Effective extraction of ventricles and myocardium objects from cardiac magnetic resonance images with a multi-task learning U-net.
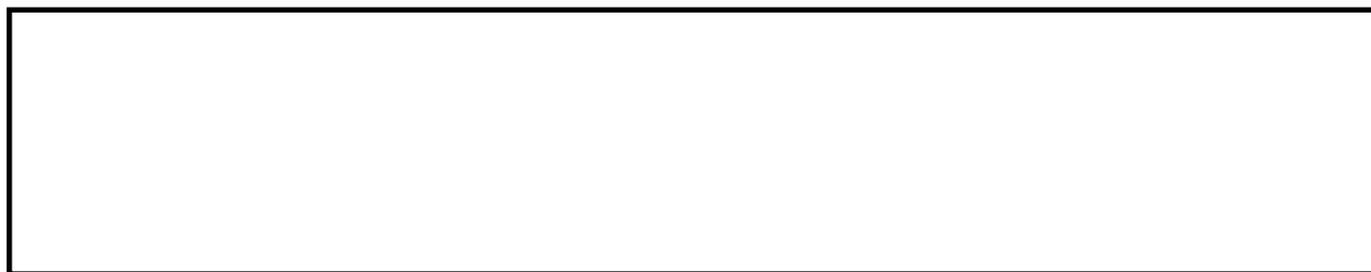
REN, J., SUN, H., ZHAO, H., GAO, H., MACLELLAN, C., ZHAO, S. and LUO, X.

2021

Effective Extraction of Ventricles and Myocardium Objects from Cardiac Magnetic Resonance Images with a Multi-Task Learning U-Net

Jinchang Ren[a,b*], He Sun[c], Huimin Zhao[a], Hao Gao[d], Calum Maclellan[e], Sophia Zhao[e], Xiaoyu Luo[d]

[a] *School of Computer Sciences, Guangdong Polytechnic Normal University, Guangzhou, 510655, China*
[b] *National Subsea Centre, Robert Gordon University, Aberdeen, U.K.*
[c] *School of Computer Science, Beijing Institute of Technology, Beijing, 100081, China*
[d] *School of Mathematics and Statistics, University of Glasgow, Glasgow, U.K.*
[e] *Centre for Signal and Image Processing, University of Strathclyde, Glasgow, U.K.*

*Corresponding authors: jinchang.ren@ieee.org (J. Ren) and Zhaohuimin@gpnu.edu.cn (H. Zhao).

A B S T R A C T

Accurate extraction of semantic objects such as ventricles and myocardium from magnetic resonance (MR) images is one essential but very challenging task for the diagnosis of the cardiac diseases. To tackle this problem, in this paper, an automatic end-to-end supervised deep learning framework is proposed, using a multi-task learning based U-Net (MTL-UNet). Specifically, an edge extraction module and a fusion-based module are introduced for effectively capturing the contextual information such as continuous edges and consistent spatial patterns in terms of intensity and texture features. With a weighted triple loss including the dice loss, the cross-entropy loss and the edge loss, the accuracy of object segmentation and extraction has been effectively improved. Extensive experiments on the publicly available ACDC 2017 dataset have validated the efficacy and efficiency of the proposed MTL-UNet model.

Keywords: U-Net; multi-task learning; magnetic resonance images (MRI); ventricles and myocardium extraction; fusion-based decoder.

1. Introduction

Early detection and diagnosis are essential for cardiac diseases, for which accurate extraction of the desired objects or regions of interest (ROI) is particularly useful for reconstructing an appropriate heart model. To this end, segmentation-based accurate extraction of both ventricles and the myocardium, especially from the magnetic resonance images (MRI), has become a crucial prerequisite for computational heart modelling and following-on assessment of pump functions and diagnosis towards personalized treatment planning.

Recently, some algorithms have been developed for the segmentation of cardiac MR images. For example, Tran has proposed the first fully convolutional network (FCN) in the application of cardiac segmentation, which inferences the result at the pixel level [1]. Afterwards, Poudel et al. have developed a recurrent FCN for learning the inter-slice affiliation from a full stack of 2D slices [2].

Although the existing models have achieved some good results, there are still several unsolved challenges within the task of cardiac MRI segmentation. First, the boundaries between ventricles and the myocardium are sometimes hard to identify because of the partial volume effect. Secondly, the intensity values between the desired ROI and other surrounding regions could be very similar, which increases the difficulty of segmentation. Thirdly, medical image segmentation requires a much higher accuracy than the segmentation of natural images, where even small errors may affect the follow-on diagnosis and treatments. Due to the aforementioned difficulties in segmentation of cardiac MRI, this task remains extremely challenging.

