Learning Directional Feature Maps
for Cardiac MRI Segmentation

Feng Cheng[1], Cheng Chen[1], Yukang Wang[1], Heshui Shi[2,3], Yukun Cao[2,3],
Dandan Tu[4], Changzheng Zhang[4], and Yongchao Xu[1(✉)]

[1] School of Electronic Information and Communications,
Huazhong University of Science and Technology, Wuhan, China
`yongchaoxu@hust.edu.cn`
[2] Department of Radiology, Union Hospital, Tongji Medical College, Huazhong
University of Science and Technology, Wuhan, China
[3] Hubei Province Key Laboratory of Molecular Imaging, Wuhan, China
[4] HUST-HW Joint Innovation Lab, Wuhan, China

Abstract. Cardiac MRI segmentation plays a crucial role in clinical
diagnosis for evaluating personalized cardiac performance parameters.
Due to the indistinct boundaries and heterogeneous intensity distribu-
tions in the cardiac MRI, most existing methods still suffer from two
aspects of challenges: inter-class indistinction and intra-class inconsis-
tency. To tackle these two problems, we propose a novel method to exploit
the directional feature maps, which can simultaneously strengthen the
differences between classes and the similarities within classes. Specifi-
cally, we perform cardiac segmentation and learn a direction field point-
ing away from the nearest cardiac tissue boundary to each pixel via
a direction field (DF) module. Based on the learned direction field,
we then propose a feature rectification and fusion (FRF) module to
improve the original segmentation features, and obtain the final seg-
mentation. The proposed modules are simple yet effective and can be
flexibly added to any existing segmentation network without excessively
increasing time and space complexity. We evaluate the proposed method
on the 2017 MICCAI Automated Cardiac Diagnosis Challenge (ACDC)
dataset and a large-scale self-collected dataset, showing good segmen-
tation performance and robust generalization ability of the proposed
method. The code is publicly available at https://github.com/c-feng/
DirectionalFeature.

Keywords: Cardiac segmentation · Deep learning · Direction field

1 Introduction

Cardiac cine Magnetic Resonance Imaging (MRI) segmentation is of great impor-
tance in disease diagnosis and surgical planning. Given the segmentation results,

---

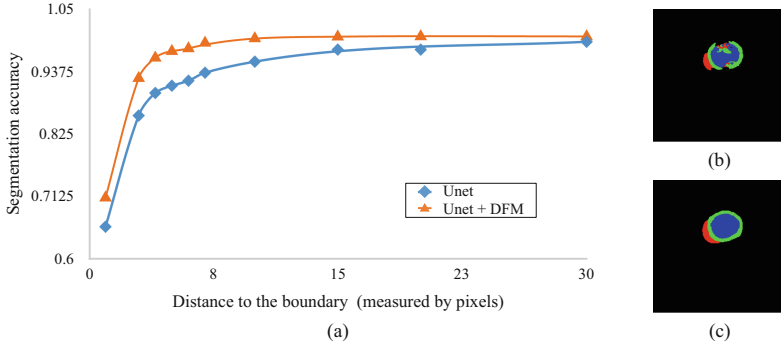F. Cheng and C. Chen—Equal contributions.

Fig. 1. (a) is the comparison of segmentation accuracy between U-Net and the proposed method at different distances from pixel to boundary; (b) and (c) are the segmentation visualizations of U-Net and the proposed method, respectively. Compared with the original U-Net, the proposed method effectively mitigate the problems of inter-class indistinction and intra-class inconsistency.

doctors can obtain the cardiac diagnostic indices such as myocardial mass and thickness, ejection fraction and ventricle volumes more efficiently. Indeed, manual segmentation is the gold standard approach. However, it is not only time-consuming but also suffers from the inter-observer variations. Hence, the automatic cardiac cine MRI segmentation is desirable in the clinic.

In the past decade, deep convolutional neural networks (CNNs) based methods have achieved great successes in both natural and medical image segmentation. U-Net [11] is one of the most successful and influential method in medical image segmentation. Recent works typically leverage the U-shape networks and can be roughly divided into 2D and 3D methods. 2D methods take a single 2D slice as input while 3D methods utilize the entire volume. nnU-Net [5] adopts the model fusion strategy of 2D U-Net and 3D U-Net, which achieves the current state-of-the-art performance in cardiac segmentation. However, the applicability is somewhat limited since it requires a high cost of memory and computation.

