

Dynamic Snake Convolution based on Topological Geometric Constraints for Tubular Structure Segmentation

Yaolei Qi¹, Yuting He¹, Xiaoming Qi¹, Yuan Zhang¹, and Guanyu Yang^{1*}

¹ LIST, Key Laboratory of Computer Network and Information Integration, Southeast University, Ministry of Education, Nanjing 210096, China

yaolei710@seu.edu.cn, yang.list@seu.edu.cn

Abstract

Accurate segmentation of topological tubular structures, such as blood vessels and roads, is crucial in various fields, ensuring accuracy and efficiency in downstream tasks. However, many factors complicate the task, including thin local structures and variable global morphologies. In this work, we note the specificity of tubular structures and use this knowledge to guide our DSCNet to simultaneously enhance perception in three stages: feature extraction, feature fusion, and loss constraint. First, we propose a dynamic snake convolution to accurately capture the features of tubular structures by adaptively focusing on slender and tortuous local structures. Subsequently, we propose a multi-view feature fusion strategy to complement the attention to features from multiple perspectives during feature fusion, ensuring the retention of important information from different global morphologies. Finally, a continuity constraint loss function, based on persistent homology, is proposed to constrain the topological continuity of the segmentation better. Experiments on 2D and 3D datasets show that our DSCNet provides better accuracy and continuity on the tubular structure segmentation task compared with several methods. Our codes will be publicly available¹.

1. Introduction

The accurate segmentation of topological tubular structures is paramount in various fields to ensure the precision and efficiency of downstream tasks. In clinical applications, a well-delineated blood vessel is a crucial prerequisite for computational hemodynamics, and it assists radiologists in locating and diagnosing lesions [13, 16]. In remote sensing applications, complete road segmentation provides a solid foundation for route planning. Regardless of the field, these structures share common features of being thin and tortuous, which make them challenging to capture due to their small proportion in images. Therefore, there is an urgent need to enhance the perception of thin tubular structures.

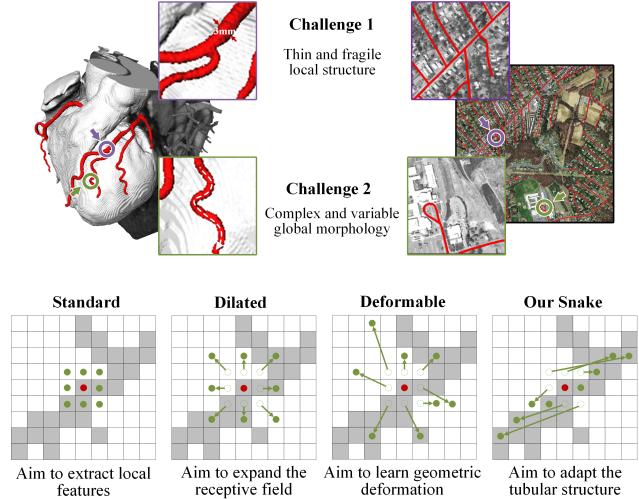


Figure 1. **Challenges.** The above figure shows a 3D heart vascular dataset and a 2D remote road dataset. Both datasets aim to extract tubular structures, but this task faces challenges due to fragile local structures and complex global morphology. **Motivation.** The standard convolutional kernel is intended to extract local features. On this basis, deformable convolutional kernels have been designed to enrich their application and adapt to geometric deformations of different targets. However, due to the aforementioned challenges, it is difficult to focus efficiently on the thin tubular structures.

However, it remains challenging due to the following difficulties: (1) **Thin and fragile local structure.** As shown in Figure 1, thin structures account for only a small proportion of the overall image with limited pixel composition. Moreover, these structures are susceptible to interference from complex backgrounds, rendering it difficult to precisely discriminate subtle target variations by the model. Consequently, the model may struggle to differentiate these structures, resulting in the fracture of the segmentation.

(2) **Complex and variable global morphology.** Figure 1

*Corresponding author

¹<https://github.com/YaoleiQi/DSCNet>

shows the complex and variable morphology of thin tubular structures, even within the same image. Morphological variations are observed in targets located in different regions, depending on the number of branches, the location of bifurcations, and the path length. The model may tend to overfit features that have already been seen, resulting in weak generalization when the data exhibits unprecedented morphological structures.

Recently, many studies have proposed incorporating domain knowledge (*e.g.* geometric topology and tree structure) to guide the model better to perceive the distinctive features of the tubular structures, thus focusing on improving the accuracy of local segmentation and maintaining the continuity of global morphology. Existing methods can be broadly classified into three categories: (1) Network-based methods [7, 12, 31, 28, 14, 8] design specific network architectures according to the characteristics of the tubular structures that guide the model to focus on critical features. However, given the small proportion of tubular structures, the network may inevitably lose the perception of the corresponding structures. (2) Feature-based methods [21, 33, 15, 35, 20] enhance the comprehension of the particular geometric and topological features of tubular structures by supplementing the model with additional feature representations. However, some redundant feature representations exacerbate the computational burden while not providing a positive influence on the model. (3) Loss-based methods [24, 29, 1, 30] incorporate measurement methods to supplement constraints during the training process, typically through loss functions. These methods reinforce the stringent constraints on segmentation. Building on this foundation, structured losses combined with continuity constraints from the topological perspective will potentially further improve the accuracy of the tubular segmentation.

To tackle the above obstacles, we propose a novel framework, DSCNet, involving a tubular-aware convolution kernel, multi-view feature fusion strategy, and topological continuity constraint loss function. (1) To address the challenge of the small proportion of thin and fragile local structures that are difficult for the model to focus on, we propose Dynamic Serpentine Convolution (DSConv) to enhance the perception of the geometric structure by adaptively focusing on the thin and curved local features of tubular structures. Unlike deformable convolution [6], which leaves the network completely free to learn geometrical changes, thus causing perceptual regions to wander, especially on thin tubular structures, our DSConv considers the snake-like morphology of tubular structures and supplements the free learning process with constraints that allow for targeted enhancement of the perception of tubular structures. (2) To address the challenge of complex and variable global morphology, we propose a multi-view feature fusion strategy. In this method, we generate multiple morphological kernel

templates based on DSConv for viewing the structural features of the target from various perspectives and achieve efficient feature fusion by summarizing typical vital features. (3) To address the problem that segmentation of tubular structures is prone to fracture, we propose a Topological Continuity Constrained Loss Function (TCLoss) based on Persistent Homology (PH). PH[9, 19, 5] responds to the process of topological features from emergence to disappearance. It obtains adequate topological information from noisy high-dimensional data. The related Betti numbers are one way of describing connectivity in the topological space. Unlike [30, 10], our TLCoss combines PH with point set similarity to pilot the network to focus on fracture regions with anomalous pixel/voxel distributions, achieving continuity constraints from a topological perspective.

To summarize, our work proposes a novel framework of knowledge fusion addressing the difficulties of the thin tubular structures, and the specific contributions are three-fold. (1) We propose a dynamic snake convolution to adaptively focus on the slender and tortuous local features and realize the accurate tubular structures segmentation on both 2D and 3D datasets. Our model is thoroughly verified using both internal and external test data. (2) We propose a multi-perspective feature fusion strategy to supplement the attention to the vital features from multiple perspectives. (3) We propose a topological continuity constraint loss function based on Persistent Homology, which better constrains the continuity of the segmentation.