In the last few years, the encoder-decoder based deep learning architectures, such as the U-Net [3] and FCN [4], have achieved a state-of-the-art performance in image segmentation. Thanks to its strong efficiency and a simple structure, it becomes a trend to apply the U-Net based models for a variety of applications in medical image segmentation, such as the segmentation of retinal vessels [5], brain tumors [6, 7], cells [8], lesions [10], the pulmonary lobes [11], and the cardiac MRI [1, 2, 9]. With the interactions between the encoder path and the decoder path, the U-Net model can efficiently combine both the low-level and

high-level features with several down-sampling, up-sampling and a specific skip connection for discovering more fine-grained details.

Although the U-Net based methods have made a success in medical image segmentation, most of these methods consider the segmentation task as a classification problem without taking into account the spatial localization between different pixels. As a result, the segmentation accuracy is compromised especially for challenging cases of cardiac MR images. Hence, it is a new trend to make the best use of the contextual information with an extended U-Net for improving the segmentation accuracy.

For improving the performance of deep learning-based image segmentation, quite a few models have been introduced recently and focused on the extraction of contextual spatial information. In order to extract more semantic features, most existing methods are designed based on rich spatial information and sufficient receptive field. For achieving a superior balance between rich contextual information and the inference speed, Yu et al. have proposed the BiseNet to extract more useful contextual information whilst preserving the spatial information [12]. In [13], a dilated convolutional kernel is designed to increase the receptive field, where multiscale contextual information is obtained without losing the spatial information.

In addition, there are also some other approaches proposed in this context. However, it is not our intention to provide a comprehensive survey, as summarized by Schuster et al in [14], which covers not only different approaches but also relevant datasets, such as a recent MS-SMSeg dataset [15]. As accurate segmentation and extraction of the ventricles and the myocardium is critical for 3D heart modelling [16], this has continuously motivated new approaches in the area. Some recently developed approaches include the multivariate mixture model based myocardial segmentation [17], multiobjective adaptive convolutional neural network segmentation [18], rough-fuzzy K-Means classification [19], and shape-transfer GAN [20] et al.

However, there is a trade-off between the spatial resolution and the receptive field. For example, the U-Net utilizes four down-sampling operations to increase the receptive field and extract higher-level features. However, the spatial resolution of the feature map after four times of down-sampling becomes much smaller than the original image, which may result in the loss of the spatial information. Although some operations with an increased size of convolutional kernels or dilated convolutions can provide a large receptive field, they may inevitably result in the huge extra computational burden. In addition, a straightforward interpretation of the designed method is important for the medical image segmentation applications, which is often neglected.

In this paper, we propose a multi-task deep learning model based on the U-Net (MTL-UNet) for automatic segmentation of the cardiac MRI. The contribution of this paper can be summarized as follows. Firstly, we have introduced an edge detection module along with an edge loss to derive more contextual features from different spatial scales. After that, a novel fusion-based module is implemented to combine the derived contextual features with the low-level and high-level features extracted from the original U-Net. To validate the efficiency of the proposed method, we have evaluated the performance of the designed MTL-UNet framework on the ACDC 2017 dataset. Experimental results have shown its significant superiority and great potential in segmentation of the cardiac images.

This remaining paper is organized as follows. Section 2 presents in detail the proposed model, especially the architecture and implementation of the MTL-UNet. Section 3 summarizes the experimental results. Finally, some concluding remarks are drawn in Section 5 along with directions for future investigation.

2. The proposed model

The architecture of the MTL-UNet framework is based on the U-Net, a popularly used fundamental model for the two-dimensional (2D) medical image segmentation. In the MTL-UNet, automatic cardiac segmentation is implemented to extract four semantic classes from each image, i.e. the background, the right ventricle, the myocardium, and the left ventricle.

Fig. 1 shows the flowchart of our proposed MTL-UNet framework, in which two new modules are introduced in the conventional U-Net model, i.e. an edge detection module and a feature fusion module. The edge detection module is to extract edge features from each spatial size in the encoder path as the context information in the spatial domain. In the feature fusion module, the extracted edge features are combined with the low-level and high-level features extracted from the original U-Net. Taking the extracted segmentation map and the edge map as the output, the proposed MTL-UNet model is trained by using a mixture loss function including a specific edge loss function.

