B. (2026) Distribution-Constrained Semi-Supervised Learning for Medical Organ Segmentation. *Information Technology and Control*, 55(2), 423-446. https://doi.org/10.5755/j01.itc.55.2.42493

Distribution-Constrained Semi-Supervised Learning for Medical Organ Segmentation

Linjiang Liu
Changshu Hospital Affiliated to Nanjing University of Chinese Medicine, Changshu 215500, China

Shuting Huang*
School of Electronic Information, Guilin University of Electronic Technology, Beihai 536000, China

Bojian Yu
School of Marine Engineering Equipment, Zhejiang Ocean University, Zhoushan 316022, China

Corresponding author: hst2022hannah@gmail.com

Medical organ segmentation plays a crucial role in disease diagnosis and treatment planning. However, most existing methods focus primarily on the segmentation objects themselves while overlooking the positional correlations within images. These correlations can effectively align object features and provide boundary constraints for segmentation tasks. To address this limitation, this paper proposes a Distribution-Constrained Semi-Supervised Segmentation Model (DCSSM). Firstly, a Category Center of Mass Calculation Module (CCMCM) is introduced to capture positional correlations in medical images. Secondly, the model employs a shared encoder and a distribution network, which are pretrained using unlabeled samples to align input image features. The aligned feature boundaries then serve as constraints to guide mask generation during segmentation. Additionally, a Lightweight Multiscale Feature Extraction Module (LMFEM) is designed to efficiently transmit multiscale features to the decoder. Experimental results demonstrate that DCSSM achieves the best performance across multiple segmentation models on 5 medical segmentation datasets.

KEYWORDS: Medical organ segmentation, Distribution network, Segmentation network, Multiscale feature extraction

1. Introduction

Organ segmentation technology has become a critical component in medical imaging analysis, attracting considerable attention from both clinical and research communities [32]. This technique utilizes image processing algorithms to identify and delineate organ regions in medical images, enabling the differentiation between pathological and healthy tissues [3]. By precisely localizing and quantifying the occurrence and extent of diseases, organ segmentation assists clinicians in achieving a comprehensive understanding of patient conditions, thereby facilitating more accurate diagnostic decisions and personalized treatment planning.

The organ segmentation task primarily involves extracting key regions of interest, such as specific organs, from complex backgrounds or surrounding tissues in medical images. These extracted regions can then be spatially registered and aligned across different time points, imaging modalities, or anatomical locations, enabling subsequent comparative analysis, quantitative assessment, and multimodal data fusion. However, existing segmentation methods often rely predominantly on local image features while overlooking global structural information and contextual relationships. This limitation can lead to the loss of critical spatial distribution information during the segmentation process, consequently compromising the accuracy and consistency of segmentation results. Furthermore, the high cost and labor-intensive nature of medical image annotation result in a scarcity of labeled training samples, which adversely affects the performance of machine learning models and restricts their generalization capability and predictive accuracy [31]. Additionally, conventional segmentation techniques exhibit limitations in capturing precise organ positional information and fail to adequately model the intricate spatial relationships among multiple organs.

To address these challenges, this paper proposes a novel distribution-constrained semi-supervised segmentation model DCSSM. The proposed approach employs a Category Center of Mass Calculation Module (CCMCM) to capture positional correlation information across image slices. In addition, DCSSM incorporates a shared encoder and a custom-designed distribution network for pretraining purposes. Subsequently, a Lightweight Multiscale Feature Extraction Module LMFEM is utilized to construct the segmentation network architecture. Finally, the pretrained shared encoder and distribution network work collaboratively to align features from input images, with the aligned distribution boundaries serving as constraints to guide mask generation in the segmentation network.

The main contributions of this paper are as follows:

1. The custom-designed distribution network for pretraining refines the distribution learning task, enabling the model to segment a large number of unlabeled samples using only a limited number of labeled samples. This approach effectively addresses the challenge of high annotation costs in medical imaging.
2. The lightweight multiscale feature extraction module LMFEM employed in constructing the segmentation network achieves rapid feature extraction while maintaining high segmentation performance.
3. The pretrained shared encoder and distribution network collaboratively align the features of input images. This alignment process provides more accurate spatial distribution information, thereby improving the model's overall computational accuracy and segmentation consistency.
4. Comprehensive comparative and ablation experiments were conducted on the publicly available datasets, comparing the proposed method against eight baseline models. The experimental results demonstrate the effectiveness of the proposed approach.

2. Related Work

Traditional medical image segmentation relies on mathematical principles, using feature extraction, description, and morphological theory as the basis for segmentation techniques. These models do not typically depend on large amounts of labeled data; instead, they use basic features such as image intensity, textures, edges, and regions for segmentation. Consequently, they are easy to understand, computationally efficient, and highly flexible. Common

examples of these models include the snake model [1], thresholding methods [24], and others. As science and technology advance and medical standards improve, the demand for modeling and methods in the medical field has gradually increased. Traditional medical image segmentation methods perform poorly in tasks such as multi-organ and multi-structure segmentation, making it difficult to address the more complex and noise-prone medical tasks encountered today.

In recent years, the rapid development of artificial intelligence and deep learning technologies has significantly enhanced traditional medical image segmentation methods through integration of deep learning techniques. These methods leverage convolutional neural networks (CNNs) and their variants to learn, recognize, and extract features of target organs and tissue structures in medical images. The relationships between features and pixel values are used to delineate target regions and achieve segmentation [12, 13]. Compared to traditional methods, modern deep learning methods show significant improvements in both accuracy and efficiency, with their variant models widely applied in various fields [4,35].

The Segmentation Anything Model (SAM) [11, 19, 28], a typical deep learning-based segmentation method, offers interactive features and flexibility, enabling zero-shot learning and segmentation across various medical images. Roy et al. [21] evaluated the SAM model on an abdominal CT organ segmentation task, demonstrating its capability to segment without specific category samples. Wu et al. [26] proposed the Med-SA segmentation architecture, an adaptation of the original SAM model for medical use, which effectively improves the original model's training performance across multiple datasets. Zhang et al. [30] further developed a medical image segmentation model, SAMed, based on SAM. By fine-tuning the SAM encoder with a low-rank adaptation strategy and combining a prompt encoder and mask decoder, the model is trained on labeled medical datasets. Through optimization strategies, this approach achieves significant results on multi-organ segmentation datasets. Chen et al. [7] proposed the SAM-OCTA segmentation model for segmenting regions like blood vessels. This model refines segmentation areas by fine-tuning the base model and using low-rank adaptation techniques. It also uses prompt point generation strategies for multi-task segmentation to handle optical coherence tomography angiography (OCTA) segmentation. Building on this, Wang et al. [23] enhanced the SAM-OCTA framework by using parallel boxes to process original low-resolution images, employing a traditional network to predict rough masks, which are then refined into fine boundary boxes. A merging strategy reduces computational resources, enabling fine-tuning of feature information in low-resolution images.

SegNet, a medical image segmentation technique, is a pixel-level segmentation method [14]. The innovation of this model lies in optimizing the encoder-decoder architecture, making it highly suitable for semantic segmentation tasks. Building on this, Gai et al. [9] proposed the GL-SegNet model. In this model, the embedded feature encoder uses multi-scale convolution (MSC) and pooling (MSP) modules to represent information by encoding global semantics in the shallow layers. Simultaneously, multi-scale feature fusion enriches local geometric details across layers, achieving efficient segmentation. Kibriya et al. [14] introduced a semantic segmentation model based on SegNet. Using the encoder-decoder structure, the model leverages the indices of the maximum pooling layers from the encoder for upsampling and as inputs to the decoder. This approach enables the model to segment melanoma in dermoscopy images. Later, Ahmed et al. [2] proposed the Twin-SegNet framework. This model introduced several innovations to the original SegNet, using partial channel re-calibration techniques to compute channel attention weights for feature exchange and gradient propagation [29], significantly enhancing generalization while maintaining high segmentation accuracy and boundary quality.

Mask R-CNN is a model for instance segmentation in modern medical image processing [10]. Its main advantage is overcoming the limitation of previous models, which could only perform single-object segmentation. The Mask R-CNN-based segmentation framework and its variants effectively address this issue. Building on this, Lu et al. [25] proposed an improved Mask R-CNN network that enhances detection of small objects and feature pyramid fusion by reusing low-level feature information. This improvement enables precise and effective segmentation based on the characteristics of the or-

gans to be segmented. Later, Li et al. [16] proposed an automatic segmentation and detection model based on 3D-Mask R-CNN. This model uses the RoI Align method, an improvement over RoIPooling, to align proposed regions with feature maps. Yao et al. [27] introduced a transfer learning-based DenseSE-Mask R-CNN model that transfers tumor-related information from PET to MR images, providing prior knowledge and improving sensitivity to relevant information, thereby enhancing feature extraction and fusion for target images. Kiernan et al. [15] applied the Mask R-CNN to carotid intima-media thickness image segmentation. This network uses MimickNet B-mode as the primary framework, converting delayed-sum images into dynamically enhanced tissue contrast images. Singular value decomposition filtering is used to remove higher-order singular values and suppress random noise. This approach reduces speckle noise, providing the model with high accuracy and automation.

U-Net, a recently emerging image segmentation model, integrates skip connections, enabling better retention of global context information in images [38]. Building on this, Petit et al. [20] combined the U-shaped fully convolutional network with the Transformer attention mechanism, proposing the U-Transformer network. This network applies both self-attention and cross-attention modules to extract global structural features and filter out non-semantic features, improving segmentation accuracy for complex, low-contrast anatomical structures. Additionally, Wang et al. [22] proposed a hybrid Transformer U-Net model. By applying a local-global Gaussian-weighted self-attention mechanism, the model captures both local and global dependencies at different granularities, facilitating feature interaction. Later, Chen et al. [6] combined U-Net with Transformer and a multi-level attention mechanism to introduce the TransAttUnet segmentation model. This model uses Transformer self-attention and global spatial attention to capture long-range dependencies and spatial relationships between encoder features. To address the limitation of using Transformer solely as an auxiliary module, Zhou et al. [37] proposed the nnformer model. This model combines interleaved convolutions and self-attention operations, introducing a volume-based local and global self-attention mechanism to learn volumetric representations, making it highly complementary in ensemble learning. Cao et al. [5] proposed an improved deep neural network called RASNet, which captures multi-scale information using a multi-scale spatial perception module, retaining the target's original shape. The model uses a decoding module with attention connections to suppress background noise and refine target boundary prediction, improving segmentation accuracy.

Recent advances in lightweight network architectures have also demonstrated promise for medical image segmentation [34]. Zheng et al. [33] proposed MobileRaT, a lightweight radio transformer method that efficiently processes communication signals while maintaining high classification accuracy. Similarly, Fei et al. [8] introduced a real-time constellation image classification method based on MobileViT, demonstrating that lightweight networks can achieve competitive performance with significantly reduced computational costs. These developments in automatic classification technology [18] have inspired efficient feature extraction strategies in medical imaging. Furthermore, recent work on multi-scale feature hierarchies in detection transformers [17] has provided insights into effective feature pyramid designs. The concept of reconstruction error-based implicit regularization [36] has also been applied to medical diagnosis tasks, including lung cancer detection, demonstrating the importance of robust feature learning in clinical applications.

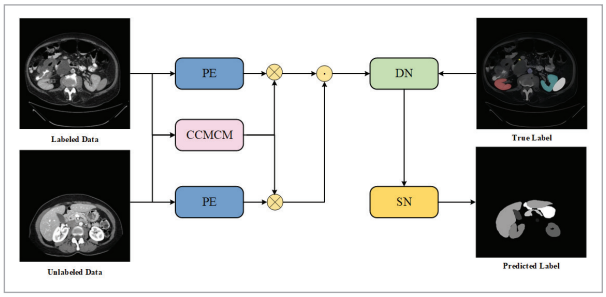
3. Method

3.1. Overview

As shown in Figure 1, DCSSM first inputs labeled and unlabeled volume data from each patient into the pretrained Public Encoder for feature extraction. Meanwhile, input data is fed into the pretrained Category Center of Mass Calculation Module (CCMCM) to compute category-specific spatial centroids.