2. Related Work

2.1. Methods based on Network Design

Various methods have been proposed to achieve better performance by designing particular network architectures and modules according to the morphology of the tubular structures. (1) Methods based on the convolutional kernel design, represented by the famous dilated convolution [32] and deformable convolution [6], are proposed to deal with the inherent limited geometric transformation in CNNs, showing outstanding performance on sophisticated detection and segmentation tasks. These methods [7, 34, 12, 31] are also designed to dynamically perceive the geometric features of the objects to adapt to the structure with changeable morphology. For example, DUNet proposed in [12] integrates the deformable convolution into the U-shape architecture and adaptively adjusts the receptive field according to the vessels’ scales and shapes. (2) Methods based on the network architecture design are proposed to learn the special geometric topological features of the tubular structures. PointScatter [28] is proposed to represent the tubular structure with points set, which is an alternative to the segmentation models for the tubular structure extraction task. [14] proposed a tree-structured convolutional gated recur-

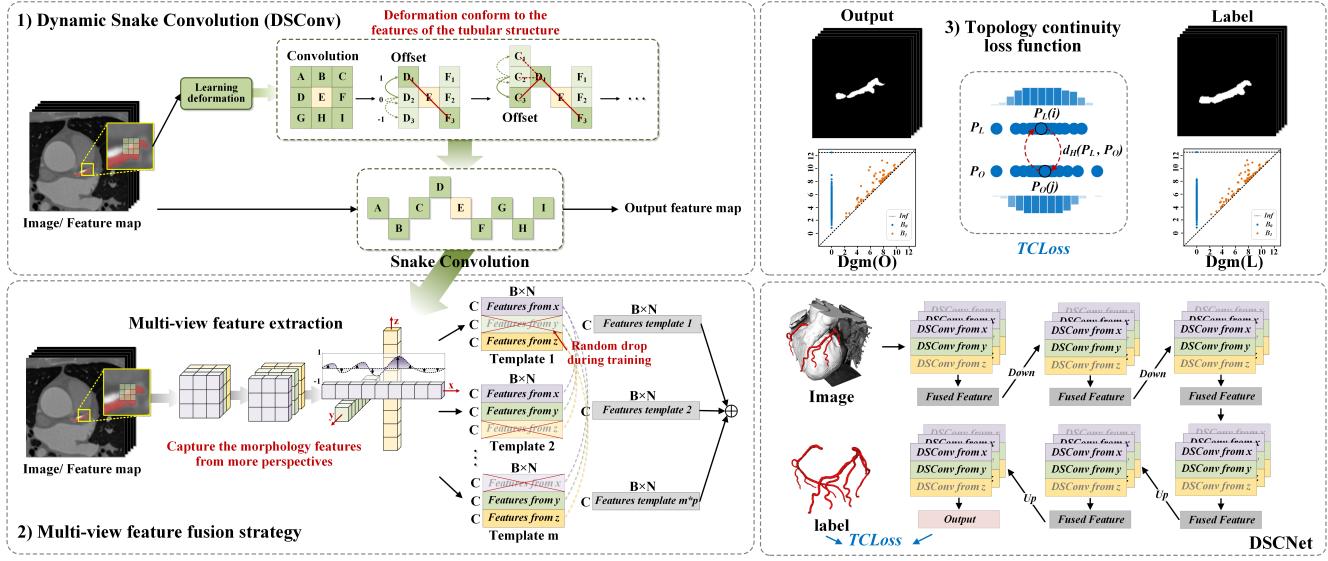


Figure 2. Methodology. Schematic overview of our proposed method illustrated on an example of the 3D coronary artery segmentation. Our method has three sections: (1) Dynamic snake convolution (DSConv), which learns the deformation according to the input feature map, adaptively focus on the slender and tortuous local features under the knowledge of the tubular structure morphology. (2) Multi-view feature fusion strategy, which generates multiple morphological kernel templates based on our DSConv and is used to observe the structural characteristics of the target from multiple perspectives. (3) Loss function, called topological continuity constraint loss function (TCLoss), based on Persistent Homology to guide the network to focus on the fracture regions with abnormally low pixels/voxels distribution and realize continuity constraint.

rent unit to model the coronary artery's topological structures explicitly. Different from the ideas mentioned above of allowing the model to learn the geometric changes completely freely, considering the limitation of the convergence difficulty caused by excessive randomness and the possibility that the model will focus on the unexpected regions of the target. Our work integrates the domain knowledge of tubular structure morphology to stably enhance the perception of the tubular structure in the feature extraction process.

2.2. Methods based on Features Fusion

Methods based on features fusion [35, 15, 33, 21, 20] strengthens the representation of the tubular structure by supplementing additional feature information to the model. Considering the topology and the sparsity of the tubular structures, [35] proposed a cross-network multi-scale feature fusion method performed between two networks to support high-quality vessel segmentation effectively. In [15], a global transformer and dual local attention network via deep-shallow hierarchical feature fusion are investigated to simultaneously capture the global and local characterizations. [33] proposed to fuse the contextual anatomical information and vascular topologies for accurate tubular structure segmentation. In our work, we propose a multi-perspective feature fusion strategy to supplement the attention to the vital features from multiple perspectives. In this strategy, we generate numerous morphological kernel tem-

plates based on our DSConv to observe the structural characteristics of the target from multiple perspectives and realize the feature fusion by summarizing the essential standard features, thus improving the performance of our model.

2.3. Methods based on Loss Function

Methods based on loss function [24, 29, 1] introduce measurement methods to supplement constraints in the training process. These methods strengthen the strong constraints on the tubular structures segmentation. [24] introduced a similarity measurement termed centerline Dice, which is calculated on the intersection of the segmentation masks and the skeleton. [29] proposed a geometry-aware tubular structure segmentation method, Deep Distance Transform (DDT), which combines intuitions from the classical distance transform for skeletonization and tubular structure segmentation. These methods focus on the continuity of the tubular structure segmentation, but the skeleton's inaccuracy and offset will affect the constraints' precision. [1] proposed a similarity index that captures the topological consistency of the predicted segmentation and designs a loss function based on the morphological closing operator for tubular structure segmentation. In [30], topological data analysis methods are incorporated with a geometric deep learning model for fine-grained segmentation for 3D objects. These methods will capture the features of the topological objects. Drawing inspiration from this, our

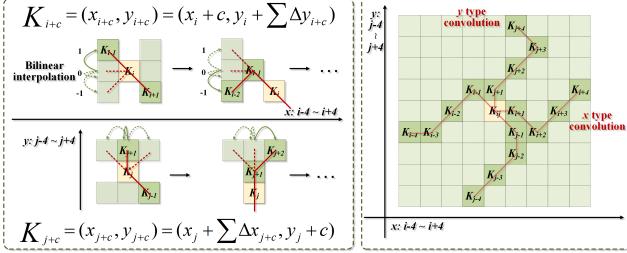


Figure 3. **Left:** Illustration of the coordinates calculation of the DSConv. **Right:** The receptive field of the DSConv.

work proposes a topological continuity constraint loss function (TCLoss) that better constrains the continuity of the segmentation from a topological perspective. Our TLCoss gradually introduces constraints based on Persistence Homology [27, 2] during the training process to guide the network to focus on the fracture regions and realize continuity.

3. Methodology

Our method is designed to deal simultaneously with 2D and 3D feature maps of thin tubular structures. For simplicity, our modules are described in 2D, and the detailed extension to 3D is also provided in our open source.