By jointly learning the output edge map and the segmentation map, the MTL-UNet can combine these contextual features and improve the segmentation accuracy. Detailed implementation of relevant modules is presented as follows.
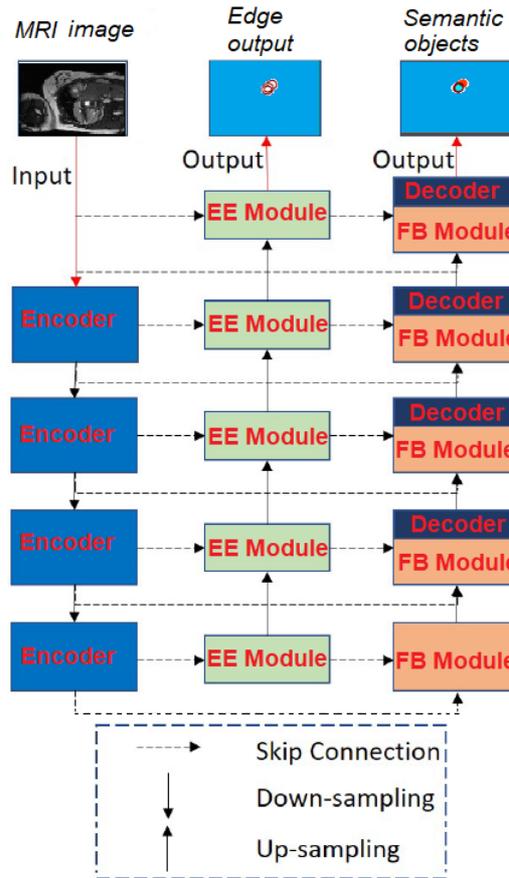

Fig. 1. The flowchart of the proposed MTL-UNet: EE and FB denote the Edge Extraction and Feature Fusion modules, respectively.

2.1. Edge Extraction Module

Image edges are generally very useful to define the boundaries of different objects within an image, which is also a natural way for image segmentation. In our MTL-UNet model, we introduce an edge extraction (EE) module motivated by the conventional edge detector. As both the ventricles and the myocardium are always adjacent to each other, the introduced edge information can help to refine the localized boundaries between different classes and thus improve the segmentation accuracy. Additionally, the desired inter-class edge information can be easily derived from the ground truth, which can thus enable easy coupling of the edge information with the end-to-end training model.

In Fig. 2, we illustrate examples of the original MR images, the corresponding ground truth and the edge maps. As can be seen, for some images the ground truth maps are clearly visible in the original images, indicating relatively easy for detection. However, for other cases it is very unclear, sometimes the expected objects are completely missing in the original images despite of similar spatial structures, which can be false alarms in modelling. This has shown the challenge of the problem.

(a)   (b)   (c)
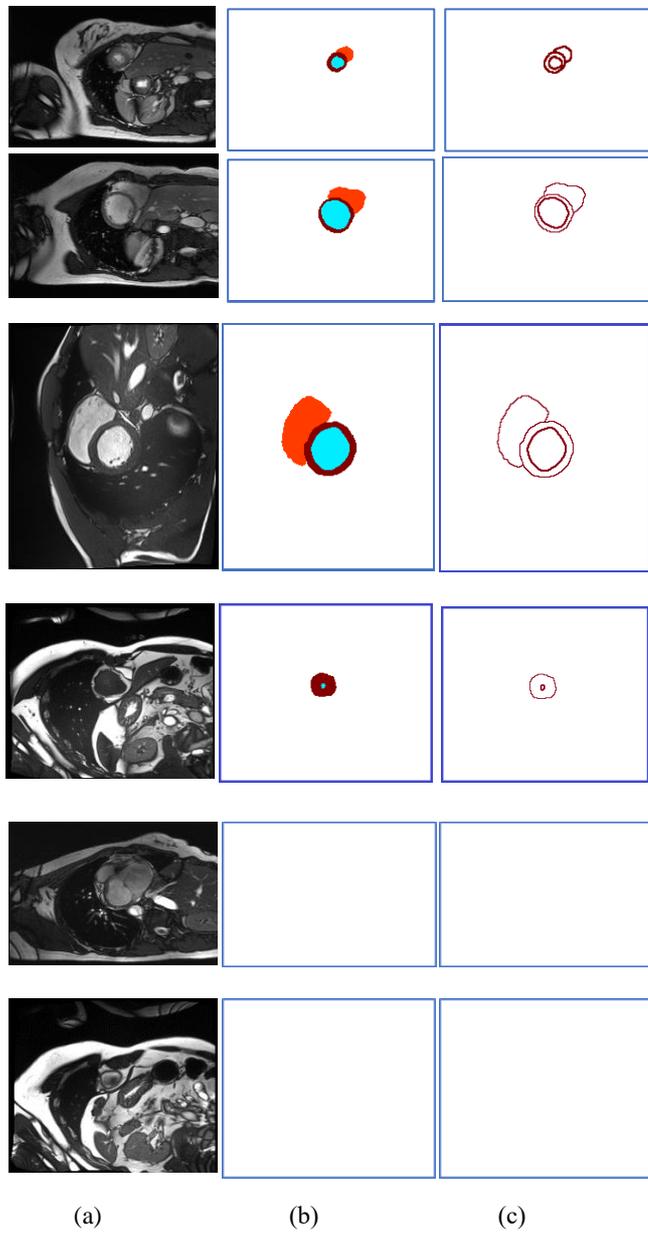
Fig. 2. The illustration of the edge map: (a) the raw data. (b) the ground truth, where the white, the red, the cyan and the brown represent the background, the right ventricle, the left ventricle and the myocardium, respectively. (c) The generated edge map from the ground truth, where the darker and the white color represents the edge and the non-edge. Note for the last two images there is no ground truth defined, indicating no real objects in the image hence potential false alarms.

