

DCFFSNet: Deep Connectivity Feature Fusion Separation Network for Medical Image Segmentation

Mingda Zhang^{a,1}, Xun Ye^{a,1}, Ruixiang Tang^a, Jingru Qiu^a, Haiyan Ding^b

^a*School of Software, Yunnan University, Kunming 650500, China*

^b*School of Information Technology and Engineering, Yunnan University, Kunming 650500, China*

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. To address these issues, we propose DCFFSNet (Dual-Connectivity Feature Fusion-Separation Network). It introduces an innovative feature space decoupling strategy. This strategy quantifies the relative strength between connectivity features and other features. It then builds a deep connectivity feature fusion-separation architecture. This architecture dynamically balances multi-scale feature expression. Experiments were conducted on the ISIC2018, DSB2018, and MoNuSeg datasets. On ISIC2018, DCFFSNet outperformed the next best model (CMUNet) by 1.3% (Dice) and 1.2% (IoU). On DSB2018, it surpassed TransUNet by 0.7% (Dice) and 0.9% (IoU). On MoNuSeg, it exceeded CSCAUNet by 0.8% (Dice) and 0.9% (IoU). The results demonstrate that DCFFSNet exceeds existing mainstream methods across all metrics. It effectively resolves segmentation fragmentation and achieves smooth edge transitions. This significantly enhances clinical usability.

Keywords: Medical image segmentation, Topological connectivity, Feature space decoupling, DCFFSNet, Multi-scale fusion

1. Introduction

Medical image segmentation aims to perform pixel-level classification of organs or lesion regions in medical images, a task that poses significant chal-

lenges in terms of technical requirements and time costs for physicians. Traditional methods primarily rely on manually designed feature extraction. While they can effectively handle geometric deviations [1, 2, 3], their performance in segmenting complex structures (such as organs or lesion regions with variable shapes and textures) is suboptimal.

In topology, connectivity describes how adjacent pixels are interrelated [4]. By defining the adjacency relationships between pixels, connectivity establishes spatial continuity constraints. This mathematical description effectively addresses two key shortcomings of traditional segmentation methods. First, by enhancing the modeling of correlations among edge pixels, it mitigates edge blurring caused by insufficient utilization of gradient information. Second, by establishing topological dependencies between regions, it improves the spatial consistency of segmentation results. Segmentation techniques based on connectivity models have achieved significant progress [5, 6, 7]. These connectivity-based techniques demonstrate stronger capabilities in handling images with complex structures, preserving internal structural continuity, reducing fragmentation in segmentation results, and optimizing segmentation edges [8]. Early applications of connectivity in deep learning primarily involved using traditional image segmentation methods for post-processing, such as applying the watershed algorithm [9] to refine the output of the DeepLab [10] model to enhance performance. These models did not truly integrate connectivity at a deeper level. However, with the advancement of connectivity technology, connectivity-based models have emerged. Traditional pixel-based segmentation mainly emphasizes categorical features, such as boundaries, whereas connectivity-based models exhibit superior capabilities in processing images with complex structures. In the field of deep learning for image segmentation, new applications leveraging connectivity continue to emerge.

Nevertheless, these connectivity-based networks often model connectivity as an additional feature injection and significantly enhance it within modules [8, 11, 12, 13]. 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 connectivity-based medical image segmentation networks lack a standardized method to measure the feature strength of different features in the feature space or to distinguish between them. These diverse features aggregate in the space, collectively forming the feature map.

Therefore, to address these issues:

1. This work decouples feature spaces by quantifying connectivity features relative to other features through standardized metrics.
2. We propose a Deep Connectivity Feature Fusion-Separation Network (DCFFSNet) based on feature space decoupling. By adaptively balancing the relative strength between connectivity scales and multi-scale features, it effectively resolves edge detail delineation challenges.

2. Related Work

2.1. *Connectivity Mask*

Traditional pixel-based segmentation methods primarily focus on categorical features, whereas connectivity-based models demonstrate superior performance in processing images with complex structures. After connectivity was widely adopted in classical image processing methods to describe topological properties [14, 15], it also found new applications in deep learning-based image segmentation [17, 18, 19, 20]. Connectivity-based segmentation networks employ connectivity masks as labels. Typically, these masks consist of 8 channels (for 2D images, corresponding to the x and y axes) or 26 channels (for 3D images, corresponding to the x, y, and z axes) [21], where each channel indicates whether a pixel in the original image belongs to the same class as its neighboring pixel in a specific direction.

A special case arises when converting standard masks to connectivity masks at image boundaries. For example, when performing upward connectivity operations on the top row of pixels, conventional methods treat non-existent neighboring positions as background (label 0). However, in this paper, such positions are assigned the same label as the pixel itself. This approach ensures that all eight channels in the connectivity mask’s boundary regions retain their classification labels, thereby imposing stronger constraints on segmentation results near edges (where, under the original computation, at least three channels would be set to background).

2.2. *Bilateral Voting (BV) Module and Region-guided Channel Aggregation (RCA)*

After obtaining the connectivity mask, since the validation process requires converting the connectivity mask back to a general image mask, Yang et al. [16] were the first to propose a bilateral voting mechanism and a

region-guided channel aggregation method to address the conversion between connectivity masks and general image masks.

The connectivity mask enhances consistency between adjacent pixels through the bilateral voting module. Because the connectivity mask only multiplies adjacent pixels of the same class—meaning the values of the two corresponding pixels in the connectivity mask should be closely related—the values in the corresponding two channels must belong to the same category to indicate consistent classification results. In bilateral voting, each pair of elements in the connectivity output X^C is multiplied, and the result is assigned to these two elements. The resulting bilateral voting output X^B , referred to as the Bicon map, is treated as the connectivity mask Y^C in the final output.

Each channel in the Bicon map displays the bilateral class probability distribution in a specific direction, reflecting the categorical relationship between a pixel and its neighboring pixels. The region-guided channel aggregation (RCA) method converts the Bicon map into a single-channel output probability through a mapping function.

$$X_i^B \text{Mask}(x, y) = \{f\{X_{9-i}^B(x, y)\}\}_{i=1}^8 \quad (1)$$

Here, f can be any aggregation function. In this paper, we employ the max operation for channel aggregation, selecting the highest probability from the eight channels as the final predicted probability distribution.