The CCMCM serves as a spatial prior generator that encodes the anatomical position distribution of each organ category across slices. By capturing the sequential appearance pattern and spatial locality of different organs in volumetric data, CCMCM provides category-aware positional embeddings that guide the network to focus on anatomically relevant

Figure 1
Overall Architecture of DCSSM.



regions. The category centers and features of both labeled and unlabeled data are multiplied in matrices and concatenated. This matrix multiplication operation projects the extracted features onto the category center space, enabling the model to establish explicit associations between pixel-level features and organ-specific spatial priors, thereby enhancing the discriminability of features for different anatomical structures. The data is then distributed through the distribution network, and a loss is calculated using the True Label to optimize the distribution process. Finally, the self-designed segmentation network outputs predicted labels for the unlabeled data. The overall structure of DCSSM is shown in Figure 1 and can be formulated as:

$$F_l = Encoder(X_l) \tag{1}$$

$$F_u = Encoder(X_u) \tag{2}$$

$$C = CCMCM(X_l, X_u) \tag{3}$$

$$H = Concat\left(F_l \cdot C^T, F_u \cdot C^T\right) \tag{4}$$

$$PL = S\left(D(TL, H)\right), \tag{5}$$

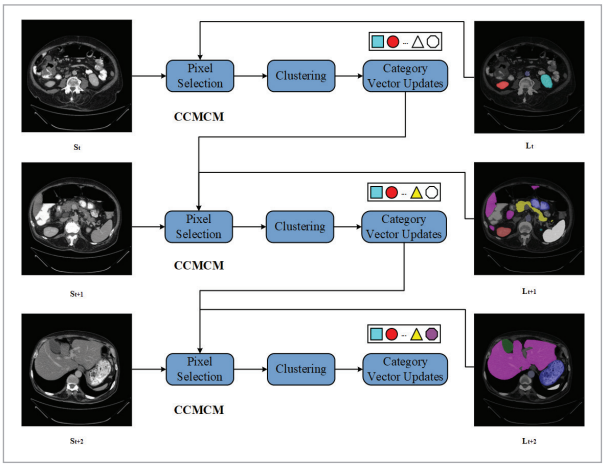
where X_l represents labeled volume data, X_u represents unlabeled volume data, F_1 denotes extracted labeled features, F_u denotes extracted unlabeled features, C contains category centers, and H is the combined representation. PL is the predicted label, TL is the true label, D is the distribution network, S is the segmentation network.

3.2. Category Center of Mass Calculation

Multi-organ segmentation in volumetric medical images faces the challenge of capturing both intra-slice spatial relationships and inter-slice anatomical continuity. Different organs exhibit characteristic spatial distributions and appearance orders along the axial direction. To explicitly model this anatomical prior knowledge, the Category Center of Mass Calculation Module (CCMCM) is designed to extract and maintain category-specific spatial representations that evolve across slices. To better capture the sequential information of input volume data and leverage the inherent spatial regularity of organ distributions, the Category Center of Mass Calculation Module is designed, shown in Figure 2.

First, the input volume data undergoes Pixel Selection on each slice based on corresponding labels. The coordinates of category pixels and slice numbers are merged into triplets. Then, K-means clustering is performed on each triplet, where the K value is set equal to the number of organ categories present in each slice (i.e., $K = N_c$, where N_c denotes the number of categories). This design choice is based on the principle that each organ category requires exactly one centroid to represent its spatial position, ensuring a one-to-one correspondence between clusters and anatomical structures. The clustering process dynamically calculates the centroids of each category label. Finally, the centroids of each category are stored in the category vector and updated iteratively. As slices are input sequentially, different categories

Figure 2
Structure of the CCMCM.



retain their centroids according to their appearance order, serving as the starting category vector for the next slice. The Category Center of Mass Calculation Module represents the spatial and slice position information of different category features and, to some extent, reflects the order of category appearance.

The centroid calculation for category k at slice t is:

$$c_k^t = \frac{1}{|P_k^t|} \sum_{(x,y,z)\in P_k^t} (x, y, z), \tag{6}$$

where P_k^t denotes the set of pixels belonging to category k in slice t. The category vector V_k is updated as:

$$V_k^t = \alpha V_k^{t-1} + (1-\alpha)c_k^t, \tag{7}$$

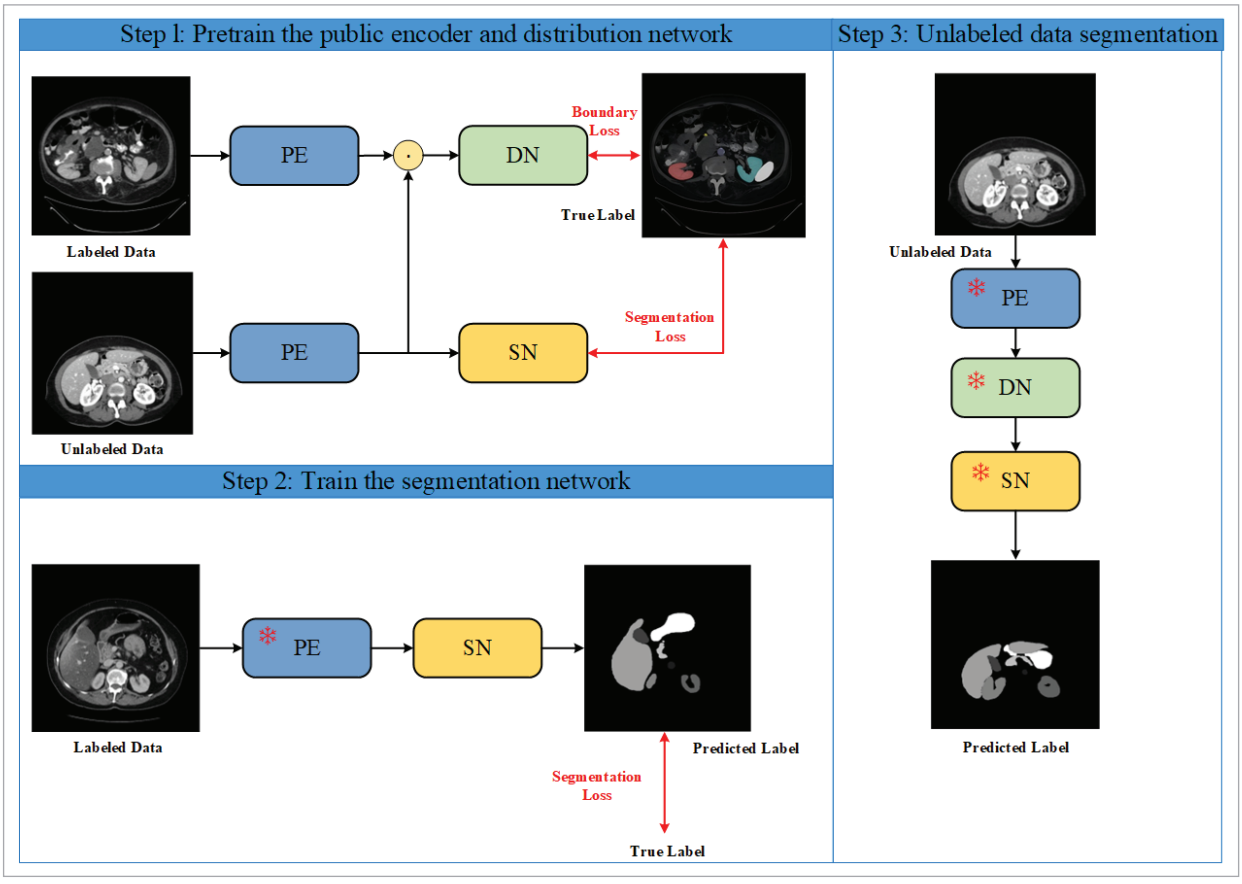
with α controlling the update rate.

By encoding where each organ typically appears in the volumetric space, CCMCM provides essential spatial prior guidance that informs the network about anatomically plausible organ locations. Furthermore, through the subsequent matrix multiplication with encoder features, CCMCM enables category-aware feature enhancement that amplifies responses in anatomically expected regions while suppressing irrelevant areas. Importantly, CCMCM also facilitates knowledge transfer from labeled to unlabeled data by using the learned category centers as a bridge—the spatial priors computed from labeled data can be applied to unlabeled volumes sharing similar anatomical structures, thereby supporting semi-supervised learning.

3.3 Multi-stage Fusion Training

To better leverage the distribution-assisted segmentation effects, a multi-stage fusion training method is designed as shown in Figure 3.

Figure 3
Workflow of the multi-stage fusion training.



The first step involves using the same Public Encoder to perform feature extraction on both labeled and unlabeled data, followed by concatenation. The concatenated features are input into the Distribution Network, and the edge loss is calculated based on labeled data labels. The parameters of the Public Encoder and the Distribution Network are updated until convergence. Meanwhile, the features of the unlabeled data are input into the Segmentation Network for loss calculation, and its parameters are updated accordingly. The second step involves inputting labeled data into the pretrained Public Encoder and the self-designed Segmentation Network to obtain the predicted label. The segmentation loss is calculated using labeled data labels, and the parameters of the Segmentation Network are updated until convergence. In the third step, the pretrained Public Encoder, Distribution Network, and Segmentation Network predict labels for the unlabeled data. This process uses a large amount of unlabeled data features to initialize the Segmentation Network's parameters and obtains a Public Encoder that bridges feature differences between labeled and unlabeled data, along with a well-distributed Distribution Network. Then, a small amount of labeled data is used to fine-tune the Segmentation Network and obtain the final model parameters. This training approach requires only a small amount of labeled data to achieve good distribution and segmentation performance, enabling collaborative optimization of both. The three-stage training process can be formulated as:

Stage 1 (Distribution Learning):

$$L_{dist} = E_{(x_l,y_l)\sim D_l}\left[CE_w\left(D\left(E(x_l)\right), y_l\right)\right], \quad (8)$$

where CE_w denotes the weighted cross-entropy loss designed to address the class imbalance problem in multi-organ segmentation. The class weight w_c for category c is computed using inverse frequency weighting:

$$w_c = \frac{N_{total}}{C \cdot N_c}, \quad (9)$$

where N_{total} represents the total number of pixels, C is the number of organ categories, and N_c denotes the number of pixels belonging to category c. The weighted cross-entropy loss is then formulated as:

$$CE_w = -\sum_{c=1}^{C} w_c \cdot y_c \cdot \log log(\hat{y}_c), \quad (10)$$

This weighting strategy assigns higher importance to under-represented organs (e.g., gallbladder and pancreas), thereby improving the model's ability to segment small organs accurately.

Stage 2 (Segmentation Fine-tuning):

$$L_{seg} = E_{(x_l,y_l)\sim D_l}\left[Dice\left(S\left(E(x_l)\right), y_l\right)\right], \quad (11)$$

where *Dice* denotes the Dice loss function, which directly optimizes the overlap between predicted segmentation masks and ground truth labels.

Stage 3 (Joint Optimization):

$$L_{final} = \lambda L_{dist} + (1-\lambda) L_{seg}, \quad (12)$$

where E, D, and S denote the Encoder, Distribution Network, and Segmentation Network.

3.4. Segmentation Network Architecture

As presented in Figure 4, the Segmentation Network's architecture is a U-shaped encoder-decoder structure.

The input passes through a Patch Embedding layer for embedding segmentation, followed by a Deformable Convolution layer for feature extraction. Next, the self-designed Lightweight Multiscale Feature Extraction Module performs rapid multiscale feature extraction, followed by downsampling. The features then pass through three downsampling and upsampling stages, with each stage accompanied by a Lightweight Multiscale Feature Extraction Module, restoring the features to their initial resolution. Finally, a Deformable Convolution layer and a Softmax layer are applied to produce the final segmentation result.