Based on the analysis above, the edge extraction module is composed of two layers, where each layer has one 3 × 3 convolution with the stride equals to 1, followed by group normalization and ReLU. As such, the spatial size of the output feature maps is the same as the input feature map of the edge extraction module. The detail of the edge extraction module is illustrated in Fig. 3. In order to derive more valuable features from different scales, especially from the large receptive field level, the edge extraction module is placed in every spatial scale of the skip connection path, which is also clearly shown in Fig. 1, the flowchart of the proposed MTL-UNet model.

Fig. 3. The proposed edge extraction module.

2.2. Fusion-based Decoder Module

With the edge extraction module, two kinds of features can be produced from the encoder path, including the edge features and the common encoder features. Although these two features are extracted from same spatial scales simultaneously, a simple sum operation on both feature maps may not be a good way to fuse them and represent spatial and contextual information jointly. Hence, we introduce a Fusion-based (FB) module before the decoder path used in U-Net, which is shown in Fig. 4. For both features from the same spatial size, they are concatenated first and used as the input of the decoder module. Next, the input features are halved after one convolutional layer, which includes one $3e3$ convolution with stride equal to 1, followed by group normalization and ReLU.
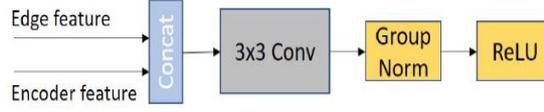
Fig. 4. The proposed fusion-based decoder.

2.3. Model Architecture

The detailed architecture of our MTL-UNet is given in Fig. 5, which takes an encoder-decoder based U-Net as the backbone. Except from employing the same feature encoder/decoder modules as the U-Net, an edge extraction module is added in the skip connection path to obtain more contextual features than the conventional U-Net [21]. The extracted edge information is combined with other common features extracted from the original U-Net via a fusion-based module.

Different from most of the existing methods, the spatial size of the input image in the MTL-UNet is arbitrary instead of downsampling or up-sampling of the input image into a given size for spatial normalization. Consequently, we set the batch size to 1 in our training model and apply the group normalization but not the widely used batch normalization in our proposed MTL-UNet framework, which has helped to improve the stabilities during the training process. The input image is firstly convoluted by two convolutional layers. The details of the implemented encoder

decoder modules are the same as the EE module, including two convolutional layers, group normalization and a ReLU.

2.4. Multitask Learning

In our MTL-UNet framework, the significance is to extract more contextual information whilst preserving the useful information from both the low-level and high-level features. Therefore, it is important to supervise the end-to-end learning process and ensure the improved segmentation accuracy. In the MTL-UNet, the training process is implemented by a commonly used multi-task learning (MTL) strategy.