2.3. Attention Mechanism

The Attention Mechanism [22] simulates the way humans focus on different regions with varying levels of attention and has been widely adopted in computer vision.

In image segmentation, the two most common and fundamental attention mechanisms are channel attention [23] and spatial attention [24]. In the channel attention mechanism, w assigns different weights to each channel by capturing the feature strength and importance within the channels. In spatial attention, w captures the feature strength and importance at different coordinates in the image or feature map, assigning distinct weights to each location.

3. Method

3.1. DCFFSNet Model

This paper proposes a DCFFSNet based on connectivity decoupling. The DCFFSNet adopts a typical U-shaped encoder-decoder architecture, and Figure 1 illustrates the structure of DCFFSNet.

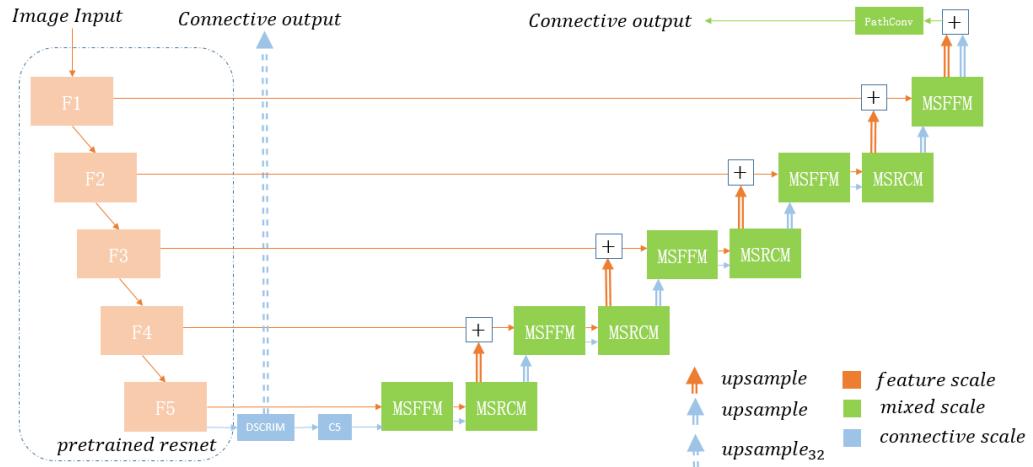


Figure 1: Structure of the proposed DCFFSNet model, which consists of four components: the backbone network, the deeply supervised connectivity representation injection module, the multi-scale residual convolution module, and the directional convolution module.

3.2. Deeply Supervised Connectivity Representation Injection Module (DSCRIM)

The 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. This module significantly enhances the connectivity representation information of the 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 the DSCRIM. Specifically, the 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.

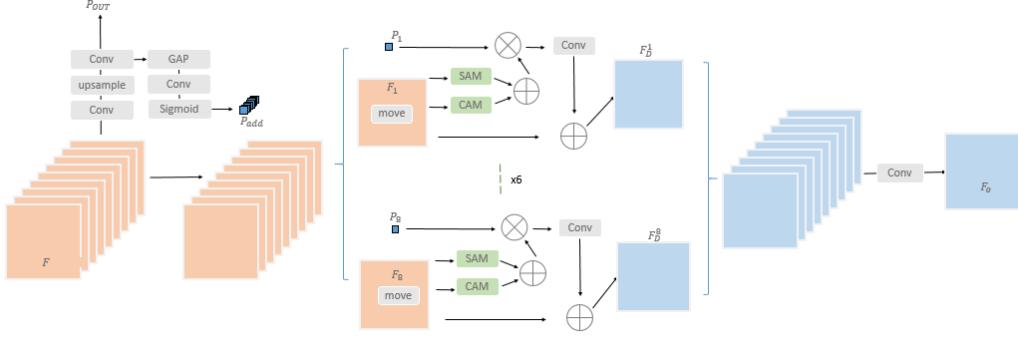


Figure 2: Structure of DSCRIM

First, the module takes the input bottom-layer feature map F_5 , performs convolution and multi-fold upsampling (Upsample₃₂), 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.

$$P_{out} = \text{Conv}_{1 \times 1}(\text{Upsample}_{32}(\text{Conv}_{1 \times 1}(F_5))) \quad (2)$$

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.
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 .

$$F_D^i = X_i + \text{Conv}_{1 \times 1}(\text{SpatialAtt}(X_i) + \text{ChannelAtt}(X_i) \odot P_i) \quad (3)$$

Here, \odot denotes element-wise multiplication, *SpatialAtt* and *ChannelAtt* represent spatial attention and channel attention, respectively.

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, yielding the final bottom-layer connectivity feature output C_5 .

$$C_5 = \text{Conv}_{1 \times 1}(\text{cat}(F_D^1, \dots, F_D^8)) \quad (4)$$

3.3. Multi-Scale Feature Fusion Module (MSFFM)

The MSFFM fuses connectivity scales and feature scales of different dimensions to obtain the next dimension’s connectivity scale features or hybrid features. It decouples the connectivity space and feature space through a self-attention mechanism.

Figure 3 illustrates the structure of the MSFFM. 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.