3.5. Lightweight Multiscale Feature Extraction

The structure of the Lightweight Multiscale Feature Extraction Module is shown in Figure 5. The input passes through a Deformable Convolution (DConv) layer for feature extraction, which does not alter the feature dimensions. The features are then split into three parts along the channels and input into DConv layers with different kernel sizes for multiscale feature extraction.

Figure 4
Structure of the Segmentation Network.

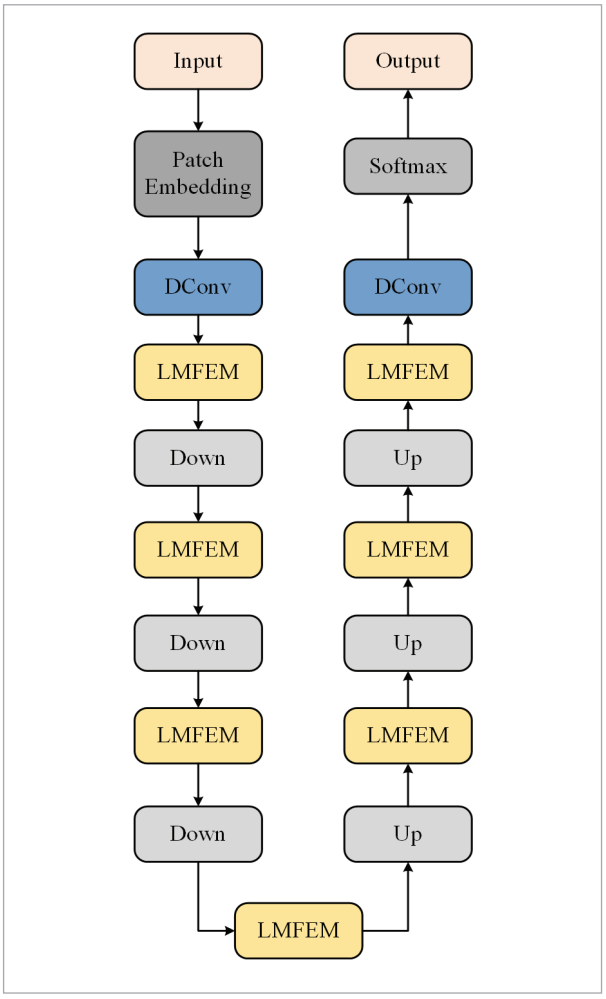
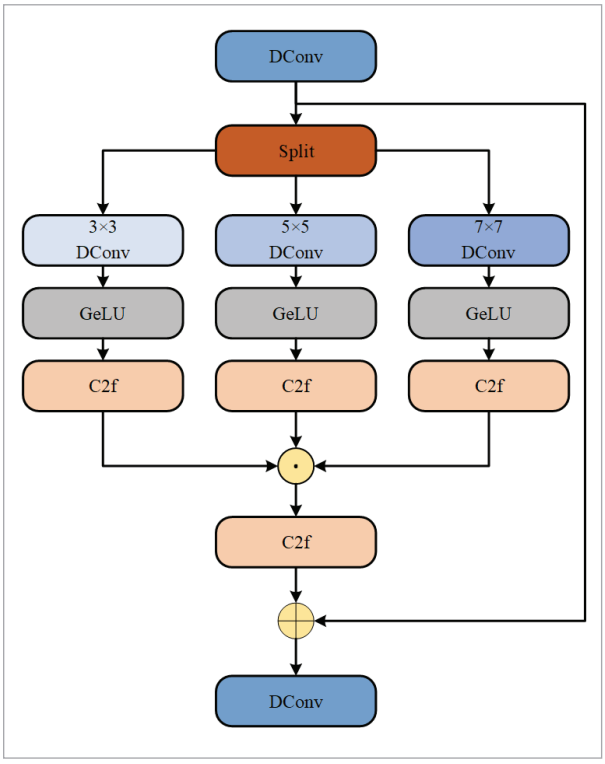


Figure 5
Structure of the LMFEM.



The channel division follows a 2:1:1 ratio, allocating more channels (50% of total) to smaller kernel sizes (3×3) for fine-grained feature extraction, while larger kernels (5×5 and 7×7) each receive 25% of channels for capturing broader contextual information. The 2:1:1 configuration achieves the optimal balance between fine detail preservation and contextual awareness, particularly beneficial for small organ segmentation where both local boundaries and surrounding anatomical context are critical. After activation with the GeLU function, the features undergo secondary extraction using the C2f module. Finally, the features from different scales are concatenated, restoring the feature dimensions to the input dimensions. The concatenated features are processed through another C2f module and a DConv layer to produce the final result. A residual connection is applied between the last DConv layer and the initial DConv layer to prevent overfitting and gradient vanishing.

4. Experiment

4.1. Datasets and Metrics

To comprehensively evaluate the performance and generalization capability of the proposed method, experiments are conducted on multiple benchmark datasets across different imaging modalities and anatomical structures. Specifically, the Synapse multi-organ CT dataset and ACDC cardiac MRI dataset are utilized to assess abdominal and cardiac structure segmentation, respectively. Additionally, three dermoscopic image datasets (ISIC 2017, ISIC 2018, and PH^2) are employed to evaluate skin lesion segmentation performance. These datasets collectively provide diverse challenges including varying organ sizes, different imaging modalities, and

distinct tissue characteristics, enabling a thorough assessment of model robustness and cross-domain adaptability.

The Synapse multi-organ segmentation dataset is a widely used benchmark for abdominal CT image segmentation. It comprises 30 abdominal CT cases with a total of 3,779 axial contrast-enhanced CT slices. The dataset provides annotations for 8 abdominal organs: aorta, gallbladder, left kidney, right kidney, liver, pancreas, spleen, and stomach. Together with the background class, the segmentation task involves 9 semantic categories. The dataset presents significant challenges due to varying organ sizes, low contrast between adjacent structures, and considerable inter-patient anatomical variability.

To validate the cross-modality generalization capability of the proposed method, additional experiments are conducted on the Automated Cardiac Diagnosis Challenge (ACDC) dataset. The ACDC dataset contains cine-MRI examinations from 100 patients, acquired using different 1.5T and 3.0T MRI scanners with varying temporal resolutions. The dataset covers patients with different cardiac pathologies including dilated cardiomyopathy, hypertrophic cardiomyopathy, myocardial infarction, and abnormal right ventricle, as well as healthy subjects. Manual annotations are provided for three cardiac structures: right ventricle (RV), left ventricle (LV), and myocardium (MYO) at both end-diastolic and end-systolic phases. All ground truth segmentations were manually delineated by clinical experts with extensive experience in cardiac MRI analysis and subsequently validated by experienced cardiologists to ensure annotation accuracy. The ACDC dataset presents unique challenges including large inter-subject anatomical variability, varying image quality, and intensity inhomogeneity typical of MRI acquisitions.

The ISIC 2017 dataset was released by the International Skin Imaging Collaboration (ISIC) for the Skin Lesion Analysis Toward Melanoma Detection challenge. It contains 2,750 dermoscopic images divided into 2,000 training images, 150 validation images, and 600 test images. The images are categorized into three diagnostic classes: melanoma, nevus, and seborrheic keratosis. Notably, images were captured by various imaging devices from multiple clinical centers, resulting in non-uniform resolutions ranging from 300×200 to 6,000×4,000 pixels. Each image is accompanied by expert manual tracings of lesion boundaries represented as binary masks, providing gold-standard ground truth for the segmentation task.

The ISIC 2018 dataset is an extension of the ISIC challenge series, primarily derived from the HAM10000 (Human Against Machine with 10,000 training images) collection. It contains 10,015 dermoscopic images categorized into seven diagnostic classes: melanoma (MEL), melanocytic nevus (NV), basal cell carcinoma (BCC), actinic keratosis (AK), benign keratosis (BKL), dermatofibroma (DF), and vascular lesion (VASC). A rigorous multi-level ground truth confirmation system is employed in this dataset: all malignant diagnoses were confirmed through histopathology, while benign diagnoses were validated via histopathology, serial imaging follow-up (3 visits or 1.5 years showing no change), or consensus of at least three expert dermatologists.

The PH^2 dataset is a dermoscopic image database developed through collaboration between Universidade do Porto, Instituto Superior Técnico Lisboa, and the Dermatology Service of Hospital Pedro Hispano, Portugal. It contains 200 dermoscopic images of melanocytic lesions with 8-bit RGB format and uniform resolution of 768×560 pixels, including 80 common nevi, 80 atypical nevi, and 40 melanomas. All images were acquired under standardized conditions using the Tuebinger Mole Analyzer system with 20× magnification. Comprehensive medical annotations are provided by board-certified dermatologists, including manual lesion segmentation, clinical and histological diagnoses, and assessment of multiple dermoscopic criteria (pigment network, dots/globules, streaks, regression areas, blue-whitish veil, and colors). The rigorous annotation protocol executed by designated expert dermatologists under standardized conditions ensures high-quality ground truth for algorithm evaluation.

The evaluation metrics on the Synapse multi-organ segmentation dataset and the Automated Cardiac Diagnosis Challenge (ACDC) dataset include the Dice similarity coefficient (DSC) and Hausdorff distance (HD), two metrics commonly used in image segmentation. The DSC measures the overlap between the predicted segmentation and the ground truth, while the HD evaluates boundary differences using the maximum surface distance at the 95th

percentile. Higher DSC scores and lower HD values typically correspond to better segmentation performance. On the skin lesion segmentation dataset, we used SE, SP, and ACC, in addition to DSC, to evaluate the model. SE indicates the model's ability to identify all actual positive samples, reflecting its capacity to capture positive samples. SP measures the model's ability to identify all actual negative samples, reflecting its ability to exclude negative samples. ACC is the proportion of samples correctly classified by the model to the total number of samples, reflecting overall classification accuracy.

4.2. Implementation Details

The experiment was conducted on an RTX 4080 GPU with 16GB of memory, using the PyTorch framework for model training. Before feeding the data into the network, augmentation techniques such as rotation and flipping were applied, with an input image size of 224×224. The batch size was set to 4, and the Adam optimizer, using Adaptive Moment Estimation (Adam), was employed with an initial learning rate of 0.02. The loss function was defined as a weighted sum of Dice loss and focal loss, with a weight ratio of 0.7:0.3. The maximum number of training epochs was set to 200. Early stopping was implemented with a patience of 20 epochs based on validation loss to prevent overfitting. Training convergence was determined when the validation loss did not improve for 20 consecutive epochs or when the maximum epoch count was reached. The average training time for DCSSM on the Synapse dataset was approximately 4.2 hours, demonstrating reasonable computational efficiency for research and clinical deployment.

To ensure fair comparison, baseline methods were configured as follows. For methods with officially released code (e.g., TransUNet, Swin-Unet, HiFormer, MISSFormer), we used the authors' recommended hyperparameters as specified in their original papers or code repositories. For methods without official implementations (e.g., UltraLightUNet), we adopted the unified training settings described above. All methods shared identical dataset splits, input resolution (224×224), data augmentation strategies (random rotation and flipping), and early stopping criteria (patience of 20 epochs) to ensure fair comparison. Table 1 summarizes the training settings for our proposed DCSSM.

Table 1
Unified Training Settings for DCSSM.

Parameter	Value
Optimizer	Adam
Initial Learning Rate	0.02
Learning Rate Decay	Step (0.9/10ep)
Batch Size	4
Input Size	224×224
Maximum Epochs	200
Early Stop Patience	20 epochs
Loss Function	Dice+Focal (0.7:0.3)
Data Augmentation	Rotation, Flip
GPU	RTX 4080
Framework	PyTorch 1.12

4.3. Performance Comparison on Multi-Organ Segmentation

4.3.1 Quantitative Comparison on the Synapse Dataset

Table 2 presents the quantitative comparison of our proposed DCSSM method with 8 existing baseline methods on the Synapse dataset. All results are reported as mean ± standard deviation across all test cases. The table demonstrates that the DCSSM method shows strong performance advantages in multi-organ segmentation tasks.