As seen in Figs. 1 and 5, there are two outputs in the MTL-UNet, i.e. one segmentation map and one edge map, where each output can be obtained by a $1 e 1$ convolution and the log SoftMax function. By applying a $3 e e$ edge detection template on the ground truth segmentation map with the central pixel weighted as 8 and all others as -1, a ground truth edge map can be generated. Hence, we can compute an edge loss from the derived edge map and the ground truth edge map.
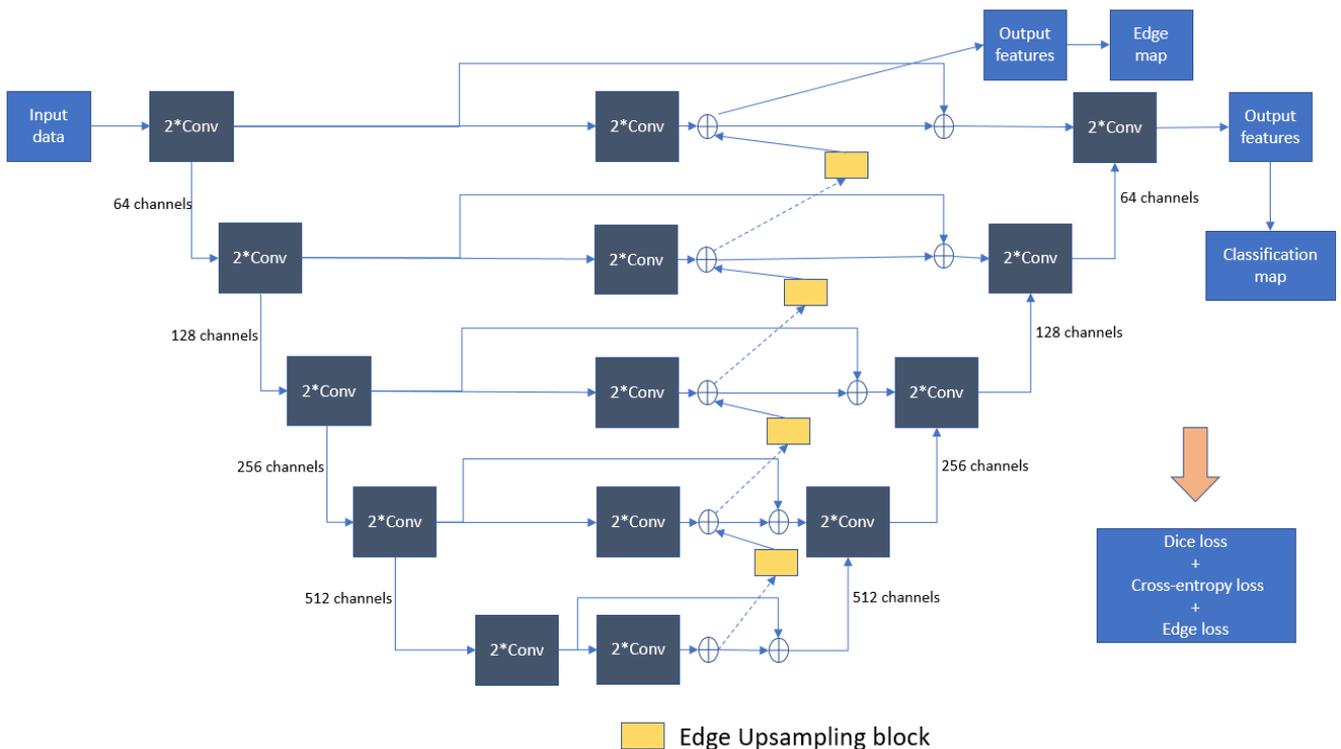
Fig. 5. Detailed architecture of the proposed MTL-UNet.

The edge loss can then be depicted as:

$$L_{edge} = \frac{1}{N} \sum_i -[\gamma_i \log(\rho_i) + (1 - \gamma_i) \log(1 - \rho_i)] \tag{1}$$

where i_i is the pixel label, which can be either 1 or 0, representing whether the pixel i is an edge pixel or not. For i_i, it is the prediction probability of i being an edge pixel.

In addition, we have utilized the dice loss and the cross-entropy (CE) loss together to evaluate the training process of image segmentation. For the dice loss, it is utilized to evaluate the segmentation performance similar to the intersection over the union (IoU) in object detection as follows [22]:

$$L_{Dice} = 1 - 2 * \frac{A \cap B}{A \cup B} \tag{2}$$

where A and B denote respectively the predicting value and the target value, i.e. the ground truth. The *intersection* represents the number of corrected prediction pixels and the *union* is total number of pixel in both the prediction and the ground truth.

The CE loss is defined by

$$L_{CE} = \frac{1}{N} \sum_{i=1}^{N} (- \sum_{c=1}^{C} y_{ic} \log(p_{ic})) \tag{3}$$

where N represents the number of total samples (pixels); and $y_{ic} = 1$ if the prediction result equals to the ground truth, otherwise it is 0. The p_{ic} represents the class probabilities of the sample i belonging to the class c.

In our proposed loss function, we have adopted three weights to normalize the effect of each loss function above. The combined joint loss is defined by:

$$L = \gamma \times L_{Dice} + \beta \times L_{CE} + \sigma \times L_{edge} \tag{4}$$

where γ, β, and σ denote the coefficient of each loss within the introduced weighting strategy for the utilized loss functions.

With the joint loss function above, a multi-task learning model can be learned, including the edge detection branch and the segmentation branch. By utilizing the edge loss to guide the learning process, more boundary information can be taken into consideration to improve the performance of segmentation simultaneously. To focus more on the potential contextual information of edges, we have set the parameters γ, β, σ to 1, 1, and 4, respectively.