3.1. Dynamic Snake Convolution

In this section, we discuss how to perform Dynamic Snake Convolution (DSConv) to extract the tubular structure's local features. Given the standard 2D convolution coordinates as K , the central coordinate is $K_i = (x_i, y_i)$. A 3×3 kernel K with dilation 1 is expressed as:

$$K = \{(x - 1, y - 1), (x - 1, y), \dots, (x + 1, y + 1)\} \quad (1)$$

To give the convolution kernel more flexibility to focus on complex geometric features of the target, inspired by [6], we introduced deformation offsets Δ . However, if the model is left free to learn the deformation offsets, the perceptual field tends to stray outside the target, especially in the case of thin tubular structures. Therefore, we use an iterative strategy (Figure 3), selecting the following position to be observed in turn for each target to be processed, thus ensuring continuity of the attention and not spreading the field of sensation too far due to the large deformation offsets.

In DSConv, we straighten the standard convolution kernel, both in the direction of the x-axis and the y-axis. We consider a convolution kernel of size 9 and take the x-axis direction as an example, the specific position of each grid in K is represented as: $K_{i\pm c} = (x_{i\pm c}, y_{i\pm c})$, where $c = \{0, 1, 2, 3, 4\}$ denotes the horizontal distance from the central grid. The selection of each grid position $K_{i\pm c}$ in the convolution kernel K is a cumulative process. Starting from the center position K_i , the position away from the center grid depends on the position of the previous grid: K_{i+1}

is augmented with an offset $\Delta = \{\delta | \delta \in [-1, 1]\}$ compared to K_i . Hence, the offset needs to be Σ , thus ensuring that the convolution kernel conforms to a linear morphological structure. Figure 1 in the direction of the x-axis becomes:

$$K_{i\pm c} = \begin{cases} (x_{i+c}, y_{i+c}) = (x_i + c, y_i + \Sigma_i^{i+c} \Delta y), \\ (x_{i-c}, y_{i-c}) = (x_i - c, y_i + \Sigma_{i-c}^i \Delta y), \end{cases} \quad (2)$$

and Equation 2 in the direction of the y-axis becomes:

$$K_{j\pm c} = \begin{cases} (x_{j+c}, y_{j+c}) = (x_j + \Sigma_j^{j+c} \Delta x, y_j + c), \\ (x_{j-c}, y_{j-c}) = (x_j + \Sigma_{j-c}^j \Delta x, y_j - c), \end{cases} \quad (3)$$

As the offset Δ is typically fractional, bilinear interpolation is implemented as:

$$K = \Sigma_{K'} B(K', K) \cdot K' \quad (4)$$

where K denotes a fractional location for Equation 2 and Equation 3, K' enumerates all integral spatial locations, and B is the bilinear interpolation kernel and it is separated into two one-dimensional kernels as:

$$B(K, K') = b(K_x, K'_x) \cdot b(K_y, K'_y) \quad (5)$$

where $b(K, K') = \max(0, 1 - |K - K'|)$.

As shown in Figure 3, our DSConv covers a 9×9 range during the deformation process due to the two-dimensional (x-axis, y-axis) changes.

3.2. Multi-view Feature Fusion Strategy

This section discusses implementing the multi-view feature fusion strategy to guide the model to complement the focus on essential features from multiple perspectives. For each K , two feature maps $f^l(K_x)$ and $f^l(K_y)$ from layer l are extracted from the x-axis and y-axis, expressed as:

$$f^l(K) = \underbrace{\{\Sigma_i w(K_i) \cdot f^l(K_i), \Sigma_j w(K_j) \cdot f^l(K_j)\}}_{f^l(K_x)} \quad (6)$$

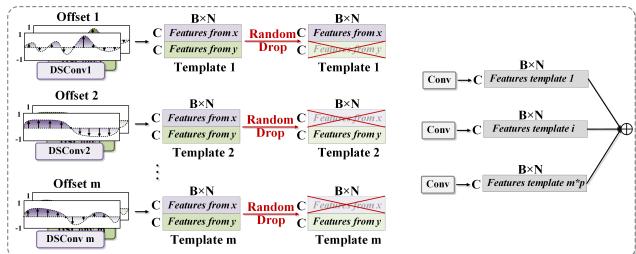


Figure 4. The multi-view feature fusion strategy.

where $w(K_i)$ denotes the weight at position K_i , and the features extracted by the l -th layer convolution kernel K are calculated using the cumulative approach.

On the basis of Equation 6, we extract m groups of features as templates T^l , which contains different morphology of the DSConv:

$$T^l = (\underbrace{f^l(K_x), f^l(K_y)}_{T_1^l}, \underbrace{f^l(K_x), f^l(K_y)}_{T_2^l}, \dots, \underbrace{f^l(K_x), f^l(K_y)}_{T_m^l}) \quad (7)$$

The feature fusion of multiple templates will inevitably bring redundant noise. Therefore, a random dropping strategy r^l is introduced during the training stage (Figure 4), to improve the performance of our model and prevent overfitting without increasing additional computational burden, then Equation 7 becomes:

$$\begin{cases} r^l \sim \text{Bernoulli}(p) \\ \hat{T}^l = r^l \cdot T^l \\ f^{l+1}(K) = \Sigma^{[m \times p]} \hat{T}_t^l \end{cases} \quad (8)$$

where p is the probability of random dropping and r^l satisfies the Bernoulli distribution. The optimal dropping strategy is saved during the training phase and guides the model to fuse key features during the testing phase.

3.3. Topological Continuity Constraint Loss

In this section, we discuss how to implement the topological continuity constraint loss (TCLoss) based on the Persistent Homology to constrain the continuity of the segmentation. The geometric and topological information in complex structures is a pivotal clue to help the models understand the continuous structures. Tools from the topological data analysis are adopted to extract the essential features hidden in the complex tubular structures.

We aim to construct the topology of the data and extract the high-dimensional relationships in the complex tubular structure, represented as persistence barcodes and Persistence Homology (PH), as illustrated in Figure 5.

Given G , its N -dimensional topological structure, homology class [9, 19] is an equivalence class of N -manifolds which can be deformed into each other within G , where 0-dimensional and 1-dimensional are connected components and handles. PH is applied to compute the evolution of topological features, and the period between the appearance time b and disappearance time d of topological features is kept [30]. Such periods are summarized in a concise format called a persistence diagram (PD), which consists of a set of points (b, d) . Each point (b, d) represents the d -th homology class that appears at b and disappears at d . Let $PD = \text{dgm}(\cdot)$ denote the persistent homology obtained from the groundtruth L and the output O .

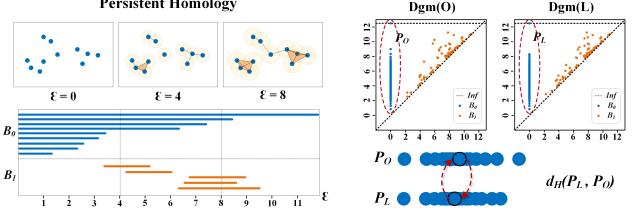


Figure 5. Illustration of the Persistent Homology and our TC Loss.

We consider the topological information in complex tubular structures, which contains the key clues to determine the presence of fractures, to be evident in the homotopy features of 0-dimensional and 1-dimensional homological features. The existing methods [30, 5, 10] use modified Wasserstein distance to compute the best match between the points generated by the output and the points generated by the groundtruth, and the outlier points without optimal pairing are matched to the diagonal and do not participate in the loss calculation. However, in our task, the outlier points represent anomalous appearing or disappearing time and imply wrong topological relations that play an important role. Therefore, we measure the similarity between the two sets of points using Hausdorff distance[26]:

$$\begin{cases} d_H(P_O, P_L) = \max_{u \in P_O} \min_{v \in P_L} \| u - v \| \\ d_H(P_L, P_O) = \max_{v \in P_L} \min_{u \in P_O} \| v - u \| \\ d_H^* = \max\{d_H(P_O, P_L), d_H(P_L, P_O)\} \end{cases} \quad (9)$$

where $P_O \in \text{Dgm}(O)$, $P_L \in \text{Dgm}(L)$ and d_H^* represents the bidirectional Hausdorff distance, which is computed in terms of n -dim points. Our used Hausdorff distance is sensitive to outliers. As shown in Equation 9, if two sets of points are similar, all the points are perfectly superimposed except only one point in P_O , which is far from any point in P_H , then the Hausdorff distance is determined by that point and is large [11].