First, DCSSM achieves the best overall segmentation performance, with a DSC of 82.89±1.26%, the highest among all methods, and an HD of 18.20±2.15, also the lowest among all compared methods. This indicates that DCSSM excels in both segmentation accuracy and boundary fitting. To validate statistical significance, we performed paired t-tests comparing DCSSM with the second-best method (UltraLightUNet) across all test cases. Specifically, for each test case, we computed the difference in DSC and HD between the two methods, then performed a two-tailed paired t-test on these paired differences. The t-statistic was calculated as $t = \bar{d}/\left(s_d/\sqrt{n}\right)$, where $\bar{d}$ is the mean difference, s_d is the standard deviation of

differences, and n is the number of test cases. The improvements in DSC ($p = 0.003$) and HD ($p = 0.008$) are statistically significant at the $\alpha = 0.01$ level, with 95% confidence intervals of [0.18%, 0.74%] for DSC improvement and [0.35, 1.95] for HD reduction. This confirms that the observed improvements are reliable and not due to random variation.

Second, DCSSM shows notable improvement in segmenting challenging small organs such as the gallbladder (72.01±2.04%) and pancreas (67.93±2.12%), outperforming the second-best gallbladder result (U-Net: 69.72±2.56%) by 2.29 percentage points. This highlights its strong capability in handling difficult segmentation cases. Additionally, DCSSM demonstrates stable performance in segmenting common large organs such as the kidney (88.89±1.45%), spleen (91.25±1.32%), and liver (95.02±0.98%), leading or matching other methods.

From a clinical perspective, the improvements have practical significance: the 2.81% enhancement in pancreas segmentation DSC (67.93% vs 65.12% for UltraLightUNet) is particularly valuable for pancreatic cancer diagnosis, where accurate tumor boundary delineation directly impacts surgical resectability assessment and treatment planning. Similarly, the improved gallbladder segmentation accuracy facilitates more precise detection of gallbladder lesions and surgical planning for cholecystectomy.

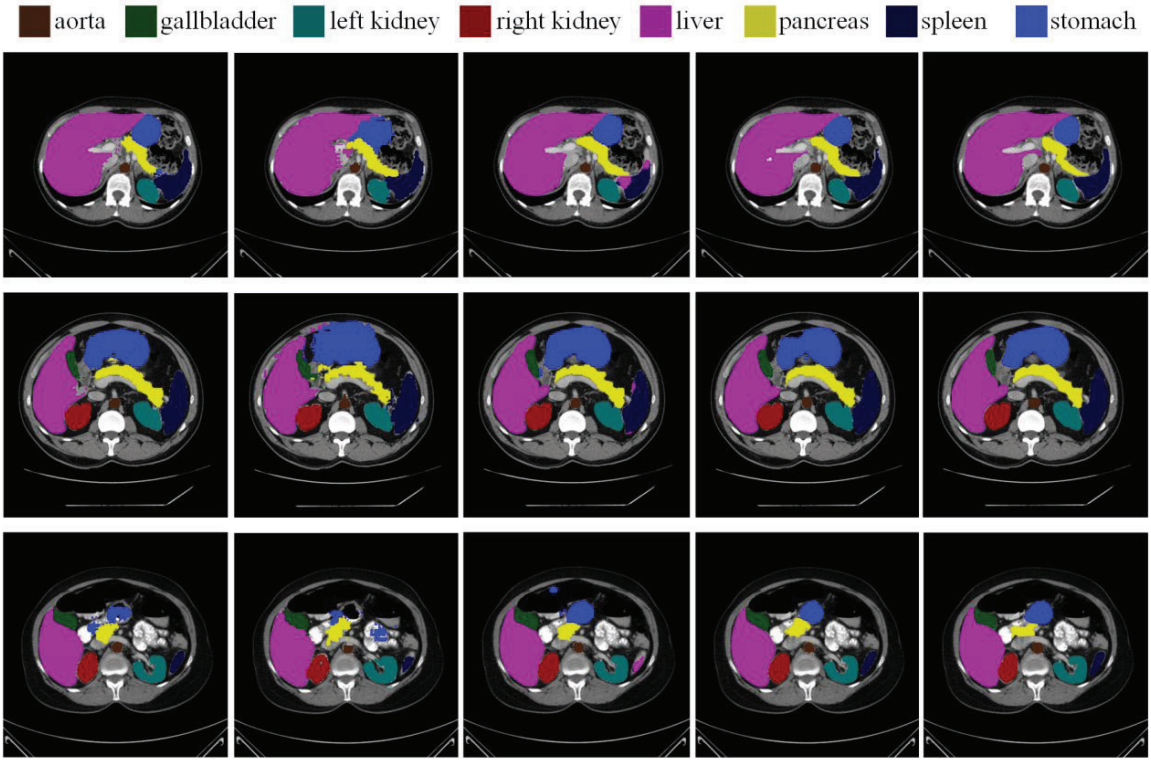
4.3.2. Visualization and Feature Analysis

The results in Figure 6 demonstrate that the DCSSM method exhibits significant advantages in multi-organ segmentation tasks. Compared to other methods, DCSSM achieves more precise boundary alignment with the ground truth in segmenting large organs such as the liver, spleen, and stomach, resulting in higher boundary accuracy. Additionally, DCSSM performs better in segmenting small organs like the gallbladder and pancreas, providing more complete results and reducing missed detections and false positives, while other methods show no-

Table 2
Comparison results on the Synapse dataset.

Methods	DSC (%) ↑	HD ↓	DSC of each organ							
			Aorta	Gal.	Kid. (L)	Kid. (R)	Liver	Pan.	Spleen	Sto.
U-Net	76.85± 2.34	39.70± 3.01	89.07± 2.12	69.72± 2.56	77.67± 2.23	68.60± 2.45	93.43± 1.12	53.98± 2.67	86.67± 1.78	75.58± 2.34
TransUNet	77.48± 2.23	31.69± 2.89	87.23± 2.01	63.13± 2.45	81.87± 2.12	77.02± 2.34	94.08± 1.08	55.86± 2.56	85.08± 1.67	75.62± 2.23
AttnUNet	75.57± 2.31	36.97± 2.98	55.92± 2.78	63.91± 2.56	79.20± 2.23	72.71± 2.45	93.56± 1.15	49.37± 2.67	87.30± 1.89	74.95± 2.34
TransDeepLab	80.16± 2.12	21.25± 2.67	86.04± 1.98	69.16± 2.34	84.08± 2.01	79.88± 2.23	93.53± 1.12	61.19± 2.45	86.83± 1.78	77.41± 2.12
Swin-Unet	79.13± 2.18	21.55± 2.73	85.47± 2.01	66.53± 2.45	83.28± 2.08	79.61± 2.28	94.29± 1.09	56.58± 2.56	90.66± 1.65	76.60± 2.21
MISSFormer	81.96± 2.01	18.98± 2.45	86.99± 1.89	68.65± 2.34	85.21± 1.98	82.00± 2.12	94.41± 1.06	65.67± 2.45	91.92± 1.56	76.81± 2.08
HiFormer	80.39± 2.09	19.70± 1.87	86.21± 1.95	65.69± 2.41	84.07± 2.03	80.64± 2.21	93.66± 1.10	62.56± 2.48	89.97± 1.63	79.87± 2.15
UltraLightUNet	82.43± 1.98	19.35± 2.34	89.53± 1.87	66.45± 2.38	88.12± 1.95	87.32± 2.08	94.78± 1.04	65.12± 2.41	91.03± 1.52	76.08± 2.18
DCSSM	82.89± 1.26	18.20± 2.15	90.01± 1.78	72.01± 2.04	88.89± 1.45	87.98± 2.01	95.02± 0.98	67.93± 2.12	91.25± 1.32	80.91± 2.01

Figure 6
Qualitative results of different methods on Synapse dataset.



ticeable errors or omissions in these small organs. For complex-shaped organs, such as the pancreas and kidneys, DCSSM outperforms other methods by preserving the overall shape and finer details, avoiding the common issues of over-smoothing or blurry edges. Moreover, DCSSM demonstrates superior noise suppression, as its segmentation results contain fewer background artifacts, while other methods exhibit artifacts or incorrect segmentation in some background regions. These findings indicate that DCSSM surpasses U-Net, MISSFormer, and UltraLightUNet in boundary fitting, small organ segmentation, complex structure recognition, and noise suppression, showcasing its stronger ability in multi-organ segmentation.

The loss of each method on the Synapse dataset is shown in Figure 7. The DCSSM method shows significant advantages in 200 rounds of training. Compared with other methods, DCSSM has the smallest training loss and the fastest convergence, basically reaching a low loss value within the first 50 rounds, and then remaining stable thereafter. This indicates that DCSSM not only learns effective features quickly in the early stage, but also avoids overfitting and maintains a more stable validation loss curve.

Figure 7
Loss curves of different methods on Synapse dataset.

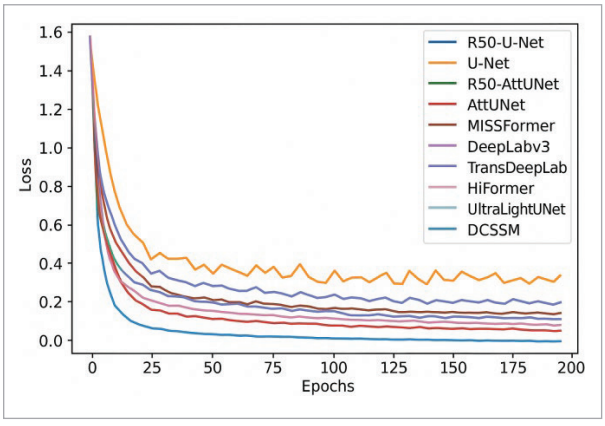


Figure 8 shows the heatmaps of the features of the DCSSM method with three baselines (U-Net, MISSFormer, UltraLightUNet) in three different examples of Synapse. By comparing these heatmaps, the advantages of DCSSM in processing medical images can be clearly seen. DCSSM performs particularly well in region localization, being able to focus more accurately on key regions and highlight more salient feature information, and the boundaries of the heat zones are more closely aligned with the real boundaries of the organs. This accurate focusing mechanism helps it outperform other methods when dealing with complex structures, especially in capturing the details of an image, which DCSSM performs more finely than U-Net and MISSFormer. In addition, compared with UltraLightUNet, DCSSM is able to achieve higher thermal image clarity and higher concentration of thermal regions in the center of the organ.

