DCFFSNET: DEEP CONNECTIVITY FEATURE FUSION SEPARATION NETWORK FOR MEDICAL IMAGE SEGMENTATION

Mingda Zhang[†‡], *Xun Ye*[†‡], *Ruixiang Tang*[‡], *Jingru Qiu*[‡], *Haiyan Ding*[*§]

[†] These authors have made equal contributions.
[‡]School of Software, Yunnan University, Kunming 650500, China
[§] School of Information Technology and Engineering, Yunnan University,Kunming 650500, China
{zhangmingda, tangruixiang, choujingru}@stu.ynu.edu.cn, ye.xun1@byd.com, dinghaiyan@ynu.edu.cn

ABSTRACT

Medical image segmentation leverages topological connectivity theory to enhance edge precision and regional consistency. However, existing deep networks integrating connectivity often forcibly inject it as an additional feature module, resulting in coupled feature spaces with no standardized mechanism to quantify different feature strengths. We propose DCFFSNet (Deep Connectivity Feature Fusion-Separation Network) with an innovative feature space decoupling strategy that quantifies the relative strength between connectivity features and other features, building a deep connectivity feature fusion-separation architecture that dynamically balances multi-scale feature expression. On ISIC2018, DSB2018 and MoNuSeg, DCFFSNet improves Dice/IoU over CMUNet, TransUNet and CSCAUNet by 1.3%/1.2%, 0.7%/0.9% and 0.8%/0.9%, respectively, indicating better edge precision and regional consistency.

Index Terms— Medical image segmentation, Topological connectivity, Feature space decoupling, Multi-scale fusion

1. INTRODUCTION

Medical image segmentation for pixel-level classification of organs or lesion regions poses significant technical challenges. While traditional methods relying on manually designed features can handle geometric deviations [1, 2], they underperform in segmenting complex structures with variable shapes and textures.

Connectivity in topology describes adjacent pixel interrelations [3], establishing spatial continuity constraints through adjacency relationships. This mathematical description addresses two key limitations: (1) mitigating edge blurring from insufficient gradient utilization by enhancing edge pixel correlations, and (2) improving spatial consistency through topological dependencies between regions. Connectivity-based techniques demonstrate superior capabilities in preserving internal continuity and optimizing edges [4].

However, existing connectivity-based networks often model connectivity as forced additional feature injection [4, 5, 6, 7]. This approach may not be optimal because feature maps contain limited feature information. While extensive enhancement of connectivity features achieves good results in connectivity feature extraction, it simultaneously affects the acquisition of other features. These networks lack standardized methods to measure feature strength in the feature space.

To address these issues, we: (1) decouple feature spaces by quantifying connectivity features relative to other features through standardized metrics, and (2) propose DCFFSNet based on feature space decoupling, adaptively balancing the relative strength between connectivity scales and multi-scale features to effectively resolve edge detail delineation challenges.

2. METHOD

2.1. DCFFSNet Architecture

DCFFSNet adopts a typical U-shaped encoder-decoder architecture with four key components: the backbone network, deeply supervised connectivity representation injection module (DSCRIM), multi-scale feature fusion module (MSFFM), multi-scale residual convolution module (MSRCM), and directional convolution (PConv). Recent advances in U-shaped architectures have demonstrated effectiveness in medical image segmentation [8, 9]. Figure 1 illustrates the complete architecture with three types of feature scales: feature scale (orange), connectivity scale (blue), and mixed scale (green).

2.2. Deeply Supervised Connectivity Representation Injection Module (DSCRIM)

DSCRIM is positioned at the bottleneck layer, where it injects connectivity features prominently into the feature space from the bottom layer through deep supervision to decouple connectivity features from classification features. Deep supervision has been shown to improve feature learning in medical image segmentation tasks [10]. This module significantly enhances the connectivity representation information of the