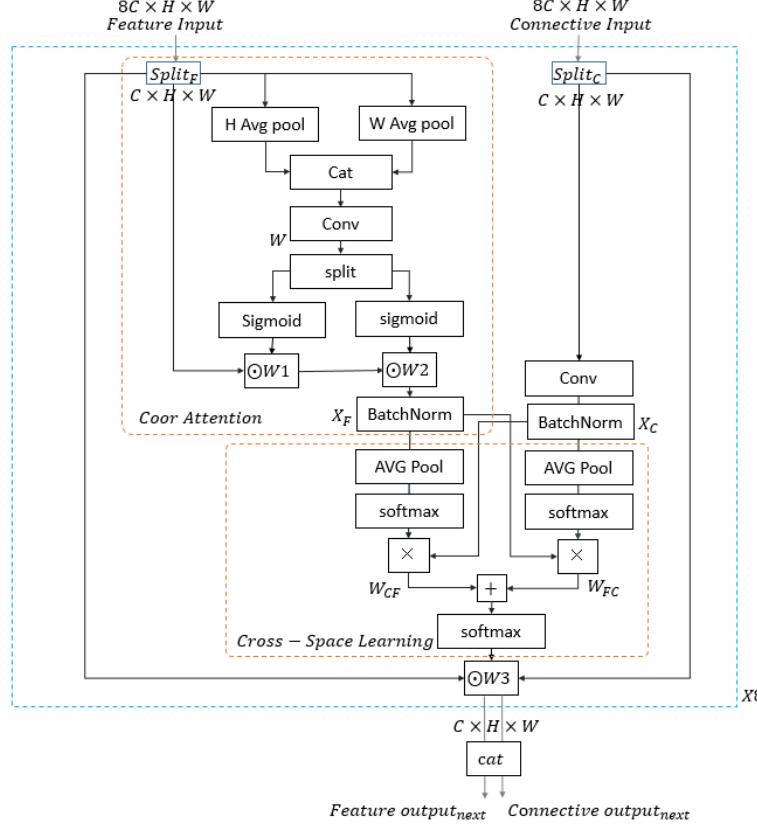


Figure 3: Structure of MSFFM

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 .

$$W = \text{Conv}(\text{cat}(\text{Avg}_H(F_i), \text{Avg}_W(F_i))) \quad (5)$$

Subsequently, crop W along the spatial direction, and reshaped back to the original structure (the horizontal and vertical directions). 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.

$$X_F = \text{BatchNormalize}(F_i \odot \text{Sigmoid}(\text{split}(W))) \quad (6)$$

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 :

$$X_C = \text{BatchNormalize}(\text{Conv}_{3 \times 3}(C_i)) \quad (7)$$

After obtaining the intermediate features between the connectivity scale and the feature scale X_C, X_F , learn about both across space. The specific operation is as follows:

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} .

$$W_F, W_{FC} = \text{reshape}(\text{Softmax}(\text{Avg}(X_F))), \text{reshape}(W_F \bullet \text{reshape}(X_C)) \quad (8)$$

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} .

$$C_{next}, F_{next} = \text{sigmoid}(W_{CF} + W_{FC}) \bullet (C, F) \quad (9)$$

Since the above operations are performed in all groups, the resulting output feature C_{next} still has the original grouping structure of the connectivity scale and can be invoked repeatedly in multiple layers.

In particular, in the last layer of the model, the model uses the final output of a mixture of both.

3.4. Multi-Scale Residual Convolution Module (MSRCM)

MSRCM performs multi-scale feature extraction in up-sampling.

Figure 4 shows the structure of MSRCM. MSRCM extracts multi-scale features through convolution kernels of different sizes and alleviates the problem of gradient disappearance through residual connection, thus improving the performance of the model.

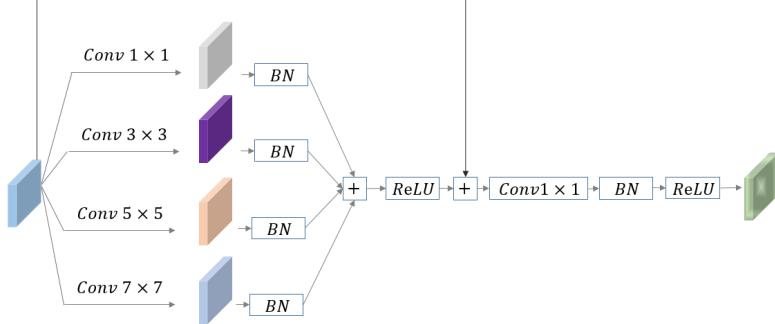


Figure 4: Structure of MSRCM

$$Y = \text{ReLU}(BN(\text{Conv}_{1 \times 1}(X + \text{ReLU}(\sum_{i=1,3,5,7} BN(\text{Conv}_{i \times i}(X)))))) \quad (10)$$

MSRCM uses four convolutional kernels of different sizes to extract input features X to obtain multi-scale feature information. Then, these features are normalized by BN and added together for fusion. After ReLU operation, they are residual, and then re-encoded with a 1×1 convolutional kernel to obtain output Y .

3.5. Directional Convolution (PConv)

The directional convolution method PConv (Path Convolution) optimizes the segmentation effect of the connectivity mask by grouping and shifting, transforming the output from the highest layer into the final prediction. Figure 5 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.

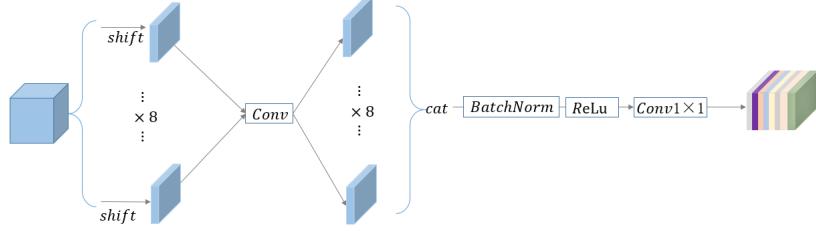


Figure 5: Structure of Directional Convolution (PConv)

$$Y = \text{Conv}_{1 \times 1}(\text{ReLU}(\text{BN}(\text{cat}(\text{Conv}_{3 \times 3}(\text{shift}(X_i)))))) \quad (11)$$

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.

3.6. Loss Function

This paper considers the connectivity output X^C as the core output of the model. In Section 2.1, it is described that the connectivity mask can be derived from the regular mask X through a simple multiplication operation. In Section 2.2, the connectivity output X^C is obtained through bilateral voting to generate the dual connectivity mask X^B , which is then transformed into the standard mask X via RCA.

Therefore, considering the above different masks, the loss function of the model is defined as follows:

$$L = L_{\text{MainBCE}} + 0.2 * L_{\text{BBCE}} + 0.8 * L_{\text{CBCE}} \quad (12)$$

$$L_{\text{MainBCE}} = L_{\text{BCE}}(X, Y) = X * Y + (1 - X) * (1 - Y) \quad (13)$$

$$L_{\text{BBCE}} = L_{\text{BCE}}(X^B, Y^C) = X^B * Y^C + (1 - X^B) * (1 - Y^C) \quad (14)$$

$$L_{\text{CBCE}} = L_{\text{BCE}}(X^C, Y^C) = X^C * Y^C + (1 - X^C) * (1 - Y^C) \quad (15)$$

Among them, the superscript B refers to the dual pass mask, C refers to the connectivity mask, and the superscript without refers to the standard mask.

$$Loss = 0.2 * L(output1) + L(output2) \quad (16)$$

The model proposed in this paper has two outputs, where *output1* is the intermediate output of deep supervised connectivity representation injection, and *output2* is the final output of the model.