3. Experiments and results

3.1. Experimental Settings

For performance evaluation of the proposed MTL-UNet framework, the MICCAI 2017 Automated Cardiac Diagnosis Challenge (ACDC 2017 [23]) dataset is used. As there is no ground truth available for the testing dataset in the ACDC 2017 dataset, only the training dataset is used, whilst the cross-validation strategy is adopted on the training dataset.

As the utilized dataset from the ACDC 2017 includes MR images from 100 patients, we randomly choose 80 patients as the training set and the rest 20 as the testing set. After extracting the 2D slices from the MR data of each patient, about 1800 slices are used for training and 370 slices for testing. In the experimental data, four semantic classes are labelled, including the background, the right ventricle, the myocardium, and the left ventricle.

For the hardware and software settings, all experiments are implemented on the Pytorch 1.1.0 package with an 8GB NVIDIA GTX 1080. The experimental results are quantitated using the mean Dice coefficients from five repeated experiments, where the training samples and testing samples are randomly selected accordingly. In our experiments, the Adam [24] is used to optimize the training process. As for hyperparameters, the learning rate, the batch size, and the training epochs are set to be 1e-4, 1, and 300.

3.2. Results

To validate the effectiveness of our proposed MTL-UNet framework, experiments with or without the proposed EE and FB modules are derived for comparison as well as the weighted joint loss function, and the performance is compared with the baseline U-Net model. Table I summarizes the performance of the proposed MTL-UNet in term of the dice index. As seen, our proposed method has improved about 7% dice coefficient in comparison to the conventional U-Net model. Within the 7% improvements, over 4% is from the edge extraction module, which has clearly shown the great potential in utilizing the contextual spatial information to improve the segmentation accuracy. In addition, the FB module and the joint loss function has further improved the results for all three interested objects, i.e. the left ventricle, the right ventricle, and the myocardium.

Table 1. Quantitative results of the proposed method.

Methods	UNet	UNet+EE	UNet+EE+FB	MTL-UNet
Dice of right ventricle (%)	62.38	67.09	68.17	**72.44**
Dice of myocardium (%)	75.83	78.87	78.61	**80.74**
Dice of left ventricle (%)	82.54	86.86	88.14	**88.18**
Dice index (%)	73.59	77.61	78.30	**80.39**

In Fig. 6, several examples of our experimental results are shown. As seen, the baseline U-Net model suffers from inaccurate boundary and misclassification problems. After combining the edge extraction module, the results have been improved with a more accurate boundary. With the fusion based-encoder module, more contextual information has been derived but the results still suffer from false alarms. Thanks to the newly introduced modules and the MTL stragegy, they have helped to significantly improve the performance of our proposed MTL-UNet model than the standard U-Net. Besides, the raw data with only the background class can be successfully detected with the introduced edge loss.
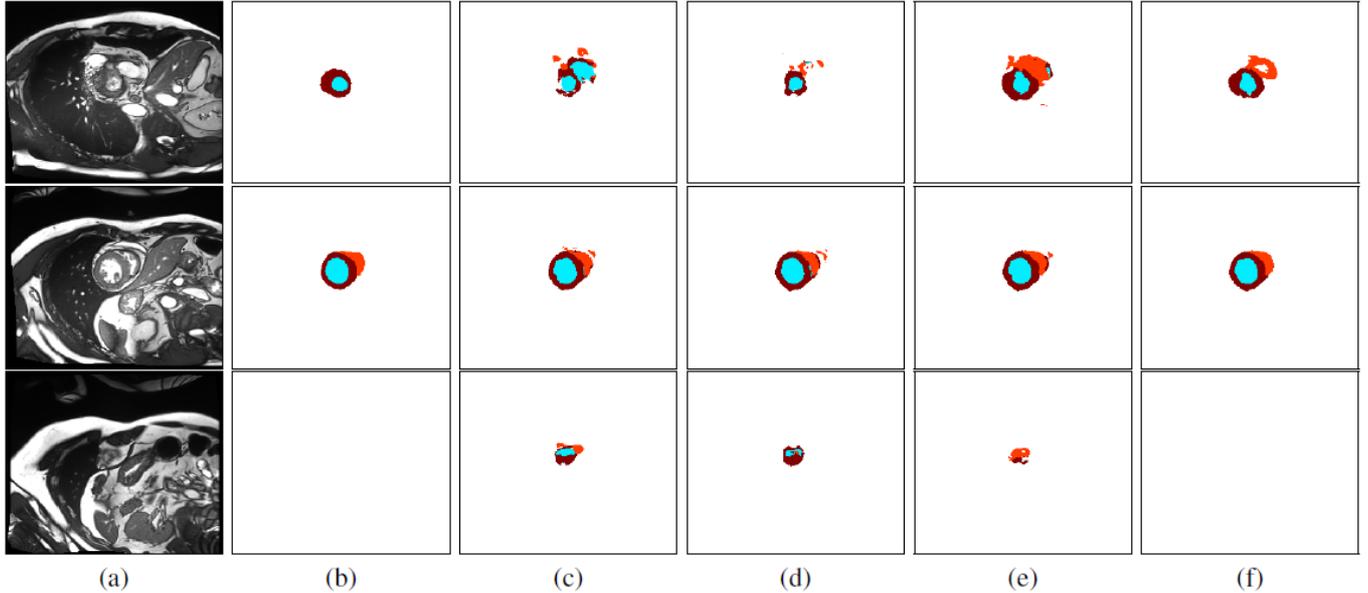
Fig. 6. Segmentation map on several examples using different techniques: (a) The raw data. (b) The ground truth. (c) The result from U-Net. (d) The result from U-Net+EE. (e) The result from U-Net+EE+FB. (f) The result from MTL-UNet (UNet+EE+FBB+weighted loss). In the segmentation map, the color represent the same classes we mentioned in Fig. 2.