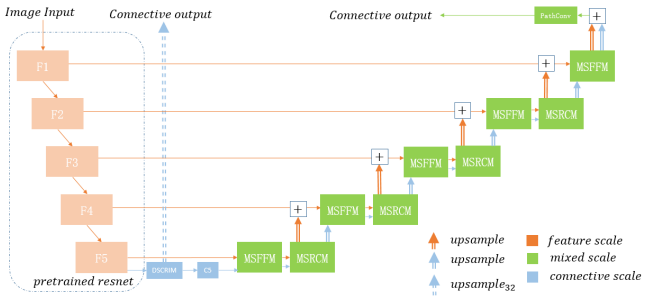

Fig. 1. DCFFSNet architecture showing the integration of DSCRIM at the bottleneck layer, MSFFM modules for feature fusion, MSRCM for multi-scale extraction, and PConv for final prediction.

feature map by employing rapid upsampling and connectivity grouping strategies, thereby capturing the connectivity scale features at the bottom layer.

Figure 2 illustrates the detailed structure of DSCRIM. Specifically, DSCRIM first performs multi-fold upsampling on the bottom-layer feature map to generate the corresponding deep supervision output and injects the connectivity representation prominently into the network. Subsequently, it learns connectivity features at the channel level through the connectivity grouping strategy, further optimizing feature representation.

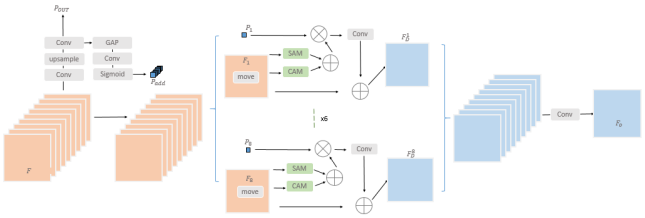

Fig. 2. Structure of DSCRIM with connectivity grouping strategy and attention mechanisms for feature decoupling.

First, the module takes the input bottom-layer feature map F_5, performs convolution and multi-fold upsampling (Upsample_32), and then applies a 1×1 convolution to obtain the output P_{OUT} of the original size. This output is used to inject prominent connectivity representations into the bottom layer through deep supervision, effectively capturing global connectivity information in the feature map. Next, global average pooling (GAP) is applied to P_{OUT}, and the number of channels is adjusted to match the original input via a 1×1 convolution. After passing through an activation function, a set of connectivity features P_{add} is generated to represent the connectivity relationships between channels.

$$P_{out} = \mathrm{Conv}_{1\times1}(\mathrm{Upsample}_{32}(\mathrm{Conv}_{1\times1}(F_5))) \quad (1)$$

P_{add} is divided into 8 groups, each corresponding to connectivity features P_i in different directions. The input feature map F_5 is also divided into 8 groups, and a shift operation is applied to each group to capture spatial dependencies in different directions, resulting in X_i.

Next, the module performs connectivity representation fusion for each group of features X_i and C_i through the following steps: (1) Pass X_i through the spatial attention module (SAM) and the channel attention module (CAM) to capture long-range spatial dependencies and inter-channel interaction information, respectively. Recent work has shown that combining spatial and channel attention improves segmentation performance [11]. (2) Add the outputs of spatial attention and channel attention to obtain a fused feature representation combining both spatial and channel attention. (3) Multiply the fused features with the connectivity representation P_i element-wise along the channel dimension to selectively enhance or suppress features with specific directional information. (4) Further optimize the processed features using a 1×1 convolution and add them to the original features to retain the original information, resulting in the intermediate feature F_D.

$$\begin{aligned} F_D^i = X_i + \mathrm{Conv}_{1\times1}(\mathrm{SpatialAtt}(X_i) \\ + \mathrm{ChannelAtt}(X_i) \odot P_i) \end{aligned} \quad (2)$$

Finally, the 8 groups of processed features are concatenated to restore the original dimensions, and a 1×1 convolution is applied to decode the feature information:

$$C_5 = \mathrm{Conv}_{1\times1}(\mathrm{cat}(F_D^1, \ldots, F_D^8)) \quad (3)$$

2.3. Multi-Scale Feature Fusion Module (MSFFM)

MSFFM fuses connectivity scales and feature scales of different dimensions to obtain the next dimension's connectivity scale features or hybrid features. Multi-scale feature fusion has been proven effective in handling organs of varying sizes [12]. It decouples the connectivity space and feature space through a self-attention mechanism. The module significantly enhances the spatial location information and global contextual dependencies of the connectivity scale feature map by fusing two types of inputs: feature scales and connectivity scales.