Then the summation over all dimensions ($n = 0, 1, 2, \dots, N$) is performed to obtain the \mathcal{L}_{PH} and the whole TC Loss is integrated with the cross entropy loss \mathcal{L}_{CE} as the final loss function \mathcal{L}_{TC} :

$$\mathcal{L}_{TC} = \mathcal{L}_{CE} + \mathcal{L}_{PH} = \mathcal{L}_{CE} + \sum_{n=0}^N d_H^* \quad (10)$$

Finally, the topology and accuracy are constrained by the combined effect of the two loss functions, contributing to continuous tubular segmentation.

4. Experiments Configurations

4.1. Datasets

We employ three datasets containing two public and one internal dataset for validating our framework. In 2D, we evaluate the DRIVE retina dataset[25] and the Massachusetts Roads dataset[17]. In 3D, we used a dataset called Cardiac CCTA Data. Details concerning the experimental setup can be found in the supplementary material.

4.2. Evaluation Metrics

We performed comparative experiments and ablation studies to demonstrate the advantages of our proposed framework. The classical segmentation network U-Net [4] and the CS²-Net [18] proposed in 2021 for vascular segmentation are compared to validate the accuracy. To validate the network design performance, we compared the DCU-net [31] proposed in 2022 for retinal vascular segmentation. To validate the advantages of feature fusion, we compared the Transunet [3] for medical image segmentation proposed in 2021. To validate the loss function constraint, we compared cIDice [24] proposed in 2021 and Wasserstein-distance-based TCLoss \mathcal{L}_{WTC} [30]. These models are trained on the same dataset with the exact implementation and were evaluated by the following metrics. All metrics were calculated for each image and averaged.

1. Volumetric scores: *Mean Dice Coefficient (Dice)*, *Relative-Dice coefficient (RDice)*[22], *CenterlineDice (cIDice)*[24], Accuracy (ACC) and AUC are used to evaluate the performance of the results
2. Topology errors: We follow [24, 28] and calculate the topology-based scores including the *Betti Errors* for Betti numbers β_0 and β_1 . Meanwhile, to objectively verify the continuity of the coronary artery segmentation, the *overlap until first error (OF)* [23] is used to evaluate the completeness of the extracted centerline.
3. Distance errors: *Hausdorff Distance (HD)* [26] is also widely used to describe the similarity between two sets of points, which is recommended to evaluate the thin tubular structures.

5. Results and Discussion

In this paragraph, we will evaluate and analyze our proposed framework's effectiveness in three ways: (1) The performance of our proposed method for the thin tubular structure segmentation task is compared and verified by the following metrics. The visual effects of different methods are simultaneously shown. (2) We analyzed the effectiveness of our proposed DSConv to guide the model to focus on the tubular structure, and the help of the TCLoss to constrain

the topology of the segmentation. (3) We provide comprehensive experiments, including ablation studies, on the DRIVE dataset as an example. Additionally, due to space constraints, we highlight some of the most important comparison experiments on other datasets. The results show the strong performance of our method on both 2D and 3D fields.

5.1. Quantitative Evaluation

The advantages of our method on each metric are demonstrated in Table 1, and the results show that our proposed DSCNet achieves better results on both 2D and 3D datasets.

Evaluation on DRIVE. On the DRIVE dataset, our DSCNet outperforms the other models regarding segmentation accuracy and topological continuity. In Table 1, our proposed DSCNet achieves the best segmentation results compared with other methods with Dice of 82.06%, RDice of 90.17%, cIDice of 82.07%, ACC of 96.87%, and AUC of 90.27% from the perspective of the volumetric accuracy. Meanwhile, from the view of the topology, our DSCNet achieves the best topological continuity compared with other methods with β_0 error of 0.998 and β_1 error of 0.803. The results show that our method better captures the specific features of the thin tubular structures and exhibits a more accurate segmentation performance and a more continuous topology. As shown in the sixth to twelfth rows of Table 1, with the addition of our TCLoss, different models both show improvement in the topological continuity of the segmentation. The results illustrate that our TCLoss accurately constrains the model to focus on thin tubular structures that lose topological continuity.

Evaluation on ROADS. On the Massachusetts Roads dataset, our DSCNet also achieves the best results. As shown in Table 1, our proposed DSCNet with TCLoss achieves the best segmentation results compared with other methods with Dice of 78.21%, RDice of 85.85%, and cIDice of 87.64%. Compared with the results of the classical segmentation network UNet, our method achieves at most 1.31% Dice, 1.78% RDice, and 0.77% cIDice improvements. The results show that our model also performs well for structurally complex and morphologically variable road datasets compared with other models.

Evaluation on CORONARY. On the Cardiac CCTA dataset, we verify that our DSCNet still achieves the same best results for segmenting the thin tubular structures in 3D. As shown in Table 2, our proposed DSCNet achieves the best segmentation results compared with other methods with Dice of 80.27%, RDice of 86.37%, and cIDice of 85.26%. Compared with the results of the classical segmentation network UNet, our method achieves at most 3.40% Dice, 1.89% RDice, and 3.83% cIDice improvements. Meanwhile, we used OF metrics to assess the continuity of segmentation. With our method, the OF metrics improved by 6.00% for LAD, 3.78% for LCX, and 3.30%