4. Experiments

4.1. Dataset

In order to verify the performance of the model, we used three different datasets for experiments, including ISIC2018 dataset [25], DSB2018 dataset [26] and MoNuSeg dataset [27].

4.2. Comparative Experiments

In order to verify the effectiveness of the proposed DCFFSNet network, we selected more classical medical image segmentation networks U-Net [28], UNet++ [29], AttUNet [30], TransUNet [31], CMUNet [32], CSCAUNet [33] for comparison. In this paper, the Dice Similarity Coefficient (DSC) and Intersection over Union (IoU) were selected as the core quantitative evaluation metrics.

Table 1 shows the model performance of DCFFSNet under different datasets, where \pm represents the standard deviation in a five-fold cross-validation environment, which reflects the stability of the model to some extent. The experimental results show that our model is the best.

Table 1: Comparative experimental results of DCFFSNet

| Model | Year | ISIC2018 | | DSB2018 | | MoNuSeg | | FLOPs | Params |
|-----------------|------|----------------------------------|----------------------------------|----------------------------------|----------------------------------|----------------------------------|----------------------------------|----------------|----------------|
| | | IoU(%) | Dice(%) | IoU(%) | Dice(%) | IoU(%) | Dice(%) | | |
| UNet [28] | 2015 | 79.8 \pm 0.7 | 86.9 \pm 0.8 | 83.8 \pm 0.3 | 90.5 \pm 0.2 | 63.1 \pm 0.8 | 76.6 \pm 0.7 | 50.166G | 34.527M |
| UNet++ [29] | 2018 | 79.9 \pm 0.1 | 87.0 \pm 0.2 | 84.5 \pm 0.1 | 91.0 \pm 0.1 | 63.7 \pm 0.6 | 76.9 \pm 0.5 | 106.162G | 36.630M |
| AttUNet [30] | 2019 | 81.8 \pm 0.1 | 88.7 \pm 0.2 | 84.1 \pm 0.1 | 90.7 \pm 0.1 | 64.1 \pm 0.8 | 77.3 \pm 0.8 | 51.015G | 34.879M |
| TransUNet [31] | 2021 | 80.9 \pm 0.8 | 86.9 \pm 0.2 | 84.7 \pm 0.2 | 91.2 \pm 0.3 | 65.7 \pm 0.7 | 78.2 \pm 0.7 | 32.238G | 93.231M |
| CMUNet [32] | 2023 | 82.2 \pm 0.3 | 88.8 \pm 0.3 | 83.9 \pm 0.2 | 90.5 \pm 0.2 | 66.1 \pm 0.7 | 78.5 \pm 0.8 | 69.866G | 49.932M |
| CSCAUNet [33] | 2024 | 82.0 \pm 0.4 | 88.5 \pm 0.4 | 84.4 \pm 0.3 | 90.9 \pm 0.3 | 66.4 \pm 0.5 | 78.8 \pm 0.6 | 10.517G | 35.275M |
| DCFFSNet (Ours) | - | 83.5 \pm 0.3 | 90.0 \pm 0.2 | 85.4 \pm 0.1 | 92.1 \pm 0.1 | 67.2 \pm 0.9 | 79.7 \pm 0.9 | 21.732G | 52.717M |

By analyzing and comparing the experimental metrics, the DCFFSNet network proposed in this paper achieved the best Dice coefficient and IoU scores across three datasets. In the ISIC2018 dataset, the model scored 83.5 ± 0.3 in IoU and 90.0 ± 0.2 in Dice coefficient, which are 1.3% and 1.2%

higher than the second-best CMUNet, respectively. In the DSB2018 dataset, the model scored 85.4% in IoU and 92.1% in Dice coefficient, which are 0.9% and 0.7% higher than the second-best TransUNet. In the MoNuSeg dataset, the model scored 67.2% in IoU and 79.7% in Dice coefficient, which are 0.8% and 0.9% higher than the second-best CSCAUNet.

DCFFSNet achieves an effective balance between computational efficiency and model performance. With a computational load of only 21.732 GFLOPs, it performs significantly lower than models like U-Net, UNet++, and CMUNet, while marginally exceeding CSCAUNet. This indicates reduced computational resource requirements during inference, enabling more efficient segmentation tasks—particularly suitable for deployment in computation-limited scenarios. Regarding parameters (52.717M), although DCFFSNet doesn't attain the lowest count, it delivers high segmentation performance without substantially increasing parametric burden. This demonstrates its optimal balance between model complexity and effectiveness. Collectively, while DCFFSNet doesn't dominate in absolute computational/parametric metrics, its lower computational load enhances deployment feasibility. The architecture demonstrates promise as an efficient segmentation model by maintaining favorable performance-efficiency trade-offs.

The visualization results in Figures 6, 7, and 8 demonstrate that DCFFSNet significantly outperforms comparative models including UNet and UNet++ in edge refinement, internal topology preservation, and overall segmentation quality. Unlike other models that frequently suffer 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, resulting in optimal morphological coherence and detail representation. By combining superior segmentation accuracy with operational efficiency through low computational load and moderate parameters, DCFFSNet establishes the best performance-efficiency balance among all evaluated models, highlighting its strong potential as an efficient segmentation solution and demonstrating significant practical value for flexible deployment in real-world medical image analysis applications.