For feature-scale inputs, they are divided into eight groups, with the features in each group referred to as group features. The features undergo average pooling along both the horizontal and vertical directions. This process captures long-range dependencies in the horizontal dimension while retaining positional information in the vertical dimension. Subsequently, the pooled results from both directions are

concatenated and encoded into intermediate weights using a convolution operation W.

$$W = \mathrm{Conv}(\mathrm{cat}(\mathrm{Avg}_H(F_i), \mathrm{Avg}_W(F_i))) \tag{4}$$

Subsequently, crop W along the spatial direction, and reshape back to the original structure. Finally, a gating mechanism is used to generate the weight representation. After multiplying with the input F_i, the intermediate features at the feature scale are obtained through normalization. The use of gating mechanisms for feature selection has shown improvements in medical image segmentation [13].

$$X_F = \mathrm{BatchNormalize}(F_i \odot \mathrm{Sigmoid}(\mathrm{split}(W))) \tag{5}$$

For connectivity scale input C, divide it into 8 groups, each group is characterized by C_i. The difference is that here 3×3 Conv is used to get the intermediate feature of connectivity scale X_C:

$$X_C = \mathrm{BatchNormalize}(\mathrm{Conv}_{3\times3}(C_i)) \tag{6}$$

After obtaining the intermediate features between the connectivity scale and the feature scale X_C, X_F, learn about both across space. On features X_F, the 2D global average pooling layer is used to encode the global spatial information, and then Softmax get the normalized channel descriptor W_F. Multiply W_F with features X_C. That is, the weighted sum of all channel features at each location is obtained to obtain the global spatial attention representation in the connectivity scale W_{FC}.

$$\begin{aligned} W_F, W_{FC} =& \mathrm{reshape}(\mathrm{Softmax}(\mathrm{Avg}(X_F))), \\ & \mathrm{reshape}(W_F \bullet \mathrm{reshape}(X_C)) \end{aligned} \tag{7}$$

Similar to the above operation, processing X_C obtain global spatial attention representation on the feature scale W_{CF}. Finally, the two kinds of spatial attention are aggregated and the final weight representation W_a is obtained using the gating mechanism. The two inputs F and C are calibrated to get the outputs C_{next} and F_{next}.

$$\begin{aligned} C_{\text{next}} &= \mathrm{sigmoid}(W_{CF} + W_{FC}) \bullet C \\ F_{\text{next}} &= \mathrm{sigmoid}(W_{CF} + W_{FC}) \bullet F \end{aligned} \tag{8}$$

2.4. Multi-Scale Residual Convolution Module (MSRCM)

MSRCM performs multi-scale feature extraction in upsampling. Multi-scale processing with different kernel sizes enables better feature representation at various scales [14]. MSRCM extracts multi-scale features through convolution kernels of different sizes and alleviates the problem of gradient disappearance through residual connection:

$$\begin{aligned} Y = \mathrm{ReLU}(\mathrm{BN}(\mathrm{Conv}_{1\times1}(X + \mathrm{ReLU}(\\ \sum_{i=1,3,5,7} \mathrm{BN}(\mathrm{Conv}_{i\times i}(X)))))) \end{aligned} \tag{9}$$

2.5. Directional Convolution (PConv)

The directional convolution method PConv optimizes the segmentation effect of the connectivity mask by grouping and shifting, transforming the output from the highest layer into the final prediction. Figure 3 illustrates the structure of PConv. PConv groups and shifts the feature maps, applies a completely identical convolution kernel within each group, concatenates the results, and encodes them to achieve the interpretability of the connectivity mask.