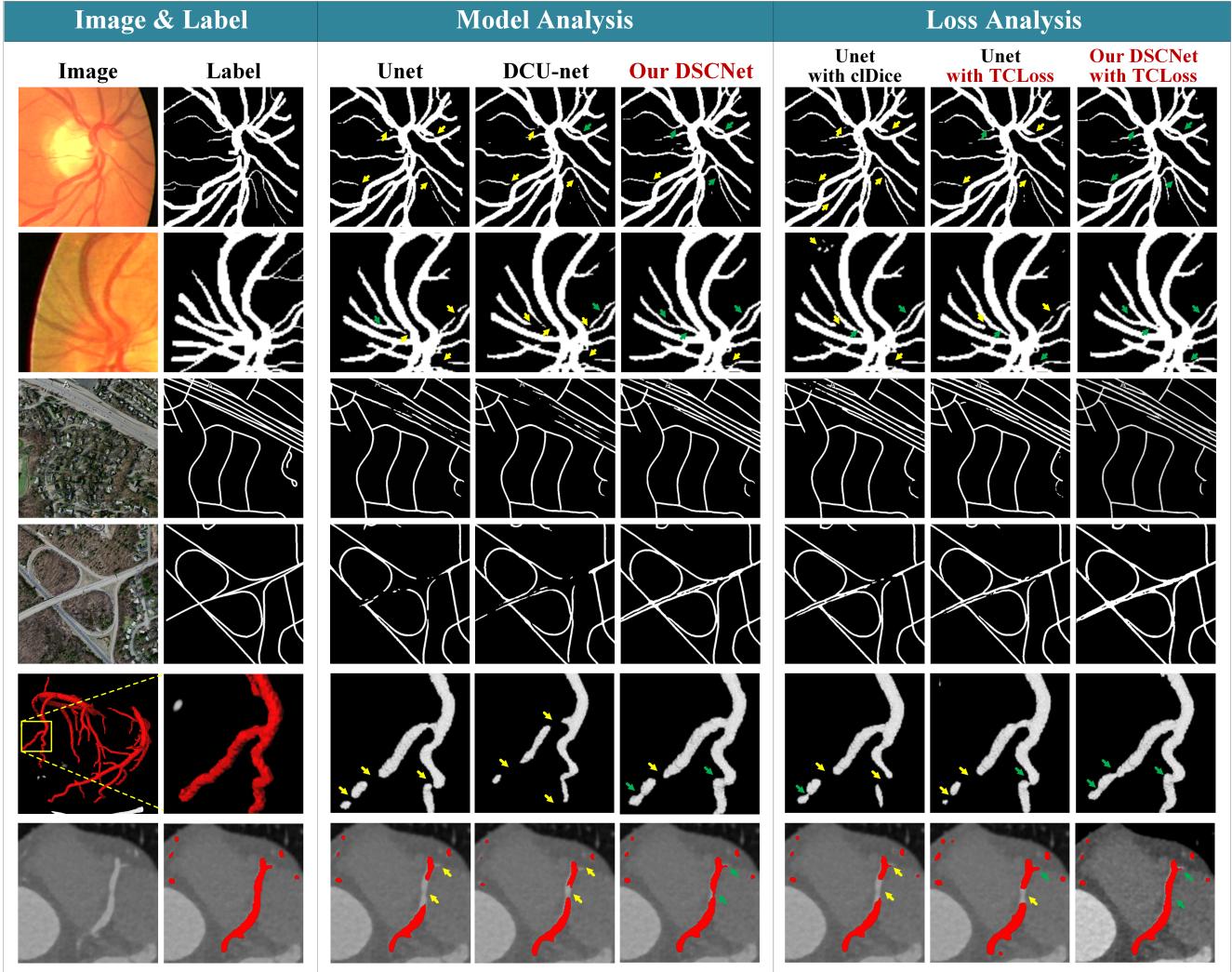


Figure 6. Qualitative results. To verify the performance of our method more objectively and efficiently, we selected representative hard-to-segment regions from each dataset. From top to bottom, we show two rows of results for the DRIVE dataset, the Massachusetts road dataset, and our internal Coronary dataset. From left to right, we show the original image, groundtruth, and the results from classical UNet, DCU-net, our DSCNet, UNet with cIDice, UNet with our proposed TCloss, and our DSCNet with TCloss. The results indicate that our DSCNet and TCloss outperform the other models regarding segmentation accuracy and topological continuity. The yellow arrows indicate the areas where the segmentation is broken, while the green arrows indicate areas where the segmentation is performing well.

for RCA (LAD, LCX, and RCA are the vital trunks of the coronary vessels). The improvement in the continuity of vessels plays a crucial role in the clinic.

Ablation Experiment Analysis. Taking the DRIVE dataset as an example, the ablation experiments prove the importance of our DSCNet and our TCloss. (1) To prove the effectiveness of our DSCNet. The results in the first five rows of Table 1 show that our method is better suited to segmenting thin tubular structures. The results show that our proposed DSConv plays a vital role in the model, helping the network better to capture the critical features of the thin tubular structures. (2) To prove the effectiveness of our TCloss. As shown in the sixth to ninth rows of Ta-

ble 1, with the addition of our TCloss, different models both show improvement in the topological continuity of the segmentation. The results illustrate that our TCloss accurately constrains the model to focus on thin tubular structures that lose topological continuity.

5.2. Qualitative Evaluation

Our DSCNet and TCloss have decisive visual superiority in arbitrary aspects (Figure 6). (1) To demonstrate the effectiveness of our DSCNet. From left to right, the third to fifth columns show the performance of the different networks in terms of segmentation accuracy. Thanks to our DSConv to adaptively perceive critical features of the

Dataset	Network	Loss	Volumetric (%) ↑					Topology ↓		Distance ↓
			Dice	RDice	cIDice	ACC	AUC	β_0	β_1	
DRIVE	UNet	\mathcal{L}_{CE}	80.73 \pm 1.77	87.94 \pm 3.32	79.66 \pm 4.00	96.74 \pm 0.28	88.57 \pm 2.44	1.209 \pm 0.342	0.883 \pm 0.135	6.86 \pm 0.56
	Transunet	\mathcal{L}_{CE}	80.56 \pm 2.14	87.14 \pm 3.82	79.02 \pm 5.05	96.75 \pm 0.32	88.02 \pm 2.79	1.210 \pm 0.309	0.844 \pm 0.157	6.83 \pm 0.52
	CS ² -Net	\mathcal{L}_{CE}	77.53 \pm 2.94	82.55 \pm 4.10	74.88 \pm 5.27	96.46 \pm 0.36	84.73 \pm 2.82	1.391 \pm 0.331	0.906 \pm 0.177	6.90 \pm 0.48
	DCU-net	\mathcal{L}_{CE}	80.83 \pm 1.99	87.73 \pm 3.60	80.19 \pm 4.80	96.77 \pm 0.31	88.45 \pm 2.67	1.104 \pm 0.327	0.817 \pm 0.166	6.84 \pm 0.58
	DSCNet(ours)	\mathcal{L}_{CE}	81.85\pm1.74	88.93\pm3.36	81.16\pm4.54	96.91\pm0.28	89.38\pm2.54	1.094\pm0.301	0.780\pm0.162	6.68\pm0.49
	UNet	$\mathcal{L}_{TC}(\text{ours})$	80.93 \pm 1.97	88.00 \pm 3.41	80.28 \pm 4.41	96.78 \pm 0.30	88.63 \pm 2.56	1.117 \pm 0.286	0.797\pm0.151	6.88 \pm 0.53
	Transunet	$\mathcal{L}_{TC}(\text{ours})$	80.79 \pm 2.11	87.78 \pm 3.80	79.86 \pm 4.90	96.76 \pm 0.32	88.48 \pm 2.82	1.176 \pm 0.295	0.818 \pm 0.176	6.83 \pm 0.51
	CS ² -Net	$\mathcal{L}_{TC}(\text{ours})$	79.69 \pm 2.31	86.14 \pm 3.82	77.72 \pm 5.09	96.64 \pm 0.32	87.25 \pm 2.76	1.308 \pm 0.334	0.848 \pm 0.160	6.93 \pm 0.45
ROADS	DCU-net	$\mathcal{L}_{TC}(\text{ours})$	81.18 \pm 1.90	87.89 \pm 3.43	80.60 \pm 4.54	96.83 \pm 0.31	88.59 \pm 2.57	1.076 \pm 0.313	0.817 \pm 0.167	6.80 \pm 0.56
	UNet	cIDice	80.77 \pm 1.92	87.53 \pm 3.42	79.93 \pm 4.48	96.77 \pm 0.31	88.29 \pm 2.52	1.199 \pm 0.303	0.833 \pm 0.157	6.93 \pm 0.54
	UNet	\mathcal{L}_{WTC}	80.89 \pm 1.95	87.85 \pm 3.55	80.03 \pm 4.75	96.78 \pm 0.29	88.53 \pm 2.64	1.144 \pm 0.339	0.814 \pm 0.176	6.79 \pm 0.47
	DSCNet(ours)	$\mathcal{L}_{TC}(\text{ours})$	82.06\pm1.44	90.17\pm3.04	82.07\pm4.35	96.87\pm0.24	90.27\pm2.32	0.998\pm0.312	0.803 \pm 0.179	6.78\pm0.51
	UNet	\mathcal{L}_{CE}	76.90 \pm 6.30	84.07 \pm 6.46	86.87 \pm 6.59	97.97 \pm 1.27	98.29 \pm 1.24	1.107 \pm 0.551	1.505 \pm 0.467	8.11 \pm 2.42
	Transunet	\mathcal{L}_{CE}	75.82 \pm 6.83	81.50 \pm 6.65	86.04 \pm 7.40	97.97 \pm 1.28	98.23 \pm 1.15	1.105 \pm 0.615	1.570 \pm 0.663	8.11 \pm 2.53
	DCU-net	\mathcal{L}_{CE}	77.24 \pm 6.30	84.26 \pm 6.37	86.98 \pm 6.53	98.03 \pm 1.14	98.34 \pm 1.19	1.085 \pm 0.653	1.474 \pm 0.497	8.04 \pm 2.53
	UNet	$\mathcal{L}_{TC}(\text{ours})$	77.70 \pm 6.07	84.80 \pm 5.96	87.47 \pm 6.31	98.03 \pm 1.23	98.41 \pm 1.13	1.072 \pm 0.631	1.401 \pm 0.496	8.04 \pm 2.72
CORONARY	UNet	cIDice	77.37 \pm 5.57	84.18 \pm 5.99	87.05 \pm 6.34	98.03 \pm 1.22	98.40 \pm 1.12	1.079 \pm 0.613	1.407 \pm 0.603	8.08 \pm 2.46
	DSCNet(ours)	\mathcal{L}_{CE}	78.04 \pm 5.77	85.35 \pm 5.42	87.74\pm6.02	98.05 \pm 1.21	98.39 \pm 1.19	1.118 \pm 0.641	1.441 \pm 0.523	7.96 \pm 2.43
	DSCNet(ours)	$\mathcal{L}_{TC}(\text{ours})$	78.21\pm5.77	85.85\pm5.56	87.64 \pm 5.99	98.05\pm1.21	98.46\pm1.08	1.053\pm0.523	1.396\pm0.456	7.34\pm2.48