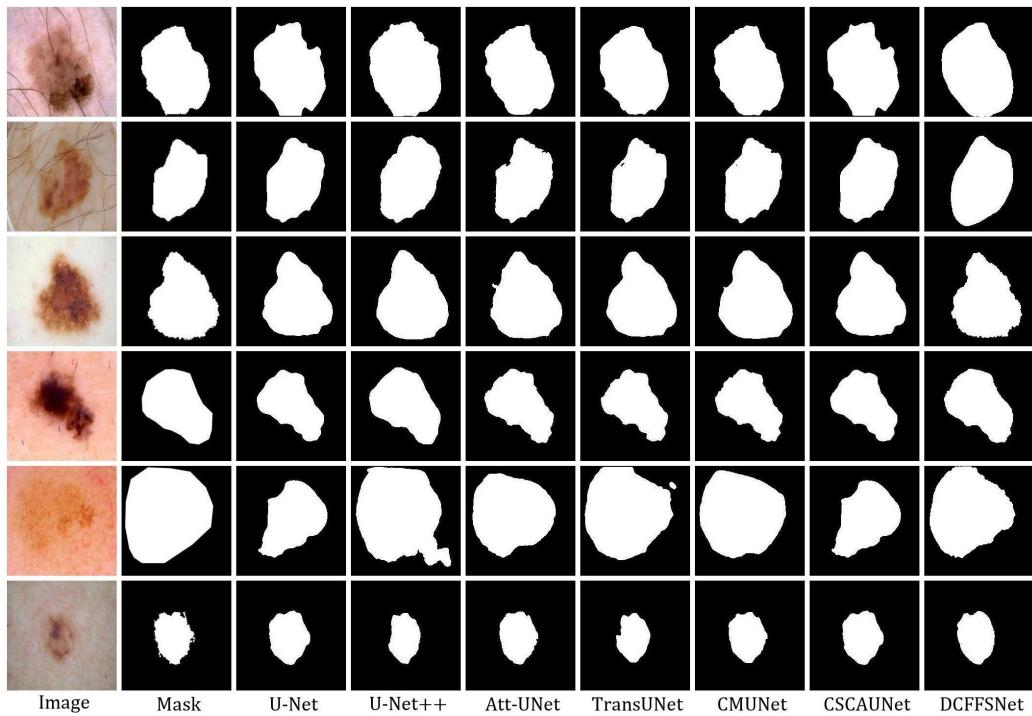


Figure 6: Visualization reference of image comparison experiment of ISIC2018 dataset

