Advances in Artificial Intelligence, Big Data and Algorithms
G. Grigoras and P. Lorenz (Eds.)

doi:10.3233/FAIA230870

Directional Feature Maps Learning for Interpretable Cardiac Image Segmentation

Tiantian YANG[1] and Yu LI
Wuhan Textile University, WuHan 430200, China

Abstract. One of the crucial duties in clinical surgery is medical image segmentation. Deep learning and fully convolutional neural networks are currently being employed for image segmentation and have achieved cutting-edge outcomes on numerous publicly available benchmark datasets. Shape is more significant than visual texture, and typical CNNs are not robust and interpretable. As a result, we suggest the Shape Attention Net (DFSNet), a new architecture that emphasizes robustness and interpretability. This architecture employs a dual-stream method to simultaneously record rich shape-related information and conventional texture streams. A dual attention decoder module is used to learn multi-resolution saliency maps, and a direction field distance prediction pipeline is utilized for fixing segmentation detail. On two sizable public datasets for heart MRI image segmentation, SUN09 and ACDC, our technique produced cutting-edge results.

Keywords. Deep learning, CNN, dual-stream method, a direction field distance prediction pipeline

1. Introduction

The heart is one of the most vital organs in the human body, but heart disease threatens the lives of many people. The medical community's standard for non-invasive assessment of cardiovascular function is cardiovascular magnetic resonance imaging (CMR) [1]. Compared with other techniques, CMR has the advantages of high spatial resolution and non-ionizing radiation, which are crucial for the diagnosis of cardiovascular diseases. Therefore, it is essential to achieve automatic and accurate segmentation of the heart region. The advancement of deep learning has made machine learning widely used in medical imaging, and convolutional neural networks have been proven to be very effective in medical image segmentation.

In the field of medical image segmentation, U-Net [2] is one of the most effective and widely used technologies at present, while nnU-Net [3] mixes 2D and 3D U-Net models to achieve the most cutting-edge performance in heart segmentation. In order to solve the problem of MRI artifacts causing interclass confusion and orientation information [4], we proposed a new method to improve segmentation feature maps by using orientation information. This technique shows strong robustness between data sets and significantly reduces inconsistencies within classes and ambiguities between classes. The effectiveness of this method is verified by experiments.

[1] Corresponding Author. TianTian YANG, Wuhan Textile University, China;
Email: 2115363047@mail.wtu.edu.cn.

2. Method

We propose a new interpretable architecture for natural image segmentation and medical image segmentation called the Directional Feature Shape Flow Net (DFSNet) architecture. DFSNet consists of two parts: a fine segmentation part and a direction field distance prediction part. The direction field learning module and the feature integration module make up the direction field prediction component. The fault segmentation network's shared direction field feature is learned by the direction learning module, and the final segmentation result is obtained by fusing the learned direction field feature with the initial segmentation feature.

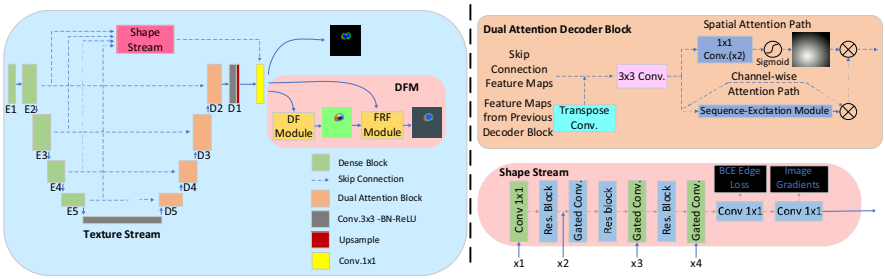

Figure 1. We propose a directional feature-shaped flow network (DFSNet). The model consists of a shape flow for processing boundary information, a texture flow, and pipelines for predicting directional field distances. The directional distance prediction field is introduced under the fine and rough segmentation of shape flow and texture flow, and feature reconstruction and fusion are used to obtain the final segmentation results.

2.1 Finely Segmented Parts

The fine segmentation part of DFSNet consists of two flows: the texture flow and the gated shape flow. The texture flow adopts a U-shaped framework. The encoder part is composed of dense blocks from DenseNet-121, and the decoder part is composed of dual-attention decoder blocks. The texture flow learns dense pixel information and features. Based on generating more precise segmentation, the gated shape flow allows the model to learn object shapes.