In Table 2, we have also compared our proposed method with the Dense-UNet [23], which has a more complicated architecture than our proposed MTL-UNet. Although the Dense-UNet has a better performance in the right ventricle, it fails to produce a comparable result in both the myocardium and the left ventricle. With the EE and FB module, our method can achieve a better performance with a more efficient neural network, which has again validated the robustness and efficacy of our proposed MTL-UNet.

Table 2. Comparison result.

Methods	UNet	Dense-UNet	MTL-UNet
Dice of right ventricle (%)	62.38	**73.44**	72.44
Dice of myocardium (%)	75.83	78.99	**80.74**
Dice of left ventricle (%)	82.54	87.21	**88.18**
Dice index (%)	73.59	79.88	**80.39**

4. Conclusions

In this paper, we have proposed a multi-task learning based deep neural network, MTL-UNet for end-to-end cardiac MRI segmentation based accurate extraction of the ventricles and myocardium. By encoding more contextual information such as the edge information, the accuracy of the segmentation has been significantly improved. With the designed edge extraction module and the joint edge loss function, the training process can be effectively guided to obtain more accurate boundary information and refined segmentation results. Afterwards, the extracted features can be fused with the features extracted from the encoder by a proposed fusion-based module.

With more extracted contextual information, it is anticipated further improved boundary and refined segmentation can be achieved, using other related techniques such as change detection [25], abnormality detection [26], brain tumor segmentation and recognition [27], mammogram recognition [28], corneal injury detection [29] and feature fusion network [30]. In addition, the anatomy and physiology of the heart will also be integrated for improving the model accuracy and efficacy [31]. These will be the focus of our future work, where more spatial information including texture features and structural similarity will be explored to further improve the efficacy of the MTL-UNet model.

Acknowledgments

References

[1] P. V. Tran. "A fully convolutional neural network for cardiac segmentaion in short-axis MRI" [Online]. Available: https://arxiv.org/abs/1604.00494, 2017

[2] R. P. K. Poudel, P. Lamata, and G. Montana. "Recurrent fully convolutional neural networks for multi-slice MRI cardiac segmentation," in Proc. RAMBO, HVSMR, 2016, pp. 83-94.

[3] O. Ronneberger, P. Fischer, and T. Brox. "U-net: Convolutional networks for biomedical image segmentation," in Proc. MICCAI, 2015, pp. 234-241.

[4] J. Long, E. Shelhamer, and T. Darrell. "Fully convolutional networks for semantic segmentation," in Proc. CVPR, 2014, pp. 3431–3440.

[5] X. Jiang, M. Lambers, and H. Bunke. "Structural performance evaluation of curvilinear structure detection algorithms with application to retinal vessel segmentation," Pattern Recognition Letters, 2012, 33(15), pp. 2048-2056.

[6] J. Amin, et al, "Brain tumor classification based on DWT fusion of MRI sequences using convolutional neural network," Pattern Recognition Letters, 2020, 129, pp. 115-122.

[7] T. H. Nguyen, et al., "Efficient Brain Tumor Segmentation with Dilated Multi-fiber Network and Weighted Bi-directional Feature Pyramid Network," in International Conference on Digital Image Computing: Techniques and Applications (DICTA), 2020.

[8] Z. Zhou, et al, "UNet++: A nested U-Net architecture for medical image segmentation," in Proc. MICCAI, 2018, pp. 3–11.