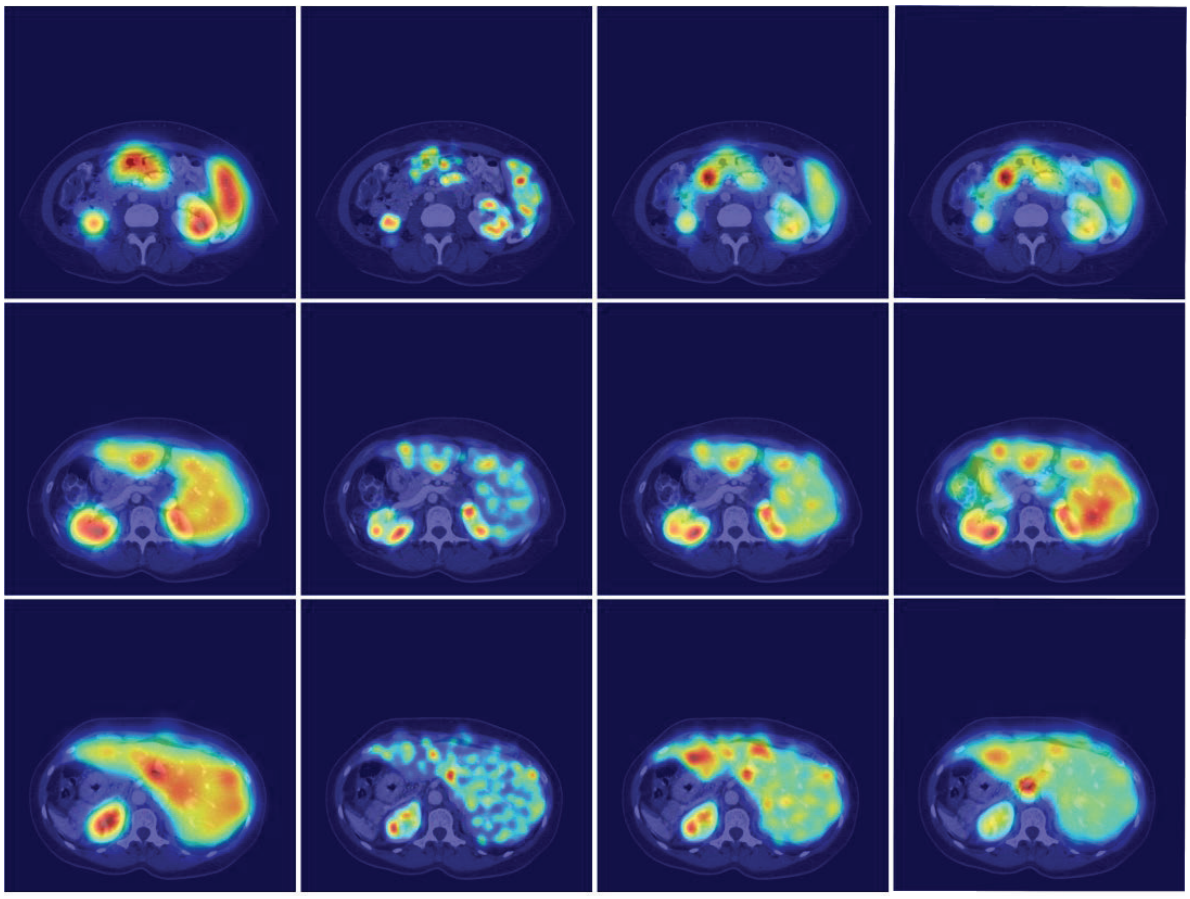
4.4. Validation on Cross-Task and Cross-Modality Generalization

4.4.1. Skin Lesion Segmentation on ISIC and PH² Datasets

In order to verify the generalization of the proposed method, experiments are conducted on three datasets ISIC 2017, ISIC 2018, and PH² in comparison with the baseline methods.

Table 3 presents the quantitative comparison results of our proposed DCSSM method with 8 baseline methods on the ISIC 2017, ISIC 2018, and PH²

Figure 8
Heatmaps of the different methods in the Synapse.



(a) DCSSM (b) U-Net (c) MISSFormer (d) UltraLightUNet

datasets. All results are reported as mean ± standard deviation. As shown in the table, the DCSSM model demonstrated consistent performance advantages across all three dermatological lesion segmentation benchmark datasets.

In terms of the core metric DSC, DCSSM achieved the highest scores on all three datasets, reaching 93.12±0.89%, 91.30±1.20%, and 94.79±1.12% on ISIC 2017, ISIC 2018, and PH², respectively. Compared to the second-best method HiFormer, DCSSM achieved improvements of 0.95%, 0.74%, and 0.77% on the three datasets, respectively. To validate the statistical significance of these improvements, we conducted paired t-tests between DCSSM and HiFormer across test samples. The results confirmed significant improvements on all three datasets: ISIC 2017 ($p = 0.002$, 95% CI: [0.38%, 1.52%]), ISIC 2018 ($p = 0.015$, 95% CI: [0.15%, 1.33%]), and PH² ($p = 0.038$, 95% CI: [0.04%, 1.50%]). The improvements on ISIC 2017 and ISIC 2018 are significant at the $\alpha = 0.01$ and $\alpha = 0.05$ levels respectively, while PH² shows marginal significance due to its smaller sample size. These results demonstrate that DCSSM's improvements are statistically reliable across diverse skin lesion datasets. For the sensitivity (SE) metric, DCSSM also demonstrated superior performance, achieving 94.10±1.09%, 92.47±0.92%, and 95.41±1.28% on the three datasets, respectively. Notably, on the PH² dataset, DCSSM achieved a sensitivity improvement of 7.07% over the conventional U-Net (95.41% vs. 88.34%), demonstrating its enhanced capability in capturing fine-grained features of lesion regions. The high sensitivity indicates that DCSSM can effectively identify true positive lesion pixels, which is crucial for clinical applications where missing lesion areas could lead to misdiagnosis.

In terms of specificity (SP), DCSSM achieved 97.80±0.27%, 96.66±0.40%, and 98.10±0.27% on ISIC 2017, ISIC 2018, and PH², respectively, all ranking first among the compared methods. The relatively low standard deviations indicate stable performance across different test samples.

Table 3

Comparison results on ISIC 2017, ISIC 2018, and PH² datasets.

Methods	ISIC 2017				ISIC 2018				PH2			
	DSC	SE	SP	ACC	DSC	SE	SP	ACC	DSC	SE	SP	ACC
U-Net	87.23± 1.78	87.94± 1.63	94.56± 0.82	93.42± 0.89	85.67± 1.58	86.45± 1.72	93.12± 0.85	92.78± 0.98	89.45± 1.48	88.34± 1.56	95.23± 0.72	93.12± 0.78
TransUNet	89.56± 1.42	90.28± 1.35	95.89± 0.61	95.23± 0.72	87.92± 1.45	88.76± 1.52	94.78± 0.68	94.45± 0.78	91.56± 1.21	90.67± 1.35	96.45± 0.48	94.89± 0.58
AttnUNet	88.14± 1.65	88.92± 1.58	95.02± 0.75	94.18± 0.81	86.53± 1.52	87.34± 1.64	93.89± 0.79	93.56± 0.89	90.23± 1.38	89.12± 1.48	95.78± 0.62	93.78± 0.72
Swin-Unet	91.03± 1.21	91.86± 1.27	96.72± 0.48	96.08± 0.58	89.34± 1.31	90.21± 1.45	95.56± 0.58	95.42± 0.68	93.12± 1.05	92.45± 1.28	97.34± 0.38	96.02± 0.48
TransDeepLab	89.87± 1.38	90.67± 1.45	96.01± 0.68	95.47± 0.74	88.26± 1.42	89.08± 1.55	95.02± 0.72	94.87± 0.79	91.89± 1.18	91.02± 1.38	96.72± 0.52	95.23± 0.62
MISSFormer	91.45± 1.15	92.31± 1.18	96.84± 0.52	96.25± 0.62	89.78± 1.28	90.69± 1.32	95.84± 0.62	95.68± 0.65	93.45± 1.12	92.89± 1.20	97.45± 0.42	96.18± 0.52
HiFormer	92.17± 1.08	93.02± 0.98	97.21± 0.46	96.89± 0.55	90.56± 1.18	91.42± 1.25	96.21± 0.55	96.12± 0.58	94.02± 0.98	93.56± 1.25	97.72± 0.32	96.58± 0.48
UltraLightUNet	91.89± 1.12	92.74± 1.22	97.08± 0.55	96.71± 0.58	90.21± 1.24	91.05± 1.38	96.03± 0.60	95.94± 0.62	93.78± 1.02	93.21± 1.30	97.56± 0.38	96.42± 0.50
DCSSM	93.12± 0.89	94.10± 1.09	97.80± 0.27	97.64± 0.51	91.30± 1.20	92.47± 0.92	96.66± 0.40	96.85± 0.52	94.79± 1.12	95.41± 1.28	98.10± 0.27	97.07± 0.60

Furthermore, DCSSM demonstrated leading performance in overall accuracy (ACC), achieving 97.64±0.51%, 96.85±0.52%, and 97.07±0.60% on the three datasets, respectively. These results validate the model's ability to maintain a balanced trade-off between sensitivity and specificity when processing complex skin lesion images with varying morphological characteristics.

It is worth noting that the performance gap between different methods is smaller on the PH^2 dataset, where most transformer-based methods achieve DSC scores above 93%. This can be attributed to the higher image quality and smaller dataset size of PH^2. Nevertheless, DCSSM still maintains its leading position, demonstrating robust generalization capability.

These experimental results indicate that DCSSM not only exhibits strong multi-organ segmentation capabilities as shown in previous experiments, but also effectively addresses the unique challenges in skin lesion segmentation, such as irregular lesion boundaries, diverse lesion morphologies, and low contrast between lesion and surrounding skin regions. The consistent improvements across three different datasets further validate the strong generalization capability of the proposed method.

4.4.2. Cardiac MRI Segmentation on the ACDC Dataset

To validate DCSSM's cross-modality generalization capability beyond CT and dermoscopic imaging, we conducted comprehensive experiments on the ACDC cardiac MRI dataset. Table 4 presents quantitative comparisons with baseline methods.

The results demonstrate that DCSSM maintains strong performance on cardiac MRI data, achieving 90.24±1.76% average DSC and 6.92±1.05 HD, outperforming all baseline methods. Statistical significance testing using paired t-tests confirmed the improvements over the second-best method HiFormer: DSC improvement ($p = 0.006$, 95% CI: [0.28%, 1.58%]) and HD reduction ($p = 0.021$, 95% CI: [0.08, 0.90]) are both statistically significant, with DSC improvement reaching the $\alpha = 0.01$ significance level.

Notably, DCSSM achieves consistent improvements across all cardiac structures: left ventricle (95.18±1.24%), right ventricle (88.32±1.96%), and myocardium (87.21±2.19%). The performance hierarchy across structures (LV > RV > MYO) aligns with clinical expectations, as the left ventricle typically exhibits clearer boundaries while the thin myocardium presents greater segmentation challenges. The 0.93% DSC improvement over HiFormer (90.24% vs 89.31%) and 6.6% reduction in HD (6.92 vs 7.41) demonstrate DCSSM's robust generalization to MRI despite being primarily designed on CT data.

Interestingly, while HiFormer and UltraLightUNet achieve comparable overall performance (89.31% vs 89.29% average DSC), they exhibit different strengths across cardiac structures: HiFormer performs better on LV (94.63% vs 94.41%) and MYO (86.35% vs 85.89%), whereas UltraLightUNet shows advantages on RV segmentation (87.56% vs 86.94%). In contrast, DCSSM consistently outperforms both methods across all three structures, demonstrating more balanced and robust segmentation capabilities.

The distribution network's boundary alignment mechanism proves effective in handling MRI-specific challenges including intensity inhomogeneity and varying contrast patterns. The CCMCM module successfully captures positional correlations of cardiac structures across different cardiac phases (ED/ES), validating its applicability beyond abdominal CT imaging.

Table 4
Comparison on ACDC cardiac MRI dataset.

Methods	DSC (%) ↑	HD ↓	LV	RV	MYO
U-Net	87.42±2.30	8.73±1.45	93.15±1.67	85.23±2.45	83.88±2.78
TransUNet	88.56±2.12	7.94±1.28	93.87±1.52	86.78±2.31	85.02±2.53
HiFormer	89.31±1.89	7.41±1.15	94.63±1.41	86.94±2.18	86.35±2.47
UltraLightUNet	89.29±1.94	7.53±1.21	94.41±1.38	87.56±2.08	85.89±2.38
DCSSM	**90.24±1.76**	**6.92±1.05**	**95.18±1.24**	**88.32±1.96**	**87.21±2.19**

These cross-modality results confirm that DCSSM's architectural design—combining distribution constraints with lightweight multiscale feature extraction—generalizes effectively across diverse medical imaging modalities (CT, dermoscopic imaging, and MRI), supporting its potential for broader clinical deployment.

4.5. Ablation Studies

4.5.1. Effectiveness of Model Components

To validate the effectiveness of each component in the proposed DCSSM, we conduct comprehensive ablation experiments on the Synapse dataset. The ablation study focuses on two key architectural components: the CCMCM (Category Centroid-guided Channel Modulation) module and the downsampling depth (SN_dn). Table 5 presents the quantitative results, with all metrics reported as mean ± standard deviation.