2.1.1 Gated Convolution Layer and a Gated Shape Stream's Output

The feature map x is processed using the normalized 1x1 convolution function $C_{1x1}(x)$ and the residual block function R(x). The residual block consists of two normalized 3x3 convolutions and a skip connection. The result of $C_{1x1}(x)$ is a feature map with decreased channels and the same spatial dimensions as x. The attention map for the boundary is determined by a gating convolution layer utilizing data from the shape stream and texture stream. The layer in the shape stream is marked by the letter l, while the encoder block that produces the texture stream feature map is denoted by the letter t. Together, these two descriptors stand for the shape stream feature map and the texture stream feature map. Bilinear interpolation is used on T_t in order to adjust it if necessary to match the dimension of S_l. Pooling layers shouldn't be employed in the form stream to get precise shape limits. In the shape stream, each residual block is designated as a layer. The calculation method for l is as follows:

$$\hat{S}_l = S_l \otimes (\sigma(C_{1x1}(S_l \parallel C_{1x1}(T_t)))) \quad (1)$$

The $\parallel$ denotes the concatenation of feature maps across channels, and the σ symbolizes the sigmoid function. $\otimes$is the Hadamard product. You can determine the feature map of the layer beneath the form as follows:

$$S_{l+1} = R(\hat{S}_l) \quad (2)$$

$S_l + 1$ can be improved using the same method.

Our model is to accurately learn the class's shape and texture. The initial image and the result of the gated shape stream are cascaded through the appropriate channel classes in order to anticipate the shape feature map. The feature mapping of the texture flow is then coupled with the output of the gated shape flow after being normalized by a 3x3 convolution.

2.1.2 Double Attention Decoder Block

In order to increase interpretability, the double attention decoding block contains two interpretability components: an interpretable spatial attention path after a normalized 3x3 convolution and an interpretable channel attention path that has been shown to improve performance by Hu et al. [5].

Enter the spatial attention path through C channels. When a single channel's pixel values are mapped to the [0, 1] range using a sigmoid, the output is F_s'. To match the output dimension of the channel attention path, F_s' is stacked C times along the channel axis to obtain F_s. Finally, element-wise multiplication is performed. The scaling factors are values between 0 and 1 and are used to scale each channel in the skip-connected feature map by its corresponding scaling factor, resulting in the scaled feature map F_c. Channel attention and spatial attention are decoded as double attention blocks in output F:

$$\mathrm{F} = (\mathrm{F_s} + 1) \otimes \mathrm{F_c} \quad (3)$$

Operator $\otimes$ represents the Hadamard product. Adding +1 will restrict the spatial attention mechanism to only amplify features without zeroing out potentially valuable features in subsequent convolutions. Additionally, this operation ensures that attention weights are always greater than zero.

2.2 Direction Field Distance Prediction Part

In image segmentation, a common issue is the inconsistency both between and within classes. Additionally, segmentation models often only learn individual representations and therefore do not constrain the relationships between pixels. To address these problems, we employ a simple yet effective approach that utilizes the directional relationships between pixels. In order to extract the common feature direction field in the network, as illustrated in Figure 1, we add a direction field learning module based on a fine segmentation network. The network's initial fine segmentation feature is coupled with the learned direction feature in the feature integration module to get the final segmentation result.

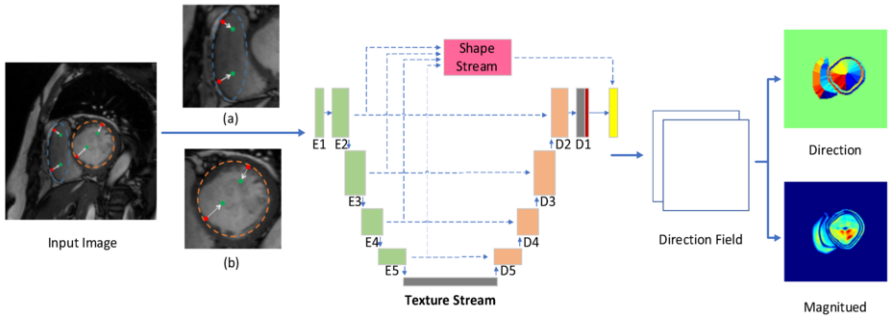

Figure 2. The DF module schematic diagram is used to predict a new direction field based on the two-dimensional vector of the image. (a) and (b) represent the vectors from the nearest boundary pixels to the current pixel. The direction and magnitude information of the direction field on the right can be calculated and visualized.