The MRI artifacts such as intensity inhomogeneity and fuzziness may make it indistinguishable between pixels near the boundary, leading to the problem of inter-class indistinction. As depicted in Fig. 1(a), we observe that the cardiac MRI segmentation accuracy drops dramatically for those pixels close to the boundary. Meanwhile, due to the lack of restriction on the spatial relationship between pixels, the segmentation model may produce some anatomical implausible errors (see Fig. 1(b) for an example). In this paper, we propose a novel method to improve the segmentation feature maps with directional information, which can significantly improve the inter-class indistinction as well as cope with the intra-class inconsistency. Extensive experiments demonstrate that the proposed method achieves good performance and is robust under cross-dataset validation.
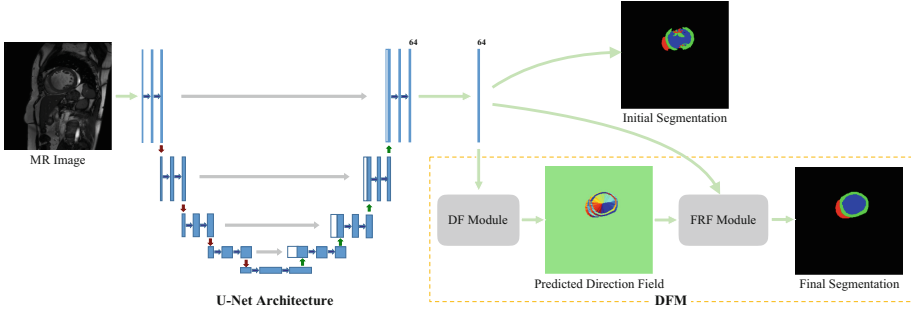
Fig. 2. Pipeline of the proposed method. Given an image, the network predicts an initial segmentation map from U-Net and a direction field (DF) (visualized by its direction information), based on which we reconstruct and fuse the original segmentation features via a feature rectification and fusion (FRF) module to produce the final segmentation.

Recent approaches in semantic segmentation have been devoted to handling the inter-class indistinction and the intra-class inconsistency. Ke *et al.* [6] define the concept of adaptive affinity fields (AAF) to capture and match the semantic relations between neighboring pixels in the label space. Cheng *et al.* [2] explore the boundary and segmentation mask information to improve the inter-class indistinction problem for instance segmentation. Shusil *et al.* [3] propose a multi-task learning framework to perform segmentation along with a pixel-wise distance map regression. This regularization method takes the distance from the pixel to the boundary as auxiliary information to handle the problem of inter-class indistinction. Nathan *et al.* [9] propose an adversarial variational auto-encoder to assure anatomically plausible, whose latent space encodes a smooth manifold on which lies a large spectrum of valid cardiac shapes, thereby indirectly solving the intra-class inconsistency problem.

Directional information has been recently explored in different vision tasks. For instance, TextField [14] and DeepFlux [12] learn similar direction fields on text areas and skeleton context for scene text detection and skeleton extraction, respectively. They directly construct text instances or recover skeletons from the direction field. However, medical images are inherently different from natural images. Such segmentation results obtained directly from the direction field are not accurate for the MRI segmentation task. In this paper, we propose to improve the original segmentation features guided by the directional information for better cardiac MRI segmentation.

2   Method

Inter-class indistinction and intra-class inconsistency are commonly found in both natural and medical image segmentation. Meanwhile, segmentation models

usually learn individual representations and thus lack of restrictions on the relationship between pixels. We propose a simple yet effective method to exploit the directional relationship between pixels, which can simultaneously strengthen the differences between classes and the similarities within classes. The pipeline of the proposed method, termed as DFM, is depicted in Fig. 2. We adopt U-Net [11] as our base segmentation framework. Given an input image, the network produces the initial segmentation map. Meanwhile, we apply a direction field (DF) module to learn the direction field with the shared features from U-Net. A feature rectification feature (FRF) module is then proposed to combine the initial segmentation feature with the learned direction field to generate the final improved segmentation result.