Table 1. Quantitative results for the DRIVE retina dataset and the Massachusetts road dataset (ROADS) are shown in this table. All experiments verified the performance of our method on three levels: volumetric accuracy, topological connectivity, and distance error. Our DSCNet and TCloss achieve the most competitive results in all metrics.

Dataset	Network	Loss	Volumetric (%) ↑			Topology OF ↑			Distance ↓	
			Dice	RDice	cIDice	LAD	LCX	RCA	HD	
CORONARY	UNet	\mathcal{L}_{CE}	76.87 \pm 5.38	84.48 \pm 4.55	81.43 \pm 6.02	0.806 \pm 0.252	0.847 \pm 0.239	0.849 \pm 0.267	7.727 \pm 3.30	
	Transunet	\mathcal{L}_{CE}	76.70 \pm 6.65	83.23 \pm 6.72	78.71 \pm 6.93	0.810 \pm 0.274	0.694 \pm 0.307	0.816 \pm 0.303	8.580 \pm 4.11	
	DCU-net	\mathcal{L}_{CE}	78.33 \pm 5.00	85.67 \pm 4.29	82.29 \pm 5.31	0.833 \pm 0.219	0.746 \pm 0.296	0.835 \pm 0.300	7.331 \pm 3.06	
	UNet	cIDice	77.86 \pm 5.25	84.42 \pm 4.65	82.37 \pm 5.54	0.817 \pm 0.256	0.845 \pm 0.234	0.859 \pm 0.265	7.412 \pm 3.68	
	DSCNet(ours)	\mathcal{L}_{CE}	79.92 \pm 5.26	85.98 \pm 4.60	84.95 \pm 5.76	0.858 \pm 0.198	0.853 \pm 0.241	0.862 \pm 0.267	6.326 \pm 2.85	
	DSCNet(ours)	$\mathcal{L}_{TC}(\text{ours})$	80.27\pm4.67	86.37\pm4.16	85.26\pm4.98	0.866\pm0.195	0.885\pm0.210	0.882\pm0.250	5.787\pm2.99	

Table 2. Quantitative results for the 3D Cardiac CCTA dataset (CORONARY) are shown in this table. Experimental metrics supplement with OF metrics for verifying the topological connectivity. The LAD, LCX, and RCA are the three main trunks of the coronary vessels and are of great clinical value.

thin tubular structure, our model focuses more accurately on special tubular features than other methods, thus showing better performance on tubular structure segmentation. (2) To demonstrate the effectiveness of our TCloss. From left to right, the sixth to eighth columns show the performance of different loss functions on the continuity of the segmentation of the thin tubular structure. With the addition of our proposed TCloss, the continuity of the segmentation is greatly improved in hard-to-segment regions. The results confirm that our method gives a stable segmentation performance with better topological continuity, especially in complex and variable morphological structures. Notably, on the Massachusetts Roads dataset, our model achieves good visualization on adjacent straight or curved roads. More visualization results can be found in the supplementary material.

5.3. Future Work

Our proposed framework copes well with the segmentation of thin tubular structures and successfully integrates morphological features with topological knowledge to guide the model to adapt to the segmentation. However, whether other morphological targets will achieve better performance with a similar paradigm is still an exciting topic. Meanwhile, more research will investigate the possibility of incorporating other types of domain knowledge or topological analysis to further improve the performance of the segmentation. Furthermore, more experimental and theoretical validation will enrich this topic more.

6. Conclusion

In this study, we pay attention to the special features of the tubular structures and use this knowledge to guide the model to enhance the perception simultaneously in three

stages: feature extraction, feature fusion, and loss constraint. Firstly, we propose a dynamic snake convolution to adaptively focus on the thin and tortuous local structure, thus accurately capturing the features of tubular structures. Secondly, we introduce a multi-view feature fusion strategy to complement the focus on features from multiple angles during feature fusion, ensuring the retention of important information from different global morphology. Lastly, we propose a topological continuity constraint loss function to constrain the topological continuity of the segmentation. Our method is verified on both 2D and 3D datasets and the results show that our method provides better accuracy and continuity on the tubular structure segmentation task compared with several methods.