We start by describing the direction field symbol, which is depicted in Figure 2(a–b). On the edge of the heart tissue, we identify each foreground pixel p that corresponds to the closest background pixel b, and we then distance-normalize the direction vector from b to p. Formally, each pixel's direction field (DF) in the image domain Ω is as follows:

$$DF(p) = \begin{cases} \frac{\overrightarrow{bp}}{|\overrightarrow{bp}|}, & p \in foreground, \\ (0,0), & otherwise. \end{cases} \tag{4}$$

Figure 2 illustrates the module we used to understand the aforementioned direction field. The input channel characteristic, which is 64, is the characteristic of the section of the network that is finely segmented. A two-channel direction field is what it outputs.

By using the initial feature graph $F^N \in R^{CxHxW}$ and the predicted direction field $F^O \in R^{CxHxW}$, the improved feature graph $\mathrm{DF} \in \mathrm{R}^{2\mathrm{xHxW}}$ is obtained step by step. The whole program form is as follows:

$$\forall_p \in \Omega, \mathrm{F^k(p)} = \mathrm{F^{(k-1)}}(\mathrm{p_x} + \mathrm{DFF(p)_x}, \mathrm{p_y} + \mathrm{DF(p)_y}) \tag{5}$$

px (resp.py) denotes the x (resp.py) coordinate of pixel p, and $1 \leq k \leq N$ is the current step, N is the total step (set to 5 if no additional instructions are given), and so forth.

After the correction process described above, we link F_N and F_O and use the final classifier to forecast the final segmentation of the heart on the connected feature map.

3. Experiments

3.1 Datasets and Implementation Details

The Automatic Cardiac Diagnostic Challenge (ACDC [6]) dataset consists of 150 cineMR images of patients.Each model trained for 180 epochs with a batch size of 10.

The SUN09[7] dataset contains separate training datasets for each of the two categories (endocardial and epicardial). The model trained for 120 epochs with a batch size of 4.

In this paper, the experiments were carried out on a Tian RTX GPU with 24 GB of memory. Z-score normalization was performed on each image slice. We applied various data augmentations . In the following sections, we will present the experimental results.

3.2 Experimental Results

Comparative experiments were conducted under the experimental settings of the paper [8]. As shown in Table 1 below, we evaluate the effectiveness of the network using several models. The performance of our model has significantly improved without pre-training weight, as shown in the table, particularly in MYO. This is attributable to the combination of fine segmentation and direction distance prediction fields, which significantly increase the network's capacity to detect and correct boundary details. The dice scores of Endocardium and Epicardium in Table 2 under identical experimental settings showed novel results in comparison to existing network segmentation models.

Table 1. ACDC test set results.

Model	LV	RV	MYO
UNet	0.910	0.901	0.888
ResUNet	0.921	0.904	0.891
SAUNet	0.925	0.914	0.887
UNetDF	0.935	0.920	0.903

Table 2. Test set Dice scores for SUN09.

Model	Endocardium	Epicardium
SAUNet	0.933	0.941
UNetDF	0.943	0.949

4. Conclusion

This article investigates a new interpretable medical image segmentation model that applies directional distance information and utilizes a novel U-shaped framework for simple and effective segmentation of cardiac MRI. The technique is more interpretable than earlier techniques that used built-in saliency maps and learned robust form characteristics of objects. The segmentation feature mapping is improved, leading to increased segmentation accuracy with the help of directional information. The effectiveness and strong generalizability of the strategy are demonstrated by the experimental results.

References

[1] Al Arif S M, Knapp K, Slabaugh G. Shape-aware deep convolutional neural network for vertebrae segmentation[C]//International Workshop on Computational Methods and Clinical Applications in Musculoskeletal Imaging. Springer, Cham, 2018: 12-24.

[2] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]//Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer International Publishing, 2015: 234-241.
[3] Isensee F, Petersen J, Kohl S A A, et al. nnu-net: Breaking the spell on successful medical image segmentation[J]. arXiv preprint arXiv:1904.08128, 2019, 1(1-8): 2.
[4] Sun, J., Darbeha, F., Zaidi, M., & Wang, B. (2020). SAUNet: Shape Attentive U-Net for Interpretable Medical Image Segmentation. *ArXiv, abs/2001.07645*.
[5] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7132-7141.
[6] Bernard O, Lalande A, Zotti C, et al. Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: is the problem solved?[J]. IEEE transactions on medical imaging, 2018, 37(11): 2514-2525.
[7] Radau P, Lu Y, Connelly K, et al. Evaluation framework for algorithms segmenting short axis cardiac MRI[J]. The MIDAS Journal, 2009.
[8] Cheng, F., Chen, C., Wang, Y., Shi, H., Cao, Y., Tu, D., Zhang, C., & Xu, Y. (2020). Learning Directional Feature Maps for Cardiac MRI Segmentation. International Conference on Medical Image Computing and Computer-Assisted Intervention.