The ablation experiments are organized into two groups. In the first group, we fix the downsampling depth at 3 (SN_dn=3) and vary the number of CCMCM modules from 0 to 3 to evaluate the contribution of the CCMCM component. In the second group, we fix the number of CCMCM modules at 2 and vary the downsampling depth from 2 to 5 to assess the impact of different feature resolution levels. The complete DCSSM model adopts CCMCM×2 combined with SN_dn×3 as the default configuration, which corresponds to the optimal settings identified through these experiments.

The results in Table 5 demonstrate that the CCMCM module plays a crucial role in improving segmentation performance. Removing the CCMCM module entirely (CCMCM×0) leads to a significant performance drop, with DSC decreasing by 2.36% and HD increasing by 3.15. This degradation confirms that the category centroid-guided channel modulation mechanism is essential for accurate organ segmentation. Introducing CCMCM modules progressively improves performance, with the optimal configuration at two modules (CCMCM×2). However, further increasing to three modules results in performance degradation, particularly for challenging organs such as Kidney(R) and Pancreas, suggesting that excessive modules may introduce redundant feature transformations and lead to overfitting.

Table 5
Ablation results of DCSSM on the Synapse dataset.

Methods	DSC (%) ↑	HD ↓	DSC of each organ							
			Aorta	Gal.	Kid. (L)	Kid. (R)	Liver	Pan.	Spleen	Sto.
CCMCM×0	80.53± 1.92	21.35± 2.68	87.45± 2.25	69.23± 3.65	86.12± 1.98	85.34± 2.52	93.68± 1.22	64.82± 4.45	89.45± 1.85	78.12± 3.28
CCMCM×1	81.47± 1.78	20.18± 2.52	88.56± 2.12	70.45± 3.42	87.35± 1.82	86.23± 2.38	94.25± 1.15	66.12± 4.18	90.12± 1.72	78.65± 3.12
CCMCM×2	**82.89± 1.26**	**18.20± 2.15**	**90.01± 1.78**	**72.01± 2.04**	**88.89± 1.45**	**87.98± 2.01**	**95.02± 0.98**	**67.93± 2.12**	**91.25± 1.32**	**80.91± 2.01**
CCMCM×3	81.32± 1.85	19.86± 2.48	89.52± 2.02	70.25± 3.55	87.68± 1.88	84.67± 2.65	94.38± 1.12	63.58± 4.52	90.32± 1.75	80.12± 3.08
SN_dn×2	81.83± 1.72	20.45± 2.55	88.24± 2.18	70.78± 3.38	87.95± 1.78	86.52± 2.32	94.45± 1.10	66.92± 4.08	90.42± 1.68	79.35± 3.05
SN_dn×3	**82.89± 1.26**	**18.20± 2.15**	**90.01± 1.78**	**72.01± 2.04**	**88.89± 1.45**	**87.98± 2.01**	**95.02± 0.98**	**67.93± 2.12**	**91.25± 1.32**	**80.91± 2.01**
SN_dn×4	82.02± 1.75	21.68± 2.72	89.18± 2.08	70.42± 3.45	87.52± 1.85	86.85± 2.28	94.58± 1.08	65.35± 4.32	90.88± 1.65	80.35± 2.95
SN_dn×5	80.92± 1.95	20.92± 2.62	87.95± 2.28	69.52± 3.58	86.42± 1.95	85.18± 2.55	93.98± 1.18	65.02± 4.48	89.85± 1.82	79.48± 3.18

The downsampling depth significantly influences the trade-off between receptive field size and spatial resolution preservation. With insufficient downsampling (SN_dn×2), the model shows limited capability in capturing global contextual information. The optimal configuration (SN_dn×3) achieves the best balance between receptive field coverage and spatial detail preservation. Increasing to four downsampling stages causes HD to increase substantially, reflecting boundary localization degradation due to resolution loss. Further deepening to five stages leads to overall performance decline, confirming that excessive downsampling compromises fine-grained boundary details critical for accurate organ delineation.

The two components address complementary aspects of the segmentation task. The CCMCM module primarily benefits small and challenging organs by providing positional priors from category centroids, with notable improvements observed for Gallbladder and Pancreas. The downsampling depth mainly affects boundary precision as reflected in the HD metric. These results validate that CCMCM enhances semantic discrimination while appropriate downsampling depth ensures spatial precision. In summary, the ablation study confirms that both the CCMCM module and the downsampling configuration are essential for DCSSM's superior performance. The optimal configuration (CCMCM×2 + SN_dn×3) achieves the best trade-off between semantic feature enhancement and spatial resolution preservation.

4.5.2. Ablation Study on K Value Selection

To validate the effectiveness of the category-constrained K value setting in CCMCM, comprehensive ablation experiments were conducted comparing different K value strategies. As shown in Table 6, the following settings were evaluated: (1) fixed offsets from the category count (K = C±1, K = C±2), and (2) conventional data-driven methods including the elbow rule and silhouette coefficient.

The results demonstrate that the proposed K = C setting achieves substantially better performance with 82.89% DSC and 18.20 HD, significantly outperforming all alternatives. When K < C, fewer centroids than organ categories are produced, forcing multiple organs to share a single centroid. This leads to severe spatial ambiguity, particularly affecting anatomically adjacent organs. The most significant degradation is observed with K = C−2 (78.42% DSC, 25.67 HD), as critical organs like the pancreas (58.91% vs. 67.93%) and gallbladder (63.28% vs. 72.01%) suffer from inadequate spatial representation.

Table 6
Ablation results of K Value on the Synapse dataset.

Methods	DSC (%) ↑	HD ↓	DSC of each organ							
			Aorta	Gal.	Kid. (L)	Kid. (R)	Liver	Pan.	Spleen	Sto.
K = C−2	78.42± 2.13	25.67± 3.21	85.13± 2.45	63.28± 2.89	83.45± 2.34	81.67± 2.78	92.34± 1.45	58.91± 3.12	86.23± 2.18	73.56± 2.67
K = C−1	80.15± 1.87	22.34± 2.89	87.23± 2.21	66.45± 2.67	85.78± 2.01	84.23± 2.45	93.45± 1.28	62.34± 2.78	88.12± 1.89	76.12± 2.41
K = C (Ours)	**82.89± 1.26**	**18.20± 2.15**	**90.01± 1.78**	**72.01± 2.04**	**88.89± 1.45**	**87.98± 2.01**	**95.02± 0.98**	**67.93± 2.12**	**91.25± 1.32**	**80.91± 2.01**
K = C+1	80.67± 1.78	21.45± 2.76	87.56± 2.12	67.23± 2.54	86.12± 1.92	84.89± 2.38	93.78± 1.21	63.12± 2.65	88.67± 1.78	77.34± 2.32
K = C+2	79.23± 1.95	23.89± 2.98	86.12± 2.32	64.78± 2.73	84.56± 2.15	82.91± 2.61	92.89± 1.35	60.45± 2.89	87.34± 1.98	75.23± 2.53
Elbow Method	79.89± 1.92	22.78± 2.85	86.78± 2.25	65.89± 2.68	85.23± 2.08	83.67± 2.52	93.12± 1.31	61.78± 2.81	87.89± 1.91	76.45± 2.45
Silhouette Coefficient	79.56± 2.01	23.12± 2.91	86.34± 2.29	65.12± 2.75	84.89± 2.12	83.12± 2.58	92.98± 1.38	61.23± 2.87	87.56± 1.95	75.89± 2.49

Conversely, when K > C, redundant clusters fragment single organs into multiple centroids, disrupting the anatomical correspondence and introducing noise in category center tracking. The performance degradation is found to be asymmetric: under-clustering (K < C) causes more severe harm than over-clustering (K > C), as missing centroids completely lose organ-specific spatial information, while redundant centroids only introduce partial noise.

Notably, conventional data-driven methods (elbow rule: 79.89% DSC; silhouette coefficient: 79.56% DSC) are observed to substantially underperform compared to K = C. These methods optimize for general clustering quality metrics rather than anatomical constraints, often selecting K values that deviate from the actual number of organ categories. These findings confirm that the deterministic K = C setting is both theoretically grounded in anatomical principles and empirically optimal for multi-organ segmentation.

4.5.3. Channel Division Strategy in Multi-Scale Extraction

To validate the effectiveness of the channel division strategy in the Lightweight Multiscale Feature Extraction Module (LMFE), ablation experiments are conducted comparing six different channel allocation ratios. The ratios are denoted as *a:b:c*, where *a*, *b*, and *c* represent the proportions of channels allocated to 3×3, 5×5, and 7×7 depthwise convolution kernels, respectively. Experimental results are presented in Table 7.

The results demonstrate that the 2:1:1 ratio achieves optimal performance across all evaluation metrics, obtaining 82.89% DSC, 18.20 HD, and the highest segmentation accuracy for all three organs. Several key observations can be drawn from these experiments.

When more channels are allocated to larger kernels (1:2:1 and 1:1:2 configurations), the segmentation performance degrades substantially. The 1:1:2 ratio yields the poorest results with 80.41% DSC and 22.35 HD, representing decreases of 2.48% and 4.15 respectively compared to the optimal configuration. This degradation is particularly pronounced for the pancreas (63.75% vs. 67.93%), indicating that small organs with intricate boundaries require sufficient fine-grained features for accurate delineation. Conversely, over-allocating channels to the 3×3 kernel (3:1:1 and 4:1:1) also leads to performance decline. The 4:1:1 configuration achieves only 81.45% DSC, 1.44% lower than the optimal ratio. This suggests that while fine-grained features are essential, capturing broader contextual information through larger receptive fields remains crucial for understanding anatomical structures and spatial relationships.

The 2:1:1 configuration strikes an effective balance by allocating 50% of channels to small kernels for detailed boundary detection while preserving 50% for medium and large kernels to capture contextual information. This balance proves particularly beneficial for challenging organs like the pancreas, where both precise local boundaries and surrounding anatomical context are critical for accurate segmentation. The equal distribution baseline (1:1:1) achieves moderate performance (81.87% DSC), outperform-

Table 7
Ablation study on the channel division strategy.

Ratio	DSC (%) ↑	HD ↓	Liver	Pan.	Sto.
1:1:1	81.87±1.45	20.34±2.67	94.51±1.18	65.92±2.53	79.85±2.28
1:2:1	80.93±1.58	21.76±2.89	93.87±1.32	64.28±2.71	78.64±2.54
1:1:2	80.41±1.67	22.35±3.12	93.42±1.45	63.75±2.94	78.02±2.68
2:1:1	**82.89±1.26**	**18.20±2.15**	**95.02±0.98**	**67.93±2.12**	**80.91±2.01**
3:1:1	82.21±1.38	19.67±2.43	94.63±1.09	66.84±2.35	80.18±2.19
4:1:1	81.45±1.52	20.58±2.71	94.15±1.25	65.47±2.58	79.32±2.41

ing configurations biased toward larger kernels but underperforming the 2:1:1 ratio. This confirms that a slight emphasis on fine-grained feature extraction is beneficial for medical image segmentation tasks involving small anatomical structures.

In summary, these ablation results validate the design choice of the 2:1:1 channel division ratio, demonstrating that prioritizing fine-grained features while maintaining adequate contextual capacity yields optimal multiscale feature fusion for abdominal organ segmentation.

4.5.4. Ablation Study of Distribution Learning Loss

To validate the effectiveness of the weighted cross-entropy loss in distribution learning stage, comparative experiments were conducted with alternative distribution losses. Specifically, the proposed weighted cross-entropy (WCE) loss was compared against standard cross-entropy (CE) loss without class balancing, Kullback-Leibler (KL) divergence loss, and Focal loss. All experiments were performed on the Synapse dataset under identical training configurations.

Table 8 presents the comparison results of different distribution losses. The weighted cross-entropy loss achieves the best overall performance with 82.89±1.26% DSC and 18.20±2.15 HD, outperforming standard cross-entropy by 3.77 percentage points in DSC and 3.40 in HD. This improvement is particularly pronounced for small organs: gallbladder segmentation improves from 67.34±2.89% to 72.01±2.04% (+4.67%), and pancreas segmentation improves from 63.21±3.01% to 67.93±2.12% (+4.72%). These results confirm that the inverse frequency weighting strategy effectively addresses the class imbalance problem inherent in multi-organ segmentation.