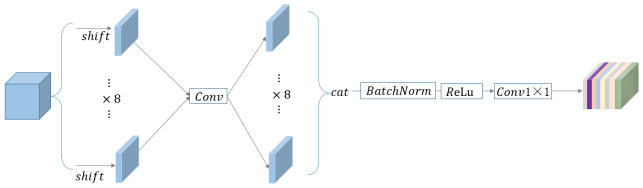

Fig. 3. Structure of Directional Convolution (PConv). The module divides input into eight groups, performs shift operations corresponding to connectivity directions, and applies convolutions for final prediction.

$$Y = \mathrm{Conv}_{1\times1}(\mathrm{ReLU}(\mathrm{BN}(\mathrm{cat}(\mathrm{Conv}_{3\times3}(\mathrm{shift}(X_i)))))) \tag{10}$$

PConv divides the input feature map X into eight groups, performs shift operations corresponding to the connectivity direction within each group, and applies a 3×3 convolution kernel for convolution. The results are then reassembled to their original size and decoded using a 1×1 convolution kernel to produce the final model prediction output.

3. EXPERIMENTS

3.1. Datasets and Implementation

Experiments were conducted on three datasets: ISIC2018 [15] , DSB2018 [16] , and MoNuSeg [17] . We used Dice Similarity Coefficient (DSC) and Intersection over Union (IoU) as evaluation metrics with five-fold cross-validation.

3.2. Comparative Results

Table 1 presents comprehensive comparative results against six state-of-the-art methods. DCFFSNet achieved the best performance across all datasets and metrics while maintaining computational efficiency.

DCFFSNet achieved 83.5±0.3 IoU and 90.0±0.2 Dice on ISIC2018, representing improvements of 1.2% and 1.3% over CMUNet. On DSB2018, it scored 85.4±0.1 IoU and 92.1±0.1 Dice, surpassing TransUNet by 0.9% and 0.7%. For MoNuSeg, it achieved 67.2±0.9 IoU and 79.7±0.9 Dice, exceeding CSCAUNet by 0.8% and 0.9%.

Table 1. Comparative experimental results of DCFFSNet

Model	Year	ISIC2018		DSB2018		MoNuSeg		FLOPs	Params
		IoU(%)	Dice(%)	IoU(%)	Dice(%)	IoU(%)	Dice(%)		
UNet [18]	2015	79.8±0.7	86.9±0.8	83.8±0.3	90.5±0.2	63.1±0.8	76.6±0.7	50.166G	34.527M
UNet++ [19]	2018	79.9±0.1	87.0±0.2	84.5±0.1	91.0±0.1	63.7±0.6	76.9±0.5	106.162G	36.630M
AttUNet [20]	2019	81.8±0.1	88.7±0.2	84.1±0.1	90.7±0.1	64.1±0.8	77.3±0.8	51.015G	34.879M
TransUNet [21]	2021	80.9±0.8	86.9±0.2	84.7±0.2	91.2±0.3	65.7±0.7	78.2±0.7	32.238G	93.231M
CMUNet [22]	2023	82.2±0.3	88.8±0.3	83.9±0.2	90.5±0.2	66.1±0.7	78.5±0.8	69.866G	49.932M
CSCAUNet [23]	2024	82.0±0.4	88.5±0.4	84.4±0.3	90.9±0.3	66.4±0.5	78.8±0.6	**10.517G**	35.275M
DCFFSNet	–	**83.5±0.3**	**90.0±0.2**	**85.4±0.1**	**92.1±0.1**	**67.2±0.9**	**79.7±0.9**	21.732G	52.717M