References

- [1] R. J. Araújo, Jaime S Cardoso, and H. P Oliveira. Topological Similarity Index and Loss Function for Blood Vessel Segmentation. *IEEE TRANSACTIONS ON MEDICAL IMAGING*, XX:1, jul 2021.
- [2] Ulrich Bauer. Ripser: efficient computation of Vietoris-Rips persistence barcodes. *J. Appl. Comput. Topol.*, 5(3):391–423, 2021.
- [3] Jieneng Chen, Yongyi Lu, Qihang Yu, et al. Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*, 2021.
- [4] Özgün Çiçek, Ahmed Abdulkadir, Soeren S. Lienkamp, et al. 3d u-net: Learning dense volumetric segmentation from sparse annotation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 424–432. Springer, 2016.
- [5] David Cohen-Steiner, Herbert Edelsbrunner, John Harer, et al. Lipschitz functions have l^1 p-stable persistence. *Foundations of computational mathematics*, 10(2):127–139, 2010.
- [6] Jifeng Dai, Haozhi Qi, Yuwen Xiong, et al. Deformable convolutional networks. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.
- [7] Shunjie Dong, Zixuan Pan, Yu Fu, et al. Deu-net 2.0: Enhanced deformable u-net for 3d cardiac cine mri segmentation. *Medical Image Analysis*, 78:102389, 2022.
- [8] Yuting He, Rongjun Ge, Jiasong Wu, et al. Thin semantics enhancement via high-frequency prior rule for thin structures segmentation. In *2021 IEEE International Conference on Data Mining (ICDM)*, pages 1096–1101. IEEE, 2021.
- [9] Edelsbrunner Herbert and Harer John. Computational topology: an introduction. *American Mathematical Soc*, 2010.
- [10] Xiaoling Hu, Fuxin Li, Dimitris Samaras, et al. Topology-preserving deep image segmentation. *Advances in neural information processing systems*, 32, 2019.
- [11] Daniel P Huttenlocher, Gregory A. Klanderman, and William J Rucklidge. Comparing images using the hausdorff distance. *IEEE Transactions on pattern analysis and machine intelligence*, 15(9):850–863, 1993.
- [12] Qiangguo Jin, Zhaopeng Meng, Tuan D. Pham, et al. Dunet: A deformable network for retinal vessel segmentation. *Knowledge-Based Systems*, 178:149–162, 2019.
- [13] Brian S. Ko, James D. Cameron, Ravi K. Munnur, et al. Non-invasive ct-derived ffr based on structural and fluid analysis. 10(6):663–673, 2017.
- [14] Bin Kong, Xin Wang, Junjie Bai, et al. Learning tree-structured representation for 3d coronary artery segmentation. *Computerized Medical Imaging and Graphics*, 80:101688, 2020.
- [15] Yang Li, Yue Zhang, Jing-Yu Liu, et al. Global transformer and dual local attention network via deep-shallow hierarchical feature fusion for retinal vessel segmentation. *IEEE Transactions on Cybernetics*, pages 1–14, 2022.
- [16] James K. Min, Jonathon Leipsic, Michael J. Pencina, et al. Diagnostic Accuracy of Fractional Flow Reserve From Anatomic CT Angiography. *JAMA*, 308(12):1237–1245, 09 2012.
- [17] Volodymyr Mnih. *Machine learning for aerial image labeling*. University of Toronto (Canada), 2013.
- [18] Lei Mou, Yitian Zhao, Huazhu Fu, et al. Cs2-net: Deep learning segmentation of curvilinear structures in medical imaging. *Medical Image Analysis*, 67:101874, 2021.
- [19] James R Munkres. *Elements of algebraic topology*. CRC press, 2018.
- [20] Xiaoming Qi, Yuting He, Guanyu Yang, et al. Mvsgan: Spatial-aware multi-view cmr fusion for accurate 3d left ventricular myocardium segmentation. *IEEE Journal of Biomedical and Health Informatics*, 26(5):2264–2275, 2021.
- [21] Xiaoming Qi, Guanyu Yang, Yuting He, et al. Contrastive re-localization and history distillation in federated cmr segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 256–265, 2022.
- [22] Yaolei Qi, Han Xu, Yuting He, et al. Examinee-examiner network: Weakly supervised accurate coronary lumen segmentation using centerline constraint. *IEEE Transactions on Image Processing*, 30:9429–9441, 2021.
- [23] Michiel Schaap, Coert T Metz, Theo van Walsum, et al. Standardized evaluation methodology and reference database for evaluating coronary artery centerline extraction algorithms. *Medical Image Analysis*, 13(5):701–714, 2009.
- [24] Suprosanna Shit, Johannes C. Paetzold, Anjany Sekuboyina, et al. cldice - a novel topology-preserving loss function for tubular structure segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 16560–16569, June 2021.
- [25] Joes Staal, Michael D Abràmoff, Meindert Niemeijer, et al. Ridge-based vessel segmentation in color images of the retina. *IEEE transactions on medical imaging*, 23(4):501–509, 2004.
- [26] Abdel Aziz Taha and Allan Hanbury. Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool. *BMC medical imaging*, 15(1):1–28, 2015.
- [27] Christopher Tralie, Nathaniel Saul, and Rann Bar-On. Ripser.py: A lean persistent homology library for python. *The Journal of Open Source Software*, 3(29):925, Sep 2018.
- [28] Dong Wang, Zhao Zhang, Ziwei Zhao, et al. Pointscluster: Point set representation for tubular structure extraction, 2022.

- [29] Yan Wang, Xu Wei, Fengze Liu, et al. Deep distance transform for tubular structure segmentation in ct scans. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.
- [30] Chi-Chong Wong and Chi-Man Vong. Persistent homology based graph convolution network for fine-grained 3d shape segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 7098–7107, October 2021.
- [31] Xin Yang, Zhiqiang Li, Yingqing Guo, et al. DCU-net: a deformable convolutional neural network based on cascade U-net for retinal vessel segmentation. *Multimedia Tools and Applications*, 81(11):15593–15607, may 2022.
- [32] Fisher Yu, Vladlen Koltun, and Thomas Funkhouser. Dilated residual networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.
- [33] Xiao Zhang, Jingyang Zhang, Lei Ma, et al. Progressive deep segmentation of coronary artery via hierarchical topology learning. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022*, pages 391–400, Cham, 2022. Springer Nature Switzerland.
- [34] Chunhui Zhao, Wenxiang Zhu, and Shou Feng. Superpixel guided deformable convolution network for hyperspectral image classification. *IEEE Transactions on Image Processing*, 31:3838–3851, 2022.
- [35] Gangming Zhao, Kongming Liang, Chengwei Pan, et al. Graph Convolution Based Cross-Network Multi-Scale Feature Fusion for Deep Vessel Segmentation. *IEEE Transactions on Medical Imaging*, pages 1–1, 2022.
- [36] K. Zuiderveld. Contrast limited adaptive histogram equalization. *Graphics Gems*, pages 474–485, 1994.

7. Appendix–Implementation details

In this section, we supplement the implementation details that are not covered in Section 4.

7.1. DRIVE

The DRIVE dataset contains 40 color images of the fundus retina, where seven images show signs of mild diabetic retinopathy. The resolution of each image is 565×584 . The DRIVE dataset contains manually annotated labels from two experts. In our experiments, we use the annotation results of the first expert as the training and validation of all models.

Data pre-processing: The DRIVE dataset provides color images with three channels. The contrast between the thin tubular vessels and the background is low, which is quite difficult for the deep learning model to discriminate. It will affect the accuracy of model segmentation. Therefore, pre-processing the original data is necessary to improve the model’s performance. As shown in Figure 7, there is a clear difference in the representation of blood vessels on different channels. The comparison reveals that the blood vessel morphology is most evident on channel G. Therefore, our

model selects the images under this channel. To improve the contrast between the blood vessels and the background, we use the CLAHE[36] algorithm to adjust the image’s contrast so that the model will focus more clearly on the structure of the blood vessels.

Experimental parameters: Due to the limited amount of data provided by the dataset, it is necessary to further sample and enlarge the input data. In our experiments, the data augmentation consisted of translation, rotation, mirroring, flipping, and random cropping. All data enhancements are performed simultaneously during the training process. The random crop size of 224×224 is used as a fixed size to feed into the model training.

Model training parameter setting: In training, the Adam optimizer is selected with default parameters. The learning rate is set to 0.0001, and the epoch is set to 200. The experimental program is based on Python 3.7 and Pytorch. It uses a Gtx2080ti graphics card to train and test.

7.2. Massachusetts road

The Massachusetts road dataset contains 1108 color images for training, 14 color images for verifying, and 49 color images for testing. The resolution of each image is 1500×1500 .

Experimental parameters: The color images provided by the Massachusetts road dataset contain three channels, and we use data information from all three channels simultaneously. Since the morphology of the data varies considerably, we used data enhancement. In our experiments, the data augmentation consisted of translation, rotation, mirroring, flipping, and random cropping. All data enhancements are performed simultaneously during the training process. The random crop size of 256×256 is used as a fixed size to feed into the model training.

