Magnetic Resonance Imaging xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Magnetic Resonance Imaging

journal homepage: www.elsevier.com/locate/mri

Original contribution

Dynamic pixel-wise weighting-based fully convolutional neural networks for left ventricle segmentation in short-axis MRI

Zhongrong Wang, Lipeng Xie*, Jin Qi

School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu, China

ARTICLE INFO

Keywords:
Cardiac MRI
Left ventricle segmentation
Convolutional neural networks
Semantic segmentation
Dynamic pixel-wise weighting strategy

ABSTRACT

Left ventricle (LV) segmentation in cardiac MRI is an essential procedure for quantitative diagnosis of various cardiovascular diseases. In this paper, we present a novel fully automatic left ventricle segmentation approach based on convolutional neural networks. The proposed network fully takes advantages of the hierarchical architecture and integrate the multi-scale feature together for segmenting the myocardial region of LV. Moreover, we put forward a dynamic pixel-wise weighting strategy, which can dynamically adjust the weight of each pixel according to the segmentation accuracy of upper layer and force the pixel classifier to take more attention on the misclassified ones. By this way, the LV segmentation performance of our method can be improved a lot especially for the apical and basal slices in cine MR images. The experiments on the CAP database demonstrate that our method achieves a substantial improvement compared with other well-know deep learning methods. Beside these, we discussed two major limitations in convolutional neural networks-based semantic segmentation methods for LV segmentation.

1. Introduction

For diagnosis of cardiovascular diseases, the cardiac magnetic resonance imaging (MRI) have been viewed as the “gold standard” in clinical practice [1]. LV is one of four chambers of the heart and located in the bottom left portion of the heart. Left ventricular (LV) segmentation of in cardiac MR images can provide various important clinical indices for grading or diagnosis of disease such as the left ventricular mass and volumes [2]. However, for most expert clinicians, manual delineation of LV in cardiac MR images is time-consuming, laborious and suffers from the intra and inter-rater variability [3]. Therefore, the need for an automated and accurate segmentation method of LV is particularly urgent, which will improve the efficiency and accuracy of diagnosis for cardiac diseases.

As illustrated in Fig. 1, the left ventricle segmentation task is to delineate the myocardial region between the endocardium (red contour) and epicardium (green contour) in short-axis cardiac MR images. Then we can extract the inner curve of myocardial region as the endocardium, which surrounds the LV cavity, and the outer curve as the epicardium, which is the boundary between the myocardium and the surrounding tissue [4]. Nevertheless, this task is challenging for several reasons: a) the high similarity of intensities between the papillary muscles and myocardium; b) the inhomogeneity of the brightness in LV cavity; c) the class imbalance between myocardium and other components especially on apical and basal slices; d) the great variability in terms of gray level intensities and shape duo to the different patients, MRI scans and balanced Fast Field Echo (bFFE) sequences cardiac MRI; e) inherent noise associated with cine MRI; etc. [3-5].

Recently, numerous effective left ventricle segmentation methods have been proposed. Mitchell et al. [6] established a fully automated framework to segment the left and right ventricles by combining a hybrid active shape model/active appearance model (ASM/AAM) with a model matching initialization approach. Katouzian et al. [7] integrated the morphological operations and threshold decomposition opening method to extract the boundary of endocardium and epicardium in the region-of-interest (ROI) of cardiac MRI. However, the segmentation performance heavily depends on the selection of ROI. In [8], the left ventricle was located by motion map and expectation maximization algorithm, then the contour was achieved by the proposed fuzzy connectedness-based image segmentation framework. Nevertheless this approach is sensitive to the variations in thickness and intensity values around the myocardium. Feng et al. [9] proposed a novel distance regularized two-layer Level set method to segment the myocardium, in which two specified level contours were utilized to represent the endocardium and epicardium respectively. A limitation of their method is that the initialization of Level set function can directly

* Corresponding author.
E-mail address: xlpflyinsky@foxmail.com (L. Xie).

https://doi.org/10.1016/j.mri.2019.08.021
Received 2 August 2018; Received in revised form 11 June 2019; Accepted 15 August 2019