[9] J. Ren, H. Sun, Y. Huang, and H. Gao. "Knowledge-based multi-sequence MR segmentation via deep learning with a hybrid U-Net++ model," in Proc. STACOM, 2019.

[10] H. Mugasa, et al. "An adaptive feature extraction model for classification of thyroid lesions in ultrasound images," Pattern Recognition Letters, 2020, 131, pp. 463-473.

[11] W. Wang, J. Chen, J. Zhao, Y. Chi, X. Xie, L. Zhang, X. Hua. "Automated segmentation of pulmonary lobes using coordination-guided deep neural networks," in Proc. ISBI, 2019.

[12] C. Yu, J. Wang, C. Peng, C. Gao, G. Yu, and N. Sang. "Bisenet: Bilateral segmentation network for realtime semantic segmentation," in Proc. ECCV, 2018, pp. 325–341.

[13] F. Yu, V. Koltun. "Multi-scale context aggregation by dilated convolutions," in Proc. ICLR, 2015.

[14] A. Schuster, et al. "Cardiovascular Magnetic Resonance Myocardial Feature Tracking Concepts and Clinical Applications," Circulation: Cardiovascular Imaging 9(4), e004077, 2016.

[15] MS-CMRSeg Multi-sequence Cardiac MR Segmentation Challenge homepage, https://zmiclab.github.io/mscmrseg19/

[16] H. Gao, Q. Long, M. Graves, J. Gillard and Z. Li, "Carotid arterial plaque stress analysis using fluid-structure interactive simulation based on in-vivo magnetic resonance images of four patients", Journal of Biomechanics, vol. 42, no. 10, pp. 1416-1423, 2009.

[17] X. Zhuang, "Multivariate mixture model for myocardial segmentation combining multi-source images", IEEE Transactions on Pattern Analysis and Machine Intelligence, 41(12): 2933-2946, 2019.

[18] M. Baldeon-Calisto and S. Lai-Yuen, "AdaResU-Net: Multiobjective adaptive convolutional neural network for medical image segmentation", Neurocomputing, 2019.

[19] A. Elmoasry, "Classification of MR medical images Based Rough-Fuzzy KMeans", IOSR Journal of Mathematics, vol. 13, no. 01, pp. 69-77, 2017.

[20] X. Tao, H. Wei, W. Xue and D. Ni, "Segmentation of Multimodal Myocardial Images Using Shape-Transfer GAN", arXiv.org, 2019. https://arxiv.org/abs/1908.05094.

[21] M. Drozdzal, E. Vorontsov, G. Chartrand, S. Kadoury, and C. Pal. "The importance of skip connections in biomedical image segmentation." In Deep Learning and Data Labeling for Medical Applications, pages 179–187. Springer, 2016.

[22] S. Jha, L. Son, R. Kumar, I. Priyadarshini, F. Smarandache and H. Long, "Neutrosophic image segmentation with Dice Coefficients", Measurement, vol. 134, pp. 762-772, 2019.

[23] O. Bernard, A. Lalande, C. Zotti, F. Cervenansky, et al. "Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: is the problem solved?" IEEE Transactions on Medical Imaging, 2018, 37(11).

[24] D. P. Kingma, and J. Ba. Adam. "A method for stochastic optimization," [Online]. Available: https://arxiv.org/abs/1412.6980, 2014.

[25] B. Zhang, Z. Wang, et al, "Short-term lesion change detection for melanoma screening with novel Siamese neural network," IEEE Trans. Medical Imaging, 2021, 40(3), pp. 732-747.

[26] A. Bhandary, et al, "Deep–learning framework to detect lung abnormality – a study with chest X-Ray and lung CT scan images," Pattern Recognition Letters, 2020, 129, pp. 271-278.

[27] M.I. Sharif, et al, "Active deep neural network features selection for segmentation and recognition of brain tumors using MRI images", Pattern Recognition Letters, 2020, 129, pp. 181-189.

[28] J Ren, D Wang, J Jiang, "Effective recognition of MCCs in mammograms using an improved neural classifier," Engineering Applications of Artificial Intelligence, 2011, 24 (4), pp. 638-645.

[29] SS Md Noor, et al, "Hyperspectral image enhancement and mixture deep-learning classification of corneal epithelium injuries," Sensors 2017, 17(11), paper id: 2644.

[30] Z Fang, et al, "A novel multi-stage residual feature fusion network for detection of COVID-19 in chest X-ray images," IEEE Transactions on Molecular, Biological and Multi-Scale Communications, 2021, in press.

[31] F. Haddad, et al, "Right ventricular function in cardiovascular disease, Part I - anatomy, physiology, aging, and functional assessment of the right ventricle", Circulation, 2008, 117, pp. 1436– 1448.