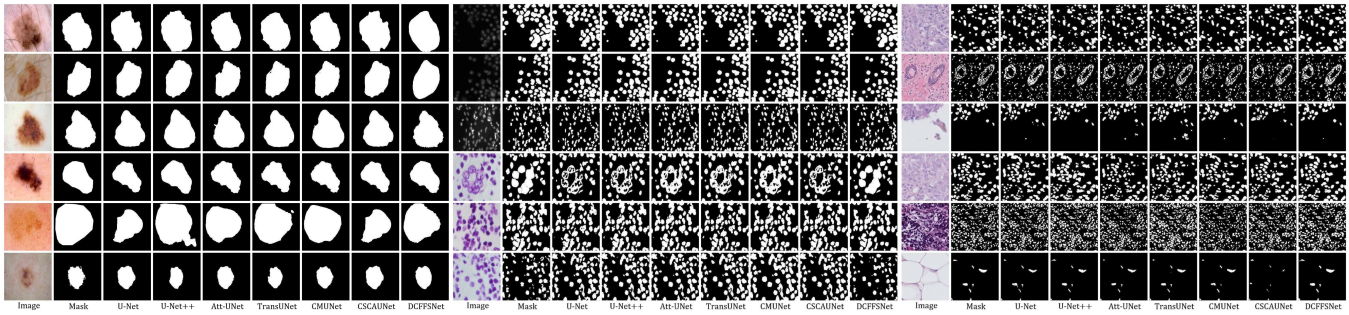

Fig. 4. Visualization comparison on ISIC2018 (left), DSB2018 (middle), and MoNuSeg (right) datasets. DCFFSNet achieves superior edge refinement, internal topology preservation, and overall segmentation quality compared to baseline methods.

Regarding computational efficiency, DCFFSNet requires 21.732 GFLOPs, significantly lower than UNet (50.166G), UNet++ (106.162G), and CMUNet (69.866G), while only marginally exceeding CSCAUNet (10.517G). With 52.717M parameters, DCFFSNet maintains a balance between model complexity and performance, demonstrating its suitability for deployment in computation-limited scenarios.

Figure 4 demonstrates DCFFSNet's superior performance in edge refinement and internal topology preservation across all three datasets. Unlike other models suffering from edge blurring, boundary misalignment, and internal discontinuities, DCFFSNet achieves precise edge localization with sharp contours highly consistent with ground-truth annotations while effectively maintaining intra-region connectivity and eliminating discontinuity artifacts.

3.3. Ablation Study

Table 2 shows ablation results validating the contribution of each module. Removing DSCRIM (w/o DS) caused the largest performance drop (2.4% Dice, 2.7% IoU on ISIC2018), confirming its crucial role in connectivity feature injection. Removing MSFFM (w/o MSF) led to 0.7% Dice and 1.1% IoU decrease, while replacing MSRCM (w/o MSR) resulted in 0.3% Dice and 0.5% IoU reduction.

The ablation visualizations indicate that each module contributes differently: DSCRIM exerts the largest effect by improving both overall feature learning and connectivity representation, while MSFFM notably enhances internal topology by fusing connectivity and feature scales. Together, these modules improve segmentation accuracy, sharpen boundaries, and preserve internal structure.

Table 2. Ablation study results

Model	ISIC2018		DSB2018		MoNuSeg	
	IoU(%)	Dice(%)	IoU(%)	Dice(%)	IoU(%)	Dice(%)
w/o DS	81.1±0.4	87.3±0.5	83.7±0.2	89.9±0.2	63.1±1.1	77.2±1.0
w/o MSF	82.8±0.1	88.9±0.1	84.8±0.1	91.1±0.1	64.2±0.7	78.3±0.7
w/o MSR	83.2±0.2	89.5±0.2	85.0±0.1	91.5±0.1	66.5±0.6	79.2±0.6
DCFFSNet	**83.5±0.3**	**90.0±0.2**	**85.4±0.1**	**92.1±0.1**	**67.2±0.9**	**79.7±0.9**

4. CONCLUSION

DCFFSNet addresses limitations of existing connectivity-based methods through systematic feature space decoupling.The method particularly excels in edge precision and regional consistency, offering significant advantages for clinical applications requiring accurate boundary delineation and internal structure preservation.

5. REFERENCES

[1] S. K. Nath, K. Palaniappan, and F. Bunyak, "Cell segmentation using coupled level sets and graph-vertex coloring," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Lecture Notes in Computer Science, vol. 4190, pp. 101–108, 2006.

[2] N. Kriegeskorte and R. Goebel, "An efficient algorithm for topologically correct segmentation of the cortical sheet in anatomical MR volumes," *NeuroImage*, vol. 14, no. 2, pp. 329–346, 2001.