2.1   DF Module to Learn a Direction Field

We first detail the notation of the direction field. As shown in Fig. 3(a, b), for each foreground pixel p, we find its nearest pixel b lying on the cardiac tissue boundary and then normalize the direction vector $\overrightarrow{bp}$ pointing from b to p by the distance between b and p. We set the background pixels to $(0,0)$. Formally, the direction field DF for each pixel p in the image domain Ω is given by:

$$DF(p) = \begin{cases} \frac{\overrightarrow{bp}}{|\overrightarrow{bp}|} & p \in foreground, \\ (0,0) & otherwise. \end{cases} \tag{1}$$

We propose a simple yet effective DF module to learn the above direction field, which is depicted in Fig. 3. This module is made up of a 1×1 convolution, whose input is the 64-channel feature extracted by U-Net and output is the two-channel direction field. It is noteworthy that we can obtain the ground truth of the direction field from the annotation easily by distance transform algorithm.

2.2   FRF Module for Feature Rectification and Fusion

The direction field predicted by the DF module reveals the directional relationship between pixels and provides a unique direction vector that points from the boundary to the central area for each pixel. Guided by these direction vectors, we propose a Feature Rectification and Fusion (FRF) module to utilize the characteristics of the central area to rectify the errors in the initial segmentation feature maps step by step. As illustrated in Fig. 4, the N-steps improved feature maps $F^N \in \mathbb{R}^{C \times H \times W}$ are obtained with the initial feature maps $F^0 \in \mathbb{R}^{C \times H \times W}$ and the predicted direction field $DF \in \mathbb{R}^{2 \times H \times W}$ step by step. Concretely, the improved feature of the pixel p is updated iteratively by the feature of the position that $DF(p)$ points to, which is calculated by the bilinear interpolation. In other words, $F(p)$ is rectified by the features of the central area gradually. The whole procedure is formalized as below:

$$\forall p \in \Omega, \ F^k(p) \ = \ F^{(k-1)}\big(p_x + DF(p)_x, p_y + DF(p)_y\big), \tag{2}$$
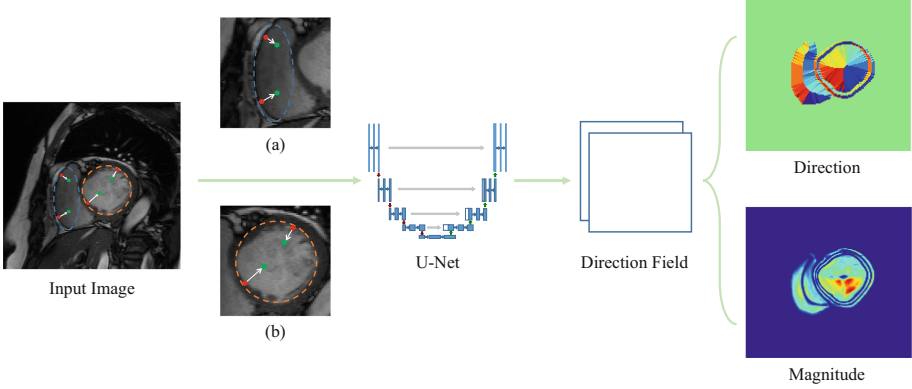
Fig. 3. Illustration of the DF module. Given an image, the network predicts a novel direction field in terms of an image of two-dimensional vectors. (a) and (b) show the vector from the nearest boundary pixel to the current pixel. We calculate and visualize the direction and magnitude information of the direction field on the right side.

where $1 \leq k \leq N$ denotes the current step, N is the total steps (set to 5 if not stated otherwise), and p_x (*resp.* p_y) represents the x (*resp.* y) coordinate of the pixel p.

After performing the above rectification process, we concatenate F^N with F^0, and then apply the final classifier on the concatenated feature maps to predict the final cardiac segmentation.

2.3   Training Objective

The proposed method involves loss function on the initial segmentation L_{CE}^i, final segmentation L_{CE}^f, and direction field L_{DF}. We adopt the general cross-entropy L_{CE} as the segmentation loss to encourage class-seperate feature, which is commonly used in semantic segmentation. Formally, L_{CE} is given by $L_{CE} = -\sum_i y_i log(\hat{y}_i)$, where y_i and $\hat{y}_i$ denote the ground truth and the prediction, respectively. For the loss to supervise the direction field learning, we choose the L_2-norm distance and angle distance as the training objective:

$$L_{DF} = \sum_{p \in \Omega} w(p)(||DF(p) - \hat{DF}(p)||_2 + \alpha \times ||cos^{-1}\langle DF(p), \hat{DF}(p)\rangle||^2) \quad (3)$$