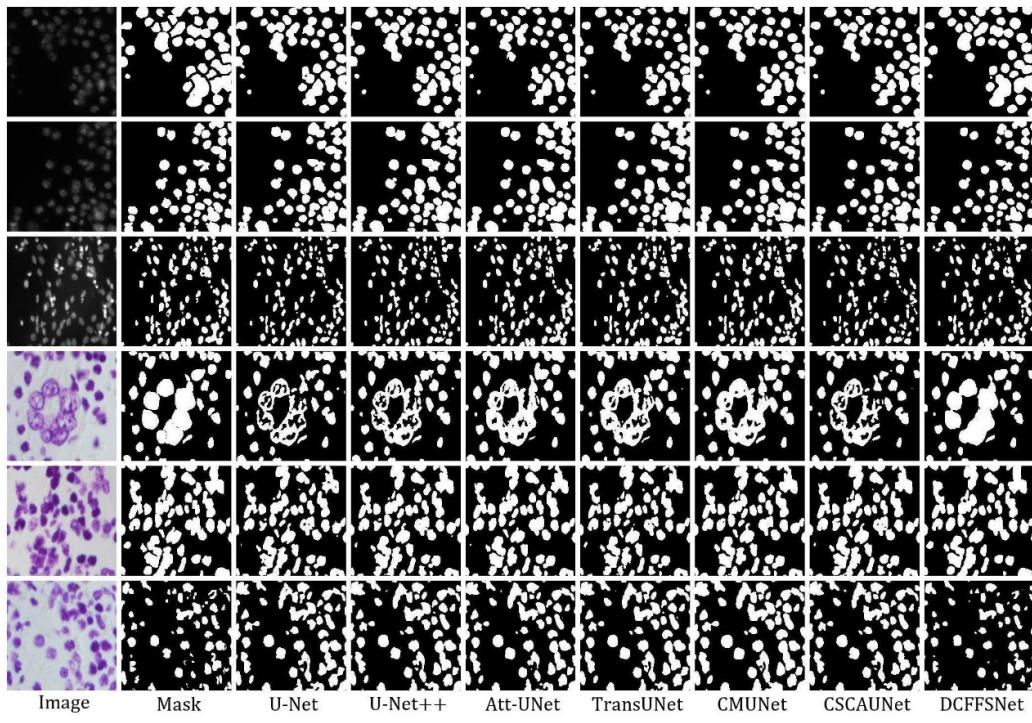


Figure 7: Visualization reference of DSB2018 dataset comparison

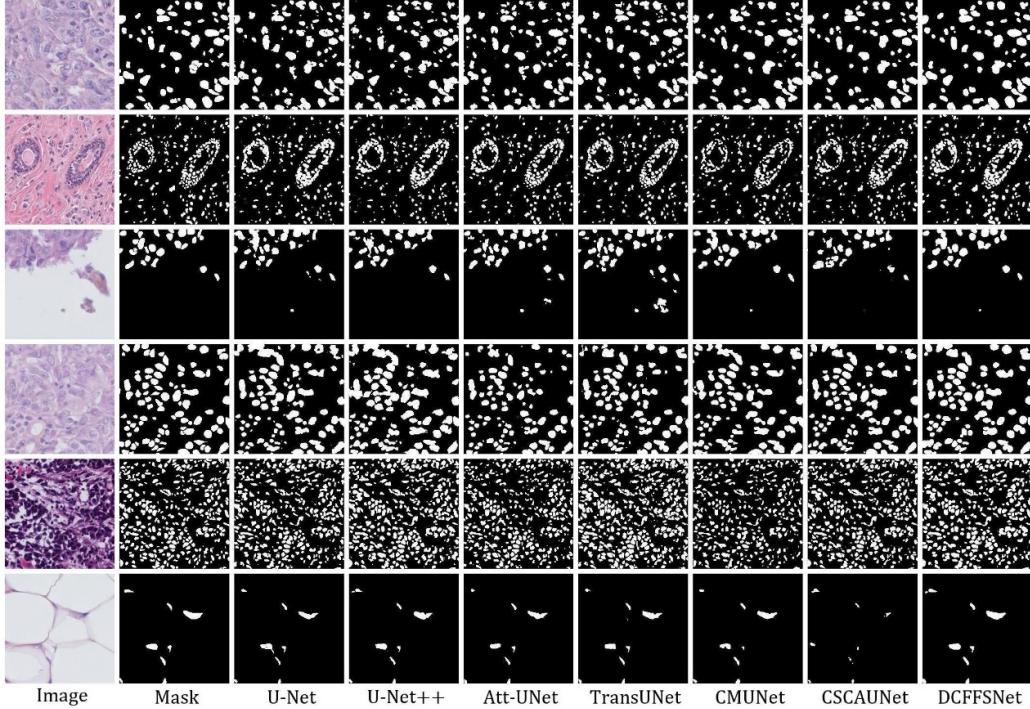


Figure 8: Visual reference of image comparison experiment on MoNuSeg dataset

4.3. Ablation Experiments

In order to verify the effectiveness of the proposed method and strictly control the variables, this section still selects ISIC2018 dataset, DSB2018 dataset and MoNuSeg dataset for ablation experiments.

Table 2 shows the results of the ablation experiment. In this paper, the ablation experiment is carried out by removing or replacing modules. The specific network structure design is as follows:

- DCFFSNet: Deep connectivity feature fusion and separation network proposed in this chapter.
- w/o DS: The deep supervision connectivity representation injection module DSCRIM is removed, and a copy of the feature map F_5 output by the main network is directly copied to the multi-scale feature fusion module.
- w/o MSF: The multi-scale feature fusion module MSFFM is removed, and the transmitted features are directly spliced with the upsampling

features in a way of addition (to ensure the consistency of channel numbers).

- w/o MSR: The multi-scale residual convolution module MSRCM is replaced with the original residual structure.

Table 2: Experimental results of DCFFSNet ablation

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

The DSCRIM, MSFFM, and MSR modules designed in this section have all positively impacted the network. For instance, in the ISIC dataset, removing the DSCRIM module resulted in a 2.4% decrease in the Dice coefficient score and a 2.7% decrease in the IoU score of DCFFSNet. Removing the MSFFM module led to a 0.7% decrease in the Dice coefficient score and a 1.1% decrease in the IoU score. Replacing the MSR module resulted in a 0.3% decrease in the Dice coefficient score and a 0.5% decrease in the IoU score.

Figures 9, 10 and 11 show the visual results of ablation experiments.

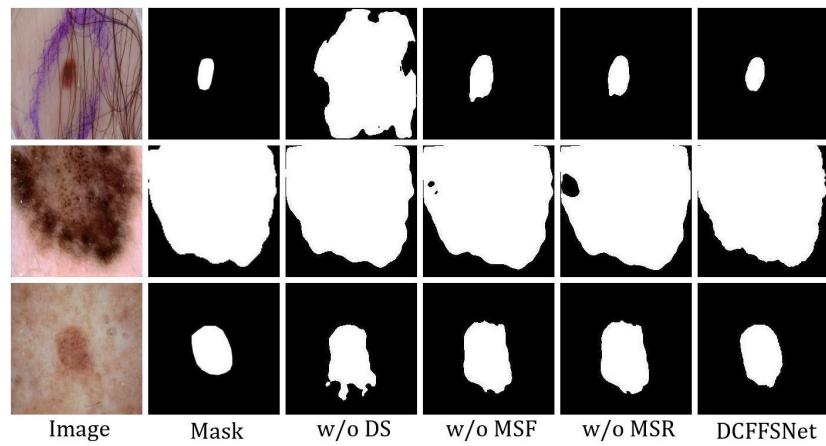


Figure 9: Ablation experiment results of ISIC2018 dataset

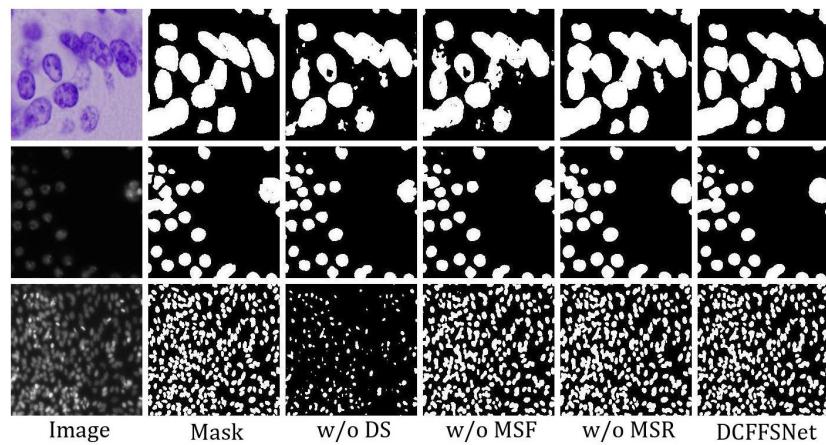


Figure 10: Ablation experiment results of DSB2018 dataset

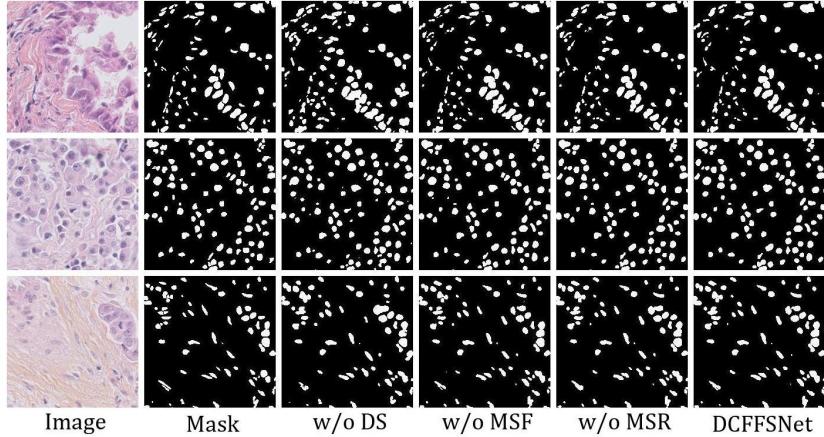


Figure 11: Results of MoNuSeg dataset ablation experiment

The visualization results from the ablation experiments clearly show that the collaborative effects of various modules have different impacts on the model proposed in this chapter. The deep supervision connectivity representation injection module significantly influences the final model results, while DSCRIM has a substantial impact on both overall feature acquisition and connectivity acquisition. The multi-scale feature fusion module (MSFFM) has a significant effect on the internal topology. By integrating connectivity and feature scales, removing it would focus more on feature scale information, leading to the destruction of both internal and edge topologies. After the collaborative effects of multiple modules, the model proposed in this chapter not only ensures segmentation performance but also accurately determines boundary positions while maintaining internal structural consistency, ultimately achieving effective segmentation of medical images.