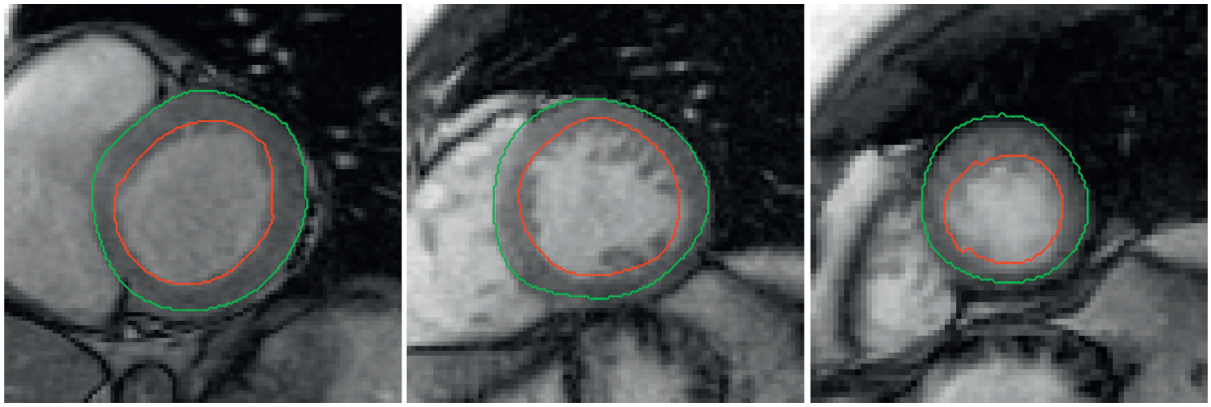


Fig. 1. The examples of manual LV segmentation in the cine MR images from base to apex: the green and red contours represent the epicardium and endocardium respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

affect the segmentation performance. Besides this, the level set approach is not robust enough for processing a variety of cine MR images. This method requires the users to know the meaning of each parameter of level set approach, and learn to adjust the parameters for improving the LV segmentation performance.

With the rise of convolutional neural networks (CNN) [10], most fields in computer vision and pattern recognition experience a huge revolution and development, including object detection, image classification and image segmentation [11]. Conforming to the trend, a variety of CNN-based left ventricle segmentation methods [12-18] were proposed and obtained good performance in clinical practice. Different from the traditional approaches, the CNN-based semantic segmentation methods utilize the hierarchical feature extracted from original images to train a pixel-wise classifier and then assign special label (e.g. “0” and “1” represent the non- and myocardial region respectively) to all pixels. In other words, CNN-based semantic segmentation methods transform the segmentation task into a simple and efficient pixel-level classification task by a reasonable way. However, these methods fail to solve the unbalance sample and unbalance class problems, which leads to that the segmentation performance for apical and basal slices is far better than middle slices, and the per-pixel classification accuracy for non-myocardial pixels is well above the myocardial pixels. These issues make bad influence on the 3D reconstruction of left ventricle.

In this paper, we develop a novel dynamic pixel-wise weighting-based fully convolutional neural network for automatic left ventricle segmentation, which can be trained end-to-end efficiently on a single GPU and perform image-to-image prediction for cardiac MR images. In our method, the main novelties can be summarized as follows: 1) integrating the hierarchical feature as a feature pyramid by the concatenation operations for improving the robustness of network, and 2) presenting a dynamic pixel-wise weighting generation approach for handling the dataset imbalance and class imbalance problems in LV segmentation task. Note that the weight map for pixels is adaptively learned from the multi-level pixel-wise classifiers in our network, and changes over the segmentation result. The main mechanism is to increase the weight value of pixels misclassified at upper classifier, and decrease the weight value of correctly classified pixel. This operation can force the network to focus on the intractable training data. The experimental data demonstrates that this dynamic pixel-wise weighting approach can improve the classification performance for the myocardial pixels and the segmentation performance for the apical and basal slices at the same time.

The remainder of our paper is organized as follows. In Section 2, the previous research work about CNN-based image segmentation and our proposed approaches are illustrated. Next, the implementation details about our method and experimental process are described in Section 3. Then we show the comparison of experimental results, and analyze the effect of our proposed pixel-wise weighting approach in Section 4. Finally, in Section 5, we make the conclusion of our work and discuss two major limitations exist in CNN-based semantic segmentation methods.