where $\hat{DF}$ and DF denote the predicted direction field and the corresponding ground truth, respectively, α is a hyperparameter to balance the L_2-norm distance and angle distance, and is to 1 in all experiments, and $w(p)$ represents the weight on pixel p, which is calculated by:

$$w(p) = \begin{cases} \frac{\sum_{i=1}^{N_{cls}} |C_i|}{N_{cls} \cdot |C_i|} & p \in C_i, \\ 1 & otherwise, \end{cases} \quad (4)$$
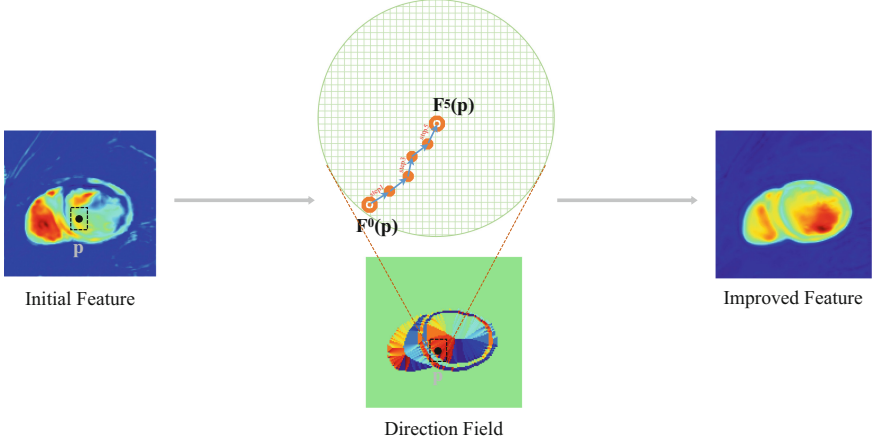
Fig. 4. Schematic illustration of the FRF module. We use the learned direction field to guide the feature rectification. The feature for pixels close to the boundary is rectified with the feature (having the same semantic category) far from the boundary.

where $|C_i|$ denotes the total number of pixels with label i and N_{cls} is the number of classes. The overall loss L combines L_{CE} and L_{DF} with a balance factor $\lambda = 1$:

$$L = L_{CE}^i + L_{CE}^f + \lambda L_{DF} \tag{5}$$

3   Experiments

3.1   Datasets and Evaluation Metrics

Automatic Cardiac Diagnosis Challenge (ACDC) Dataset contains cine-MR images of 150 patients, split into 100 train images and 50 test images. These patients are divided into 5 evenly distributed subgroups: normal, myocardial infarction, dilated cardiomyopathy, hypertrophic cardiomyopathy and abnormal right ventricle, available as a part of the STACOM 2017 ACDC challenge. The annotations for the 50 test images are hold by the challenge organizer. We also further divide the 100 training images into 80% training and 20% validation with five non-overlapping folds to perform extensive experiments.

Self-collected Dataset consists of more than 100k 2D images that we collected from 531 patient cases. All the data was labeled by a team of medical experts. The patients are also divided into the same 5 subgroups as ACDC. A series of short axis slices cover LV from the base to the apex, with a thickness of 6 mm and a flip angle of 80°. The magnitude field strength of images is 1.5 T and the spatial resolution is 1.328 mm^2/pixel. We also split all the patients into 80% training and 20% test.

Evaluation Metrics: We adopt the widely used 3D Dice coefficient and Hausdorff distance to benchmark the proposed method.

3.2   Implementation Details

The network is trained by minimizing the proposed loss function in Eq. (5) using ADAM optimizer [8] with the learning rate set to 10^{-3}. The network weights are initialized with [4] and trained for 200 epochs. Data augmentation is applied to prevent over-fitting including: 1) random translation with the maximum absolute fraction for horizontal and vertical translations both set to 0.125; 2) random rotation with the random angle between $-180°$ and $180°$. The batch size is set to 32 with the resized 256×256 inputs.

Table 1. Performance on ACDC dataset (train/val split) and Self-collected dataset.

	Methods	Dice coefficient				Hausdorff distance (mm)			
		LV	RV	MYO	Mean	LV	RV	MYO	Mean
ACDC dataset	U-Net	0.931	0.856	0.872	0.886	24.609	30.006	14.416	23.009
	Ours	**0.949**	**0.888**	**0.911**	**0.916**	**3.761**	**6.037**	**10.282**	**6.693**
Self-collected dataset	U-Net	0.948	0.854	0.906	0.903	2.823	3.691	2.951	3.155
	Ours	**0.949**	**0.859**	**0.909**	**0.906**	**2.814**	**3.409**	**2.683**	**2.957**

Table 2. Results on the ACDC leaderboard (sorted by Mean Hausdorff Distance).