In addition to the ablation experiments mentioned above, corresponding experiments were also carried out on the weight allocation of the loss function used in this paper. Specifically, the weight of $L(\text{output1})$ was adjusted, and the results are shown in Table 3. In which $w = 0$ means that the weight is cancelled, that is, the deep supervision method is no longer used to inject connectivity scale features.

Table 3: Weight ablation results of $L(\text{output1})$

| Dataset | Evaluating W | IoU(%) | Dice(%) |
|----------|--------------|----------------------------------|----------------------------------|
| ISIC2018 | 0 | 81.7 ± 0.2 | 88.1 ± 0.2 |
| | 0.1 | 82.7 ± 0.3 | 88.8 ± 0.3 |
| | 0.2 | 83.5 ± 0.3 | 90.0 ± 0.2 |
| | 0.3 | 83.1 ± 0.2 | 89.5 ± 0.3 |
| DSB2018 | 0 | 84.9 ± 0.1 | 91.5 ± 0.1 |
| | 0.1 | 85.2 ± 0.1 | 91.9 ± 0.1 |
| | 0.2 | 85.4 ± 0.1 | 92.1 ± 0.1 |
| | 0.3 | 85.1 ± 0.1 | 91.7 ± 0.1 |
| MoNuSeg | 0 | 64.8 ± 0.8 | 78.0 ± 0.8 |
| | 0.1 | 65.7 ± 0.9 | 78.5 ± 0.9 |
| | 0.2 | 67.2 ± 0.9 | 79.7 ± 0.9 |
| | 0.3 | 66.2 ± 0.9 | 78.7 ± 0.9 |

The results show that the model performs best when the weight of $L(\text{output1})$ is 0.2. However, after the use of deep supervision to inject connectivity scale features is cancelled, the performance of the model decreases significantly, which also reflects the important role played by DSCRIM module in the model to some extent.

5. Discussion

The proposed algorithm excels in preserving internal topological structures and smoothing edge regions, effectively capturing target details and generating smooth segmentation boundaries; however, it demonstrates limited overall metric improvements (e.g., Dice coefficient/IoU) compared to state-of-the-art methods, likely due to insufficient global consistency focus while optimizing local details and potential slight accuracy reduction from connectivity constraints despite enhanced topological integrity. Future research should prioritize balanced local-global optimization through enhanced consistency mechanisms, develop parameterized strategies for dynamically adapting connectivity strength across scenarios, and integrate connectivity with advanced techniques like attention mechanisms or multi-scale feature fusion to comprehensively boost performance.

6. Data Statement

The ISIC2018 (International Skin Imaging Collaboration 2018) dataset [25], released by the International Skin Imaging Collaboration, is designed for skin lesion segmentation. It comprises 2,594 dermoscopic images with corresponding annotations. Image resolutions range from 6748×4499 pixels to 718×542 pixels. The DSB2018 (Data Science Bowl 2018) dataset [26], published via the Kaggle platform, was created for a nucleus segmentation competition. It contains 670 microscopy images (stage1_train subset) with nucleus annotations. Image resolutions vary between 256×256 pixels and 1024×1024 pixels. The MoNuSeg dataset [27], released by the U.S. National Institutes of Health (NIH), specializes in nucleus segmentation of histopathology images. It consists of 44 high-resolution images (30 training + 14 test) at 1000×1000 pixel resolution, each with corresponding nucleus annotations.

7. Conflict of Interest

The authors declare no conflict of interest.

8. Acknowledgments

We extend our sincere gratitude to: the International Skin Imaging Collaboration (ISIC) for the ISIC 2018 dataset (2,594 dermoscopic images, resolution: 542×718 to $4,499 \times 6,748$ pixels), whose precise lesion annotations served as a crucial benchmark for validating cross-scale feature fusion algorithms in our skin lesion segmentation model; Kaggle for the DSB2018 dataset (670 microscopy images, resolution: 256×256 to $1,024 \times 1,024$ pixels), whose densely distributed nuclei and complex backgrounds optimized tiny target detection capabilities in our nuclear instance segmentation framework; and the U.S. National Institutes of Health (NIH) for the MoNuSeg dataset (44 histopathology images at $1,000 \times 1000$ resolution), whose high-accuracy nuclear boundary annotations played a pivotal role in developing the boundary-aware loss function and enhancing contour segmentation precision—we hereby express our profound appreciation to all institutions advancing open science.

References

- [1] Nath SK, Palaniappan K, Bunyak F. Cell segmentation using coupled level sets and graph-vertex coloring. In: Larsen R, Nielsen M, Sporrings J, editors. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2006. MICCAI 2006. Lecture Notes in Computer Science, vol 4190. Berlin, Heidelberg: Springer; 2006. p. 101-108. https://doi.org/10.1007/11866565_13.
- [2] Kriegeskorte N, Goebel R. An efficient algorithm for topologically correct segmentation of the cortical sheet in anatomical MR volumes. NeuroImage 2001;14(2):329-346. <https://doi.org/10.1006/nimg.2001.0831>.
- [3] Liu M, Colas F, Siegwart R. Regional topological segmentation based on mutual information graphs. IEEE International Conference on Robotics and Automation 2011:3269-3274. <https://doi.org/10.1109/ICRA.2011.5979672>.
- [4] Kong TY, Rosenfeld A. Digital topology: Introduction and survey. Computer Vision, Graphics, and Image Processing 1989;48(3):357-393. [https://doi.org/10.1016/0734-189X\(89\)90147-3](https://doi.org/10.1016/0734-189X(89)90147-3).
- [5] Kampffmeyer M, Dong N, Liang X, et al. ConnNet: A long-range relation-aware pixel-connectivity network for salient segmentation. IEEE Transactions on Image Processing 2018;28(5):2518-2529. <https://doi.org/10.1109/TIP.2018.2886997>.
- [6] Fu J, Liu J, Tian H, Li Y, Bao Y, Fang Z, Lu H. Dual attention network for scene segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2019:3146-3154.
- [7] Zhao H, Zhang Y, Liu S, et al. PSANet: Point-wise spatial attention network for scene parsing. European Conference on Computer Vision (ECCV) 2018:267-283.
- [8] Yang Z, Farsiu S. Directional connectivity-based segmentation of medical images. IEEE/CVF Conference on Computer Vision and Pattern Recognition 2023:11525-11535.