KL divergence loss yields suboptimal results (79.87±2.12% DSC) with relatively higher variance. Although KL divergence shows slight improvement over standard CE in overall DSC, it performs worse on several organs including Aorta, kidney (L), and spleen, indicating that distribution divergence minimization alone does not guarantee consistent improvement across all organ categories. Focal loss achieves 80.45±1.98% DSC by down-weighting well-classified samples. While Focal loss demonstrates improved performance on kidney (L) segmentation (87.23±1.98%), it shows degraded performance on gallbladder, spleen, and stomach compared to standard CE, suggesting that hard sample mining may not align well with the class imbalance characteristics of multi-organ segmentation. The standard cross-entropy loss without class balancing shows the lowest overall performance (79.12±2.34% DSC), demonstrating that explicit class balancing is essential for multi-organ segmentation tasks with significant size variations across organ categories.

4.6. Computational Efficiency Analysis

Table 9 presents a comprehensive comparison of computational costs among different methods, demonstrating DCSSM's favorable efficiency-accuracy tradeoff while maintaining competitive segmentation performance. DCSSM contains 12.36M parameters, achieving substantial reductions compared to transformer-based methods: 51.6%

Table 8
Ablation results of different loss on the Synapse dataset.

Methods	DSC (%) ↑	HD ↓	DSC of each organ							
			Aorta	Gal.	Kid. (L)	Kid. (R)	Liver	Pan.	Spleen	Sto.
CE (w/o weighting)	79.12± 2.34	21.60± 2.78	87.23± 2.45	67.34± 2.89	85.12± 2.67	74.36± 2.34	93.52± 1.56	63.21± 3.01	88.34± 2.12	76.23± 2.56
KL Divergence	79.87± 2.12	20.45± 3.12	86.89± 2.78	68.12± 1.85	86.34± 2.23	85.23± 2.89	93.12± 1.78	62.78± 2.67	83.45± 1.89	77.89± 2.78
Focal Loss	80.45± 1.98	19.78± 2.45	87.12± 2.02	66.78± 3.12	87.23± 1.98	86.12± 2.56	90.23± 1.34	64.56± 2.89	88.12± 2.44	75.45± 2.23
WCE (Ours)	82.89± 1.26	18.20± 2.15	90.01± 1.78	72.01± 2.04	88.89± 1.45	87.98± 2.01	95.02± 0.98	67.93± 2.12	91.25± 1.32	80.91± 2.01

fewer than HiFormer (25.51M), 54.5% fewer than Swin-Unet (27.17M), and 70.9% fewer than MISSFormer (42.46M). Compared to the heavyweight TransUNet (105.32M), our model reduces parameters by 88.3%, while being slightly smaller than conventional CNN-based U-Net (17.26M). This parameter efficiency is enabled by the U-shaped encoder-decoder architecture with lightweight multiscale feature extraction modules (LMFEM), which employs channel splitting strategy (2:1:1 ratio) to allocate computational resources efficiently across different receptive fields. Notably, DCSSM maintains comparable parameter count to UltraLightUNet (10.28M), with only 20.2% additional parameters to accommodate the distribution network for semi-supervised learning and the CCMCM module for category-aware spatial prior generation.

In terms of computational complexity, DCSSM requires 10.52 GFLOPs per forward pass, representing reductions of 21.1% compared to HiFormer (13.33G), 40.3% compared to MISSFormer (17.62G), and 81.1% compared to TransUNet (55.73G). The modest increase over UltraLightUNet (8.64G, +21.8%) enables the integration of the three-stage fusion training framework and deformable convolutions while preserving computational efficiency suitable for resource-constrained clinical environments. The inference time of 32±2ms achieves real-time processing at approximately 31 FPS, meeting clinical deployment requirements including intraoperative guidance and point-of-care diagnostics. While slightly slower than UltraLightUNet (24ms) due to the deformable convolution operations and the distribution network forward pass, DCSSM provides substantially faster inference than transformer-based alternatives such as TransUNet (76ms), MISSFormer (52ms), and Swin-Unet (48ms). GPU memory consumption during training is 5.24GB with batch size of 4, which fits comfortably within the 16GB memory constraint of RTX 4080 and enables deployment on consumer-grade hardware. Compared to transformer-based methods that typically require 7-11GB memory, our approach reduces memory footprint by 28-54%, allowing for potential increase of batch sizes on the same hardware configuration. The average training time for DCSSM on the Synapse dataset is approximately 4.20±0.24 hours using the Adam optimizer with an initial learning rate of 0.02 and early stopping strategy, which is longer than UltraLightUNet (3.42h) due to the multi-stage fusion training process involving distribution learning, segmentation fine-tuning, and joint optimization stages, but substantially more efficient than transformer-based methods such as TransUNet (8.46h), MISSFormer (6.15h), and HiFormer (5.72h).

The efficiency advantages of DCSSM stem from several key architectural design choices: the LMFEM employs deformable convolutions with strategic channel splitting to reduce redundant computations while capturing multiscale features through different kernel sizes (3×3, 5×5, and 7×7); the public encoder is shared across labeled and unlabeled data processing paths, eliminating duplicate feature extraction during inference; and the CCMCM performs K-means clustering only during the prepro-

Table 9
Computational Efficiency Comparison.

Method	Params (M)	FLOPs (G)	Inference (ms)	Memory (GB)	Train (h)
U-Net	17.26	24.85	28±2	5.12±0.12	3.85±0.20
TransUNet	105.32	55.73	76±4	11.35±0.28	8.46±0.48
HiFormer	25.51	13.33	46±3	7.28±0.18	5.72±0.32
Swin-Unet	27.17	11.46	48±3	7.15±0.16	5.38±0.28
MISSFormer	42.46	17.62	52±3	8.24±0.20	6.15±0.35
UltraLight.	10.28	8.64	24±2	4.85±0.12	3.42±0.18
DCSSM	12.36	10.52	32±2	5.24±0.15	4.20±0.24

cessing stage to establish category-specific spatial centroids, incurring zero additional overhead during inference since all centroids are precomputed and stored in the category vector.

4.7. Robustness Evaluation on Challenging Cases

To validate the robustness of DCSSM under challenging scenarios, the proposed method was evaluated on generated extreme cases. Since naturally occurring extreme cases are rare and lack reliable ground truth annotations, a controlled perturbation approach was adopted, which is a common practice in medical image analysis robustness studies.

Four types of extreme test cases were generated from the original Synapse test set:

1. **Anatomical Position Perturbation (APP):** To simulate cases with abnormal organ positioning (e.g., situs inversus or post-surgical anatomical changes), we applied elastic deformations with large displacement fields to the original images. Specifically, we used a deformation field with control point spacing of 64 pixels and maximum displacement of 40-50 pixels, creating realistic anatomical variations while preserving organ integrity.
2. **Boundary Blurring (BB):** To simulate cases with indistinct organ boundaries caused by pathological conditions or imaging artifacts, we applied selective Gaussian blurring (σ = 2.0-4.0) along organ boundaries. The blurring kernel was applied within a 10-15 pixel band around ground truth boundaries.
3. **Intensity Inhomogeneity (II):** To simulate poor image quality or contrast agent distribution variations, we introduced multiplicative bias fields using a combination of low-frequency sinusoidal functions, creating intensity variations of ±30% across the image.
4. **Combined Perturbation (CP):** To evaluate performance under multiple simultaneous challenges, we combined all three perturbation types with moderate intensity (APP: 40 pixels, BB: σ=2.5, II: ±20%).

Table 10 presents the performance comparison under extreme conditions, where Δ denotes the performance degradation relative to the original test set. DCSSM demonstrates the smallest performance degradation across all perturbation types. Under the most challenging Combined Perturbation condition, DCSSM maintains 72.34% DSC with only 10.55% degradation, compared to 12.31-14.51% degradation for other methods. This enhanced robustness can be attributed to the CCMCM module's ability to maintain stable category-specific spatial priors even when organ appearances are distorted. By encoding the typical spatial distribution and appearance order of organs along the axial direction, CCMCM provides anatomical prior guidance that remains partially valid under perturbations, enabling the network to focus on anatomically plausible regions despite local distortions. Furthermore, the multi-stage fusion training strategy contributes to robustness by first establishing robust feature distributions through the Distribution Network before fine-tuning the Segmentation Network, allowing the model to learn more generalizable representations that are less sensitive to input variations. DCSSM shows particular resilience to Intensity Inhomogeneity (Δ = -2.66%), as the category center representations computed by CCMCM are based on spatial coordinates rather than intensity values, making them inherently invariant to intensity changes. The Lightweight Multiscale Feature Extraction Module also contributes to boundary robustness under Boundary Blurring conditions through its multi-scale design with 2:1:1 channel allocation, where the larger receptive fields (5×5 and 7×7 kernels) can capture contextual information to compensate for blurred local boundaries.

Despite overall robustness, specific applicability boundaries are identified. When anatomical displacement exceeds the spatial tolerance encoded by CCMCM's category vectors, the learned spatial priors become less effective, leading to increased vulnerability. Under such conditions where displacement exceeds 50 pixels, all methods including DCSSM exhibit rapid performance decline (>15% degradation). Additionally, small organs with high shape variability remain challenging under combined perturbations, as the limited pixel count provides fewer samples for reliable centroid estimation in CCMCM. Based on these findings, DCSSM should be applied with caution in patients with severe anatomical abnormalities such as situs inversus totalis, images with significant motion artifacts, or cases requiring precise small organ segmentation under suboptimal imaging conditions.

Table 10
Performance Comparison on Extreme Cases.

Method	Original DSC (%)	APP DSC (%)	BB DSC (%)	II DSC (%)	CP DSC (%)
U-Net	76.85	68.42 (-8.43)	71.23 (-5.62)	72.15 (-4.70)	62.34 (-14.51)
TransUNet	77.48	70.12 (-7.36)	72.56 (-4.92)	73.89 (-3.59)	64.78 (-12.70)
AttnUNet	75.57	66.89 (-8.68)	70.12 (-5.45)	71.34 (-4.23)	60.45 (-15.12)
TransDeepLab	80.16	72.45 (-7.71)	75.23 (-4.93)	76.89 (-3.27)	66.78 (-13.38)
Swin-Unet	79.13	71.34 (-7.79)	74.56 (-4.57)	75.78 (-3.35)	65.23 (-13.90)
MISSFormer	81.96	74.23 (-7.73)	76.45 (-5.51)	78.12 (-3.84)	68.56 (-13.40)
HiFormer	80.39	72.89 (-7.50)	75.67 (-4.72)	77.23 (-3.16)	67.12 (-13.27)
UltraLightUNet	82.43	75.67 (-6.76)	77.89 (-4.54)	79.34 (-3.09)	70.12 (-12.31)
DCSSM	82.89	77.45 (-5.44)	79.12 (-3.77)	80.23 (-2.66)	72.34 (-10.55)

5. Conclusion and Discussion

This paper proposes a Distribution-Constrained Semi-Supervised Segmentation Model (DCSSM) for medical organ segmentation. The model addresses the challenges of limited labeled data and insufficient utilization of spatial positional information in medical image segmentation through three key innovations: the Category Center of Mass Calculation Module (CCMCM) for capturing spatial distribution patterns, a multi-stage fusion training strategy for leveraging unlabeled data, and the Lightweight Multiscale Feature Extraction Module (LMFEM) for efficient feature extraction.

Extensive experiments on five benchmark datasets demonstrate the effectiveness of the proposed method. On the Synapse multi-organ CT dataset, DCSSM achieves 82.89% DSC and 18.20 HD, outperforming eight baseline methods. The model also shows strong generalization capability across different imaging modalities, achieving competitive results on the ACDC cardiac MRI dataset (90.24% DSC) and three dermoscopic image datasets (up to 94.79% DSC on PH^2). Furthermore, DCSSM maintains computational efficiency with only 12.36M parameters and 32ms inference time, making it suitable for clinical deployment.