Model training parameter setting: In training, the Adam optimizer is selected with default parameters. The learning rate is set to 0.0001, and the epoch is set to 400. The experimental program is based on Python 3.7 and Pytorch. It uses a Gtx2080ti graphics card to train and test.

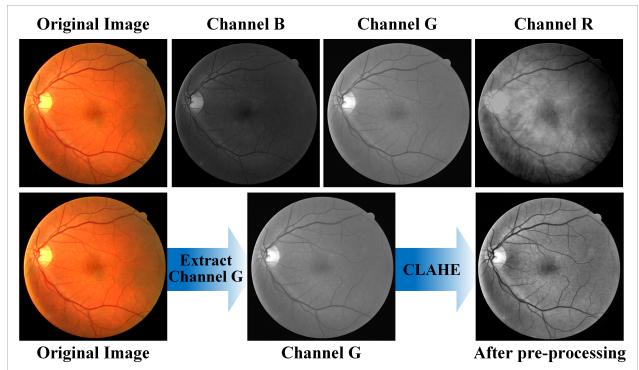


Figure 7. Visual effects before and after pre-processing.

7.3. Cardiac CCTA

The Cardiac CCTA dataset contains 288 images with contrast agents showing the coronary arteries. 192 images are used in our experiments as the training data and the remaining 96 images are used as the testing data. The original image is 512×512 per slice, with 200 to 500 slices per image. The coronary lumen boundary is delineated by a trained researcher and verified by a clinician. The region of interest (ROI) is extracted and these ROIs are cropped into small 3D patches of $128 \times 128 \times 96$ online due to the limitation of GPU memory.

Model training parameter setting: In training, the Adam optimizer is selected with default parameters. The learning rate is set to 0.0001, and the epoch is set to 200. The experimental program is based on Python 3.7 and Pytorch. It uses a Gtx3090 graphics card to train and test.

8. Appendix–Analysis of our DSCNet

Our DSConv dynamically adapts the shape to tubular structures, and the attention well fits the target. (1) **Adapt to the shape of tubular structures.** Top of Figure 8 shows the convolution kernel’s positions and shape. Visualization results show that our DSConv adapts well to tubular structures and maintains the shape, while the deformable convolution wanders outside the target. (2) **Focus on the locations of tubular structures.** Bottom of Figure 8 shows the heatmap of the attention on the given point. Results show that the brightest regions of our DSConv are concentrated in the tubular structures, which represents that our DSConv is more sensitive to tubular structures.

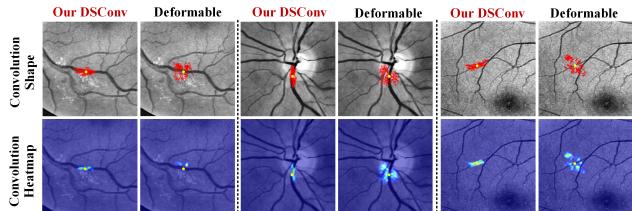


Figure 8. **Top:** We superimpose 3 layers with a total of 729 points (red) on each image to show the convolution kernel’s position and shape for a given point (yellow). **Bottom:** Heatmap shows the region of interest for each convolution.

9. Additional Qualitative Results

In this section, we provide additional qualitative analysis of our methods (Figure 9).

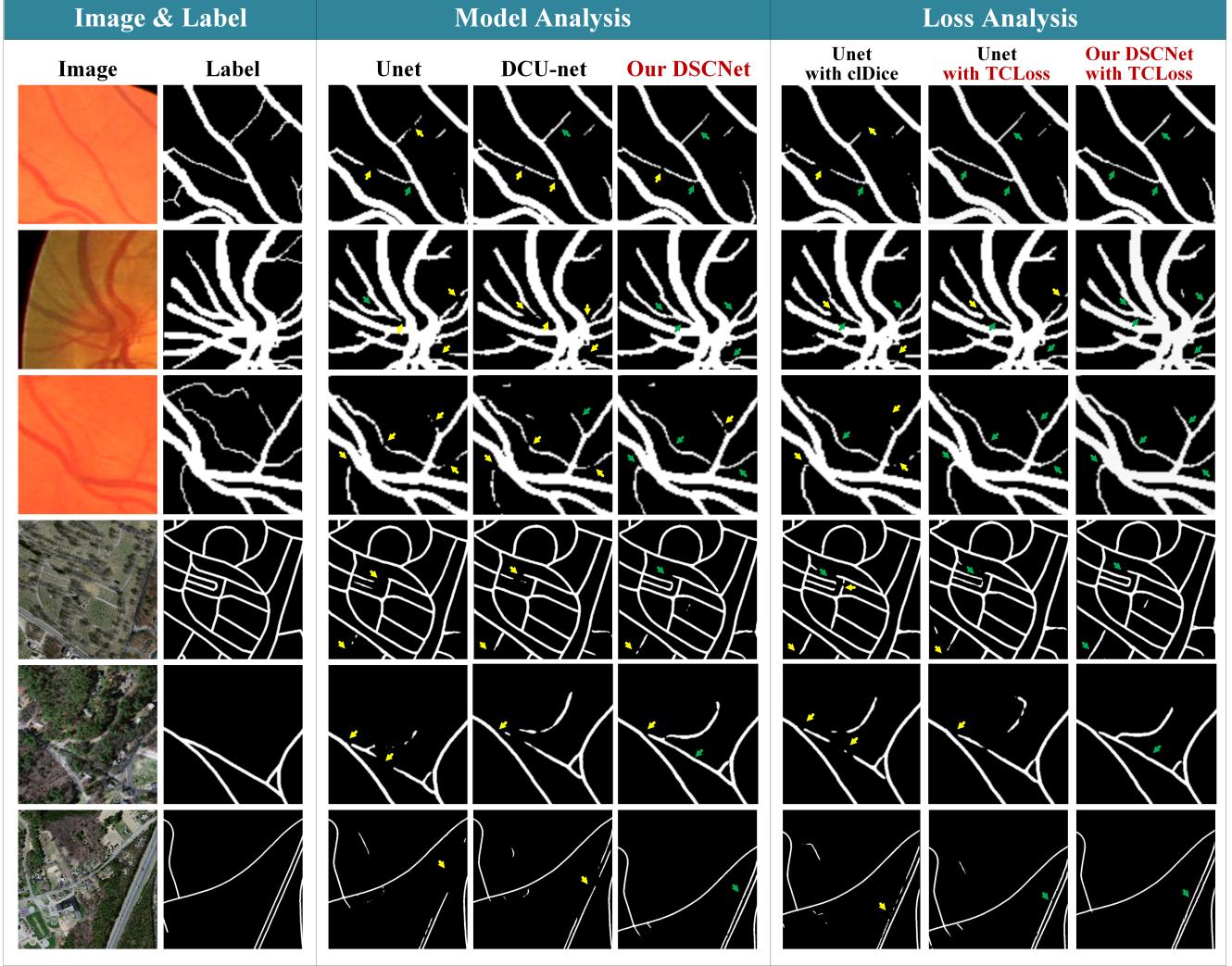


Figure 9. Qualitative results. To verify the performance of our method more objectively and efficiently, we selected representative hard-to-segment regions from each dataset. From top to bottom, we show three rows of results for the DRIVE dataset and the Massachusetts road dataset. From left to right, we show the original image, groundtruth, and the results from classical UNet, DCU-net, our DSCNet, UNet with cIDice, UNet with our proposed TCloss, and our DSCNet with TCloss. The results indicate that our DSCNet and TCloss outperform the other models regarding segmentation accuracy and topological continuity. The yellow arrows indicate the areas where the segmentation is broken, while the green arrows indicate areas where the segmentation is performing well.