[3] T. Y. Kong and A. Rosenfeld, "Digital topology: Introduction and survey," *Computer Vision, Graphics, and Image Processing*, vol. 48, no. 3, pp. 357–393, 1989.

[4] Z. Yang and S. Farsiu, "Directional connectivity-based segmentation of medical images," in *IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 11525–11535, 2023.

[5] W. Zhou, S. Dong, C. Xu, *et al.*, "Edge-aware guidance fusion network for RGB-thermal scene parsing," in *AAAI Conf. on Artificial Intelligence*, vol. 36, no. 3, pp. 3571–3579, 2022.

[6] M. Jian, R. Wu, W. Xu, *et al.*, "VascuConNet: An enhanced connectivity network for vascular segmentation," *Medical & Biological Engineering & Computing*, vol. 62, no. 11, pp. 3543–3554, 2024.

[7] Z. Qu, J. D. Wang, and X. H. Yin, "A directional connectivity feature enhancement network for pavement crack detection," *IEEE Transactions on Intelligent Transportation Systems*, 2024.

[8] C. Liu, K. Xu, L. L. Shen, *et al.*, "ImageFlowNet: Forecasting multiscale trajectories of disease progression with irregularly sampled longitudinal medical images," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2025.

[9] H. Huang, L. Lin, R. Tong, *et al.*, "UNet 3+: A full-scale connected UNet for medical image segmentation," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1055–1059, 2020.

[10] S. Li, Y. Zhang, and X. Chen, "Medical image segmentation via sparse coding decoder," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2025.

[11] J. Nam, J. Kim, H. Lee, *et al.*, "Modality-agnostic domain generalizable medical image segmentation by multi-frequency in multi-scale attention," in *IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 15234–15243, 2024.

[12] H. Cheng, Y. Wang, and Q. Liu, "Interactive medical image segmentation: A benchmark dataset and baseline," in *IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 21456–21465, 2025.

[13] Y. Liu, Z. Wang, and J. Chen, "Show and segment: Universal medical image segmentation via in-context learning," in *IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 32547–32556, 2025.

[14] X. Wang, L. Zhang, and Q. Li, "Adaptive bidirectional displacement for semi-supervised medical image segmentation," in *IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 14523–14532, 2024.

[15] N. Codella, V. Rotemberg, P. Tschandl, *et al.*, "Skin lesion analysis toward melanoma detection 2018," arXiv preprint arXiv:1902.03368, 2019.

[16] J. C. Caicedo, A. Goodman, K. W. Karhohs, *et al.*, "Nucleus segmentation across imaging experiments: The 2018 Data Science Bowl," *Nature Methods*, vol. 16, no. 12, pp. 1247–1253, 2019.

[17] N. Kumar, R. Verma, D. Anand, *et al.*, "A multi-organ nucleus segmentation challenge," *IEEE Transactions on Medical Imaging*, vol. 39, no. 5, pp. 1380–1391, 2019.

[18] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Lecture Notes in Computer Science, vol. 9351, pp. 234–241, 2015.

[19] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-Net architecture for medical image segmentation," in *Deep Learning in Medical Image Analysis*, pp. 3–11, 2018.

[20] O. Oktay, J. Schlemper, L. L. Folgoc, *et al.*, "Attention U-Net: Learning where to look for the pancreas," arXiv preprint arXiv:1804.03999, 2018.

[21] J. Chen, Y. Lu, Q. Yu, *et al.*, "TransUNet: Transformers make strong encoders for medical image segmentation," arXiv preprint arXiv:2102.04306, 2021.

[22] F. Tang, L. Wang, C. Ning, *et al.*, "CMU-Net: A strong ConvMixer-based medical ultrasound image segmentation network," in *Proc. IEEE Int. Symp. on Biomedical Imaging (ISBI)*, pp. 1–5, 2023.

[23] X. Shu, J. Wang, A. Zhang, *et al.*, "CSCA U-Net: A channel and space compound attention CNN for medical image segmentation," *Artificial Intelligence in Medicine*, vol. 150, p. 102800, 2024.