Rank	User	Mean DICE	Mean HD (mm)
1	Fabian Isensee [5]	0.927	7.8
2	Clement Zotti [15]	0.9138	9.51
3	**Ours**	**0.911**	**9.92**
4	Nathan Painchaud [9]	0.911	9.93
5	Christian Baumgartner [1]	0.9046	10.4
6	Jelmer Wolterink [13]	0.908	10.7
7	Mahendra Khened [7]	0.9136	11.23
8	Shubham Jain [10]	0.8915	12.07

3.3   In-Dataset Results

We first evaluate the proposed method on the ACDC dataset (train/val split described in Sect. 3.1) and the self-collected dataset. Table 1 presents the performance of the proposed DFM on two datasets. LV, RV and MYO represent the left ventricle, right ventricle and myocardium, respectively. Our approach consistently improves the baseline (U-Net), demonstrating its effectiveness. To further make comparison with the current state-of-the-arts methods, we submit the results to the ACDC leadboard. As shown in Table 2, compared with

Table 3. Comparisons with different methods aiming to alleviate the inter-class indistinction and intra-class inconsistency on ACDC dataset (train/val split).

Methods	Dice coefficient				Hausdorff distance (mm)			
	LV	MYO	RV	Mean	LV	MYO	RV	Mean
U-Net	0.931	0.856	0.872	0.886	24.609	30.006	14.416	23.009
AAF [6]	0.928	0.853	0.891	0.891	13.306	14.255	13.969	13.844
DMR [3]	0.937	0.880	0.892	0.903	7.520	9.870	12.385	9.925
U-Net+DFM	**0.949**	**0.888**	**0.911**	**0.916**	**3.761**	**6.037**	**10.282**	**6.693**
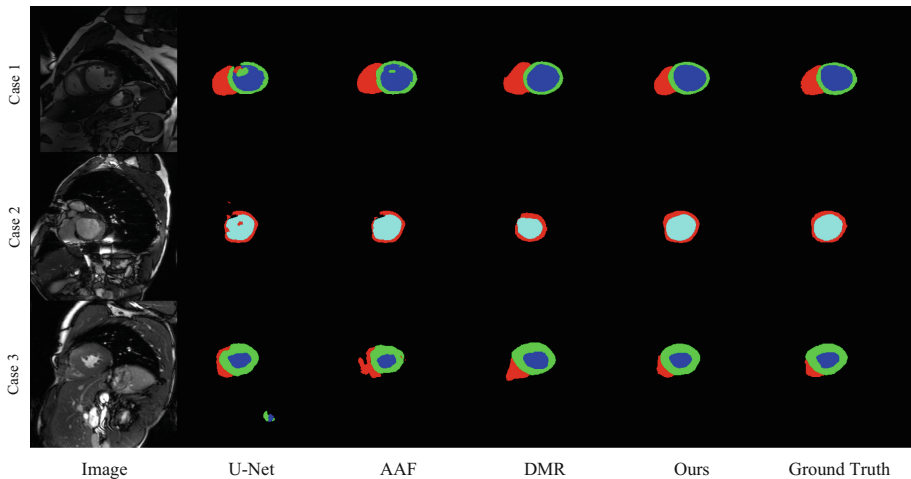
Fig. 5. Qualitative comparison on ACDC dataset. The proposed DFM achieves more accurate results along with better smoothness and continuity in shape.

those methods that rely on well-designed networks (*e.g.* 3D network in [5] and Grid-like CNN in [15]) or multi-model fusion, the proposed DFM achieves competitive performance with only two simple yet effective modules added to the baseline U-Net. We also compare our approach with other methods dedicated to alleviate the inter-class indistinction and intra-class inconsistency. As depicted in Table 3, the proposed DFM significantly outperforms these methods. Some qualitative comparisons are given in Fig. 5.