2. Proposed method

2.1. The overview of CNNs-based semantic segmentation

To the best of our knowledge, the most influential semantic segmentation method based on CNN should be the Fully Convolutional Networks (FCN) by Long et al. [19]. Note that the traditional CNN like AlexNet [10], VGG [20], and GoogLeNet [21] were designed for whole-image classification and made significant contributions to this field. However, the powerful features learned from these architectures would loss the spatial information under the action of Pooling layers, which makes it hard to track and label each pixel. In order to overcome this problem, the authors of FCN [19] replaced the Fully connected layer with convolutional layer, which allows the network to output spatial feature maps instead of classification scores. Nevertheless, the final segmentation result is very coarse because the feature map of last layer loses most local information. To tackle this issue, they presented three improvement measurements, such as reducing the stride of pooling layers, shift-and-stitch trick, and combining the fine layer with coarse layer by element-wise layer. As illustrated in Fig. 2 (a), the final output of FCN is a synthesis of outputs from previous multi-level layers.

Inspired by their work, more networks with fully convolutional architecture have been designed for semantic segmentation such as HED [22], U-net [23], SharpMask [24] and Feature Pyramid [25]. The main difference among them is how to take advantage of the hierarchical feature maps or side-output results for improving the final segmentation. Specially, HED network merged all the prediction results from side-output layers as a comprehensive feature map by concatenation layer. The final segmentation result was obtained by performing per pixel classification on the feature map. Different from this idea, the authors of U-net network utilized the concatenation operation to combine the high-resolution and low-resolution feature maps as a feature pyramid, which contains both the global and local information of image, as illustrated in Fig. 2 (c). Then the feature pyramid was applied to predict the label of pixels. The SharpMask and Feature Pyramid networks employed the similar approach to stack the multi-level feature maps as a whole. In this paper, we propose a novel network architecture with making full use of multi-scale feature maps and side-outputs simultaneously, as shown in Fig. 2 (d). The main difference between our method and the above networks is that we present a dynamic pixel-wise weighting approach block (denoted by red arrow) for