Despite these promising results, certain limitations remain. The learned spatial priors become less effective when anatomical displacement exceeds the tolerance encoded by CCMCM, and small organs with high shape variability remain challenging under suboptimal imaging conditions. Future work will focus on several directions: (1) developing adaptive spatial prior mechanisms that can dynamically adjust to severe anatomical abnormalities such as situs inversus or post-surgical anatomical changes; (2) exploring more robust centroid estimation strategies for small organ segmentation, potentially incorporating uncertainty quantification to improve reliability; (3) extending the framework to 3D volumetric segmentation to better exploit inter-slice continuity; and (4) investigating the integration of foundation models such as SAM with the proposed distribution-constrained learning paradigm to further enhance generalization across diverse medical imaging tasks.

Acknowledgement

This research was funded by the Natural Science Foundation of Nanjing University of Chinese Medicine (XZR2021093) and the Suzhou City Applied Basic Research (Health) Science and Technology Innovation General Program (SYW2024051).

References

1. Abdollahi, A., Pradhan, B., Alamri, A. M. An Ensemble Architecture of Deep Convolutional Segnet and Unet Networks for Building Semantic Segmentation from High-Resolution Aerial Images. Geocarto International, 2022, 37(12), 3355-3370. https://doi.org/10.1080/10106049.2020.1856199
2. Ahmed, S., Hasan, M. K. Twin-SegNet: Dynamically Coupled Complementary Segmentation Networks for Generalized Medical Image Segmentation. Computer Vision and Image Understanding, 2024, 240, 103910. https://doi.org/10.1016/j.cviu.2023.103910
3. Archana, R., Jeevaraj, P. S. E. Deep Learning Models for Digital Image Processing: A Review. Artificial Intelligence Review, 2024, 57(1), 11. https://doi.org/10.1007/s10462-023-10631-z
4. Arumugam, L., Arumugam, S., Chidambaram, P., Govindasamy, K. A Multi-Model Deep Learning Approach for Human Emotion Recognition. Cognitive Neurodynamics, 2025, 19(1), 123-144. https://doi.org/10.1007/s11571-025-10304-3
5. Cao, G., Sun, Z., Wang, C., Geng, H., Fu, H., Yin, Z., Pan, M. RASNet: Renal Automatic Segmentation Using an Improved U-Net with Multi-Scale Perception and Attention Unit. Pattern Recognition, 2024, 150, 110336. https://doi.org/10.1016/j.patcog.2024.110336
6. Chen, B., Liu, Y., Zhang, Z., Lu, G., Kong, A. W. K. TransAttUnet: Multi-Level Attention-Guided U-Net with Transformer for Medical Image Segmentation. IEEE Transactions on Emerging Topics in Computational Intelligence, 2024, 8(1), 55-68. https://doi.org/10.1109/TETCI.2023.3309626
7. Chen, X., Wang, C., Ning, H., Li, S., Shen, M. SAM-OCTA: Prompting Segment-Anything for OCTA Image Segmentation. arXiv preprint arXiv:2310.07183, 2024. https://doi.org/10.1016/j.bspc.2025.107698
8. Fei, H., Wang, B., Wang, H., Fang, M., Wang, N., Ran, X., Liu, Y., Qi, M. MobileAmcT: A Lightweight Mobile Automatic Modulation Classification Transformer in Drone Communication Systems. Drones, 2024, 8(8), 357. https://doi.org/10.3390/drones8080357
9. Gai, D., Zhang, J., Xiao, Y., Min, W., Chen, H., Wang, Q., Su, P., Huang, Z. GL-Segnet: Global-Local Representation Learning Net for Medical Image Segmentation. Frontiers in Neuroscience, 2023, 17, 1153356. https://doi.org/10.3389/fnins.2023.1153356
10. He, K., Gkioxari, G., Dollár, P., Girshick, R. Mask R-CNN. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, October 22-29, 2017, IEEE, 2961-2969. https://doi.org/10.1109/ICCV.2017.322
11. He, S., Bao, R., Li, J., Grant, P. E., Ou, Y. Accuracy of Segment-Anything Model (SAM) in Medical Image Segmentation Tasks. arXiv preprint arXiv:2304.09324, 2023.
12. Houssein, E. H., Emam, M. M., Ali, A. A. An Efficient Multilevel Thresholding Segmentation Method for Thermography Breast Cancer Imaging Based on Improved Chimp Optimization Algorithm. Expert Systems with Applications, 2021, 185, 115651. https://doi.org/10.1016/j.eswa.2021.115651
13. Kayalibay, B., Jensen, G., van der Smagt, P. CNN-based Segmentation of Medical Imaging Data. arXiv preprint arXiv:1701.03056, 2017.
14. Kibriya, H., Abdullah, I., Kousar, F. Melanoma Lesion Segmentation and Classification Using SegNet. In: Proceedings of 2023 4th International Conference on Advancements in Computational Sciences (ICACS), Lahore, Pakistan, February 20-22, 2023, IEEE, 1-6. https://doi.org/10.1109/ICACS55311.2023.10089716
15. Kiernan, M. J., Al Mukaddim, R., Mitchell, C. C., Maybock, J., Wilbrand, S. M., Dempsey, R. J., Varghese, T. Lumen Segmentation Using a Mask R-CNN in Carotid Arteries with Stenotic Atherosclerotic Plaque. Ultrasonics, 2024, 137, 107193. https://doi.org/10.1016/j.ultras.2023.107193
16. Li, S. T., Zhang, L., Guo, P., Pan, H. Y., Chen, P. Z., Xie, H. F., Xie, B. K., Chen, J., Lai, Q. Q., Li, Y. Z., Wu, H., Wang, Y. Prostate Cancer of Magnetic Resonance Imaging Automatic Segmentation and Detection Based on 3D-Mask RCNN. Journal of Radiation Research and Applied Sciences, 2023, 16(3), 100636. https://doi.org/10.1016/j.jrras.2023.100636
17. Liu, F., Zheng, Q., Tian, X., Shu, F., Jiang, W., Wang, M., Elhanashi, A., Saponara, S. Rethinking the Multi-Scale Feature Hierarchy in Object Detection Transformer (DETR). Applied Soft Computing, 2025, 175, 113081. https://doi.org/10.1016/j.asoc.2025.113081
18. Lou, J., Yu, X., Ying, J., Song, D., Xiong, W. Exploring the Potential of Machine Learning and Magnetic Resonance Imaging in Early Stroke Diagnosis: A Bibliometric Analysis (2004-2023). Frontiers in Neurology, 2025, 16, 1505533. https://doi.org/10.3389/fneur.2025.1505533
19. Mazurowski, M. A., Dong, H., Gu, H., Yang, J., Konz, N., Zhang, Y. Segment Anything Model for Medical Image Analysis: An Experimental Study. Medical Image Analysis, 2023, 89, 102918. https://doi.org/10.1016/j.media.2023.102918

20. Petit, O., Thome, N., Rambour, C., Themyr, L., Collins, T., Soler, L. U-Net Transformer: Self and Cross Attention for Medical Image Segmentation. In: Lian, C. et al. (Eds.), Machine Learning in Medical Imaging: 12th International Workshop, MLMI 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, September 27, 2021, Proceedings, Lecture Notes in Computer Science, 12966, Springer, Cham, 2021, 267-276. https://doi.org/10.1007/978-3-030-87589-3_28
21. Roy, S., Wald, T., Koehler, G., Rokuss, M. R., Disch, N., Holzschuh, J., Zimmerer, D., Maier-Hein, K. H. SAM.MD: Zero-shot Medical Image Segmentation Capabilities of the Segment Anything Model. arXiv preprint arXiv:2304.05396, 2023. https://doi.org/10.48550/arXiv.2304.05396
22. Wang, H., Xie, S., Lin, L., Iwamoto, Y., Han, X. H., Chen, Y. W., Tong, R. Mixed Transformer U-Net for Medical Image Segmentation. In: Proceedings of ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2022, 2390-2394. https://doi.org/10.1109/ICASSP43922.2022.9747147
23. Wang, H., Ye, H., Xia, Y., Zhang, X. Leveraging SAM for Single-Source Domain Generalization in Medical Image Segmentation. arXiv preprint arXiv:2401.02076, 2024. https://doi.org/10.48550/arXiv.2401.02076
24. Wang, X. H., Fang, L. L. Survey of Image Segmentation Based on Active Contour Model. Pattern Recognition and Artificial Intelligence, 2013, 26(8), 751-760.
25. Wei, L., Liu, D., Shao, M., Wu, Y. Application of Improved Mask R-CNN Network in Medical Image Recognition and Segmentation. Journal of Computer Engineering & Applications, 2021, 57(24).
26. Wu, J., Ji, W., Liu, Y., Fu, H., Xu, M., Xu, Y., Jin, Y. Medical SAM Adapter: Adapting Segment Anything Model for Medical Image Segmentation. arXiv preprint arXiv:2304.12620, 2023. https://doi.org/10.48550/arXiv.2304.12620
27. Yao, Y., Chen, Y., Gou, S., Chen, S., Zhang, X., Tong, N. Auto-Segmentation of Pancreatic Tumor in Multi-Modal Image Using Transferred DSMask R-CNN Network. Biomedical Signal Processing and Control, 2023, 83, 104583. https://doi.org/10.1016/j.bspc.2023.104583
28. Yuan, F., Zhang, Z., Fang, Z. An Effective CNN and Transformer Complementary Network for Medical Image Segmentation. Pattern Recognition, 2023, 136, 109228. https://doi.org/10.1016/j.patcog.2022.109228
29. Yuan, Y., Wang, X., Yang, X., Heng, P. A. Effective Semi-Supervised Medical Image Segmentation with Probabilistic Representations and Prototype Learning. IEEE Transactions on Medical Imaging, 2025, 44(3), 1181-1193. https://doi.org/10.1109/TMI.2024.3484166
30. Zhang, K., Liu, D. Customized Segment Anything Model for Medical Image Segmentation. arXiv preprint arXiv:2304.13785, 2023. https://doi.org/10.48550/arXiv.2304.13785
31. Zhang, Y., Chen, J., Ma, X., Wang, G., Bhatti, U. A., Huang, M. Interactive Medical Image Annotation Using Improved Attention U-net with Compound Geodesic Distance. Expert Systems with Applications, 2024, 237(Part A), 121282. https://doi.org/10.1016/j.eswa.2023.121282
32. Zhao, X., Zhang, P., Song, F., Ma, C., Fan, G., Sun, Y., Feng, Y., Zhang, G. Prior Attention Network for Multi-Lesion Segmentation in Medical Images. IEEE Transactions on Medical Imaging, 2022, 41(12), 3812-3823. https://doi.org/10.1109/TMI.2022.3197180
33. Zheng, Q., Tian, X., Yu, Z., Ding, Y., Elhanashi, A., Saponara, S., Kpalma, K. MobileRaT: A Lightweight Radio Transformer Method for Automatic Modulation Classification in Drone Communication Systems. Drones, 2023, 7(10), 596. https://doi.org/10.3390/drones7100596
34. Zheng, Q., Saponara, S., Tian, X., Yu, Z., Elhanashi, A., Yu, R. A Real-Time Constellation Image Classification Method of Wireless Communication Signals Based on the Lightweight Network MobileViT. Cognitive Neurodynamics, 2024, 18(2), 659-671. https://doi.org/10.1007/s11571-023-10015-7
35. Zheng, Q., Tian, X., Yu, L., Elhanashi, A., Saponara, S., Vocaturo, E. Recent Advances in Automatic Modulation Classification Technology: Methods, Results, and Prospects. International Journal of Intelligent Systems, 2025, 2025, 4067323. https://doi.org/10.1155/int/4067323
36. Zheng, Q., Tian, X., Yang, M., Han, S., Elhanashi, A., Saponara, S., Kpalma, K. Reconstruction Error Based Implicit Regularization Method and Its Engineering Application to Lung Cancer Diagnosis. Engineering Applications of Artificial Intelligence, 2025, 139, 109439. https://doi.org/10.1016/j.engappai.2024.109439