- [9] Beucher S. Use of watersheds in contour detection. International Workshop on Image Processing 1979:17-21.
- [10] Chen LC, Papandreou G, Kokkinos I, et al. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2017;40(4):834-848. <https://doi.org/10.1109/TPAMI.2017.2699184>.
- [11] Zhou W, Dong S, Xu C, et al. Edge-aware guidance fusion network for RGB-thermal scene parsing. *AAAI Conference on Artificial Intelligence* 2022;36(3):3571-3579. <https://doi.org/10.1609/aaai.v36i3.20269>.
- [12] Jian M, Wu R, Xu W, et al. VascuConNet: An enhanced connectivity network for vascular segmentation. *Medical & Biological Engineering & Computing* 2024;62(11):3543-3554. <https://doi.org/10.1007/s11517-024-03150-8>.
- [13] Qu Z, Wang JD, Yin XH. A directional connectivity feature enhancement network for pavement crack detection. *IEEE Transactions on Intelligent Transportation Systems* 2024. <https://doi.org/10.1109/TITS.2024.3488656>.
- [14] Rudra AK, Chowdhury AS, Elnakib A, et al. Kidney segmentation using graph cuts and pixel connectivity. *Pattern Recognition Letters* 2013;34(13):1470-1475. <https://doi.org/10.1016/j.patrec.2013.05.013>.
- [15] Fontaine M, Macaire L, Postaire JG. Image segmentation based on an original multiscale analysis of the pixel connectivity properties. In: *Proceedings of the International Conference on Image Processing (ICIP)* 2000:804-807. <https://doi.org/10.1109/ICIP.2000.901081>.
- [16] Yang Z, Soltanian-Zadeh S, Farsiu S. BiconNet: An edge-preserved connectivity-based approach for salient object detection. *Pattern Recognition* 2022;121:108231. <https://doi.org/10.1016/j.patcog.2021.108231>.
- [17] Qin Y, Chen M, Zheng H, et al. AirwayNet: A voxel-connectivity aware approach for accurate airway segmentation. *International Confer-*

ence on Medical Image Computing and Computer-Assisted Intervention 2019:212-220. https://doi.org/10.1007/978-3-030-32226-7_24.

- [18] Yang Z, Soltanian-Zadeh S, Chu KK, et al. Connectivity-based deep learning approach for segmentation of the epithelium in in vivo human esophageal OCT images. *Biomedical Optics Express* 2021;12(10):6326-6340. <https://doi.org/10.1364/BOE.434775>.
- [19] Li X, Wang Y, Zhang L, et al. Topology-enhanced urban road extraction via a geographic feature-enhanced network. *IEEE Transactions on Geoscience and Remote Sensing* 2020;58(12):8819-8830. <https://doi.org/10.1109/TGRS.2020.2991006>.
- [20] Ma H, Li C, Liu J, et al. Enhance connectivity of promising regions for sampling-based path planning. *IEEE Transactions on Automation Science and Engineering* 2022;20(3):1997-2010. <https://doi.org/10.1109/TASE.2022.3191519>.
- [21] Soltanian-Zadeh S, Liu Z, Liu Y, et al. Deep learning-enabled volumetric cone photoreceptor segmentation in adaptive optics optical coherence tomography images. *Biomedical Optics Express* 2023;14(2):815-833. <https://doi.org/10.1364/BOE.478693>.
- [22] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2014.
- [23] Hu J, Shen L, Sun G. Squeeze-and-excitation networks. *arXiv preprint arXiv:1709.01507*, 2018.
- [24] Wang F, Jiang M, Qian C, et al. Residual attention network for image classification. *arXiv preprint arXiv:1704.06904*, 2017.
- [25] Codella N, Rotemberg V, Tschandl P, et al. Skin lesion analysis toward melanoma detection 2018. *arXiv preprint arXiv:1902.03368*, 2019.
- [26] Caicedo JC, Goodman A, Karhohs KW, et al. Nucleus segmentation across imaging experiments: The 2018 Data Science Bowl. *Nature Methods* 2019;16(12):1247-1253. <https://doi.org/10.1038/s41592-019-0612-7>.

- [27] Kumar N, Verma R, Anand D, et al. A multi-organ nucleus segmentation challenge. *IEEE Transactions on Medical Imaging* 2019;39(5):1380-1391. <https://doi.org/10.1109/TMI.2019.2947628>.
- [28] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention – MICCAI* 2015:234-241. https://doi.org/10.1007/978-3-319-24574-4_28.
- [29] Zhou Z, Rahman Siddiquee MM, Tajbakhsh N, et al. UNet++: A nested U-Net architecture for medical image segmentation. *Deep Learning in Medical Image Analysis* 2018:3-11. https://doi.org/10.1007/978-3-030-00889-5_1.
- [30] Oktay O, Schlemper J, Folgoc LL, et al. Attention U-Net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*, 2018.
- [31] Chen J, Lu Y, Yu Q, et al. TransUNet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*, 2021.
- [32] Tang F, Wang L, Ning C, et al. CMU-Net: A strong ConvMixer-based medical ultrasound image segmentation network. *IEEE International Symposium on Biomedical Imaging (ISBI)* 2023:1-5. <https://doi.org/10.1109/ISBI53787.2023.10230609>.
- [33] Shu X, Wang J, Zhang A, et al. CSCA U-Net: A channel and space compound attention CNN for medical image segmentation. *Artificial Intelligence in Medicine* 2024;150:102800. <https://doi.org/10.1016/j.artmed.2024.102800>.