3.4   Cross Dataset Evaluation and Ablation Study

To analyze the generalization ability of the proposed DFM, we performed a cross-dataset segmentation evaluation. Results listed in Table 4 show that the proposed DFM can consistently improve the cross-dataset performance compared with the original U-Net, validating its generalization ability and robustness. We also conduct ablation study on ACDC dataset to explore how the number of steps N

Table 4. Cross-dataset evaluation compared with the original U-Net.

Methods	ACDC to self-collected dataset		Self-collected dataset to ACDC	
	Mean dice	Mean HD (mm)	Mean dice	Mean HD (mm)
U-Net	0.832	25.553	0.803	4.896
U-Net+DFM	**0.841**	**17.870**	**0.820**	**4.453**

Table 5. Ablation study on the number of steps N on ACDC dataset.

Number of steps	0	1	3	**5**	7
Mean dice	0.910	0.913	0.914	**0.916**	0.910
Mean HD (mm)	10.498	17.846	9.026	**6.693**	13.880

involved in the FRF module influences the performance. As shown in Table 5, the setting of $N = 5$ gives the best performance.

4   Conclusion

In this paper, we explore the importance of directional information and present a simple yet effective method for cardiac MRI segmentation. We propose to learn a direction field, which characterizes the directional relationship between pixels and implicitly restricts the shape of the segmentation result. Guided by the directional information, we improve the segmentation feature maps and thus achieve better segmentation accuracy. Experimental results demonstrate the effectiveness and the robust generalization ability of the proposed method.

References

1. Baumgartner, C.F., Koch, L.M., Pollefeys, M., Konukoglu, E.: An exploration of 2D and 3D deep learning techniques for cardiac MR image segmentation. In: Pop, M., et al. (eds.) STACOM 2017. Lecture Notes in Computer Science, vol. 10663, pp. 111–119. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-75541-0_12
2. Cheng, T., Wang, X., Huang, L., Liu, W.: Boundary-preserving mask R-CNN. In: Proceedings of European Conference on Computer Vision (2020)
3. Dangi, S., Linte, C.A., Yaniv, Z.: A distance map regularized CNN for cardiac cine MR image segmentation. Med. Phys. **46**(12), 5637–5651 (2019)
4. He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: surpassing human-level performance on ImageNet classification. In: Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, pp. 1026–1034 (2015)

5. Isensee, F., Petersen, J., Kohl, S.A., Jäger, P.F., Maier-Hein, K.H.: NNU-net: breaking the spell on successful medical image segmentation. arXiv preprint arXiv:1904.08128 (2019)
6. Ke, T.W., Hwang, J.J., Liu, Z., Yu, S.X.: Adaptive affinity fields for semantic segmentation. In: Proceedings of European Conference on Computer Vision, pp. 587–602 (2018)
7. Khened, M., Kollerathu, V.A., Krishnamurthi, G.: Fully convolutional multi-scale residual DenseNets for cardiac segmentation and automated cardiac diagnosis using ensemble of classifiers. Med. Image Anal. **51**, 21–45 (2019)
8. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
9. Painchaud, N., Skandarani, Y., Judge, T., Bernard, O., Lalande, A., Jodoin, P.M.: Cardiac MRI segmentation with strong anatomical guarantees. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11765, pp. 632–640. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32245-8_70
10. Patravali, J., Jain, S., Chilamkurthy, S.: 2D-3D fully convolutional neural networks for cardiac MR segmentation. In: Pop, M., et al. (eds.) STACOM 2017. LNCS, vol. 10663, pp. 130–139. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-75541-0_14
11. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
12. Wang, Y., Xu, Y., Tsogkas, S., Bai, X., Dickinson, S., Siddiqi, K.: DeepFlux for skeletons in the wild. In: Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, pp. 5287–5296 (2019)
13. Wolterink, J.M., Leiner, T., Viergever, M.A., Išgum, I.: Automatic segmentation and disease classification using cardiac cine MR images. In: Pop, M., et al. (eds.) STACOM 2017. LNCS, vol. 10663, pp. 101–110. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-75541-0_11
14. Xu, Y., Wang, Y., Zhou, W., Wang, Y., Yang, Z., Bai, X.: TextField: learning a deep direction field for irregular scene text detection. IEEE Trans. Image Process. **28**(11), 5566–5579 (2019)
15. Zotti, C., Luo, Z., Lalande, A., Jodoin, P.M.: Convolutional neural network with shape prior applied to cardiac MRI segmentation. IEEE J. Biomed. Health Inf. **23**(3), 1119–1128 (2018)