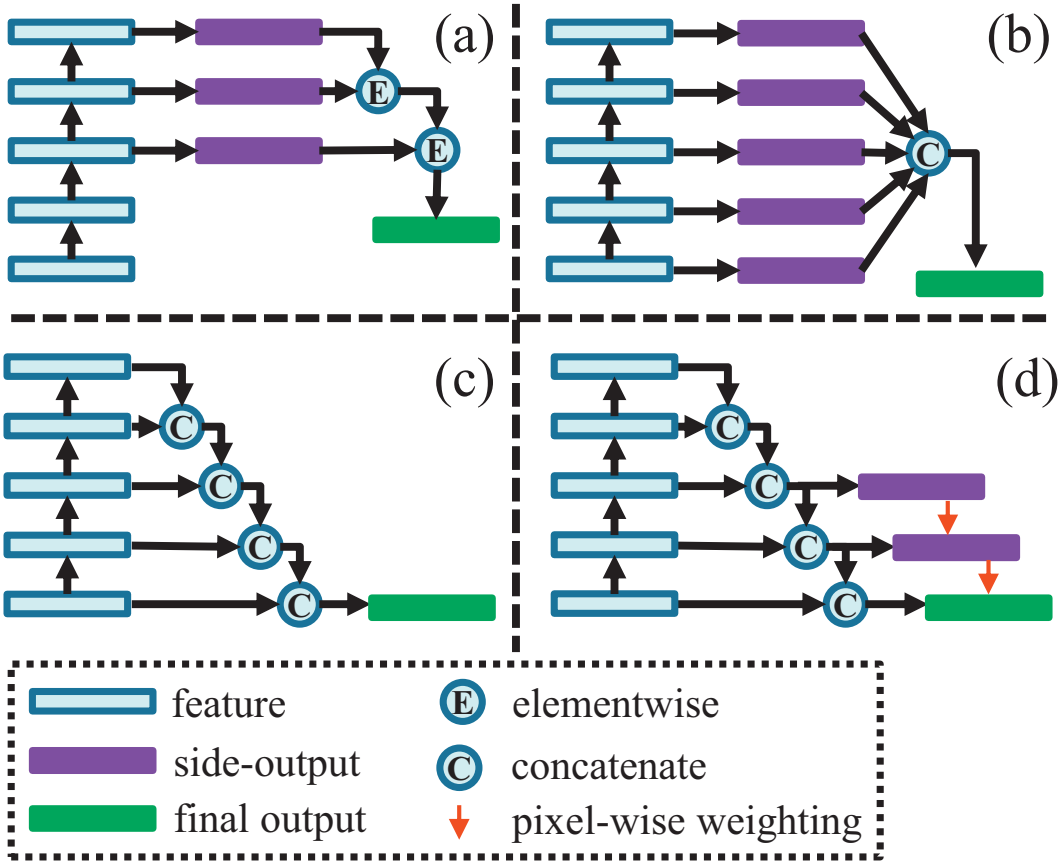


Fig. 2. Multi-scale merging strategy. (a) FCN; (b) HED; (c) U-net; (d) proposed.

utilizing the low resolution segmentation result to refine the high resolution one.

2.2. The architecture of proposed network

As illustrated in Fig. 3, our network architecture consists of a bottom-up pathway (left side) and a top-down pathway (right side), following the top-down refinement approach by Pinheiro et al. [24]. The bottom-up pathway is a modified version of VGG16 [20] without Fully connected layer. It consists of five downsampling blocks (denoted by "DS" in Fig. 3). Each DS block is composed of two or three repeated 3×3 Convolutional layers (zero padding) with stride 1, a rectified linear unit (ReLU) activation function, a Batch Normalization layer and a 2×2 Max-pooling layer with stride 2. With propagating forward along the bottom-up pathway, the feature map would be semantically stronger while the resolution decreases. In general, the coarsest resolution feature map contains object-level information, which is necessary for object classification, and the finest-resolution one contains rich spatial information that makes per-pixel classification possible. So it is natural for a variety of semantic segmentation methods to integrate the hierarchical features at multi-scale layers.

The top-down pathway is composed of four concatenating & upsampling blocks (denoted by "CU" in Fig. 3), which can upsample the feature map of low resolution and merge the multi-level feature maps together. Each CU block consists of a 3×3 Convolutional layers (zero padding) with stride 1, a Concatenating layer and a Bilinear interpolation layer with factor 2. Different from the Deconvolution method in FCN, HED and U-net, the weights of Bilinear interpolation layer are fixed, and do not require learning. Due to that the Concatenating layer amalgamates multiple feature maps into a single one along the channel axis, the number of final feature map channels would be very large. For reducing data redundancy of the final feature map, we add a Convolutional layers to each CU block with setting the output channels to 64. Under the action of top-down pathway, the feature maps from different levels could form a feature pyramid with scale-invariant characteristic.

In addition, there are three side-output blocks with two 3×3 convolutional layers for achieving the per-pixel classification result. For the side-output block at small scale, a Bilinear upsampling layer is need to make the spatial size of all output results consistent with the input image.

2.3. Dynamic pixel-wise weighting and loss function

Obviously, how to utilize these multi-scale segmentation results to improve the final segmentation performance is an interesting problem. Base on the Adaboost theory [26], we propose a novel pixel-wise weighting strategy (PW), which can make proper use of the coarse segmentation result to enhance the performance of classifier at high-resolution layer by adding a weight map to each side-output loss function. By this way, there is a progressive relationship between the small-scale and large-scale layer. The weight map and side-output result make contribution to final loss function for training the proposed network.

Formally, we define the given input image and associated segmentation ground truth as $I = \{I_x \in \mathbb{R} | x \in \Omega\}$ and $G = \{G_x \in \mathscr{L} | x \in \Omega\}$, where Ω and $\mathscr{L}$ indicates the image space domain with N pixels and label sets with M different labels. The side-output result generated by the k-th side-output layer is denoted by $O^k = \{O_x^k \in \mathbb{R}^M | x \in \Omega\}$. Note that the resolution of feature map would be higher with the k increasing. The posterior probabilities for pixel x attribute to each class are given by soft-max algorithm as follows:

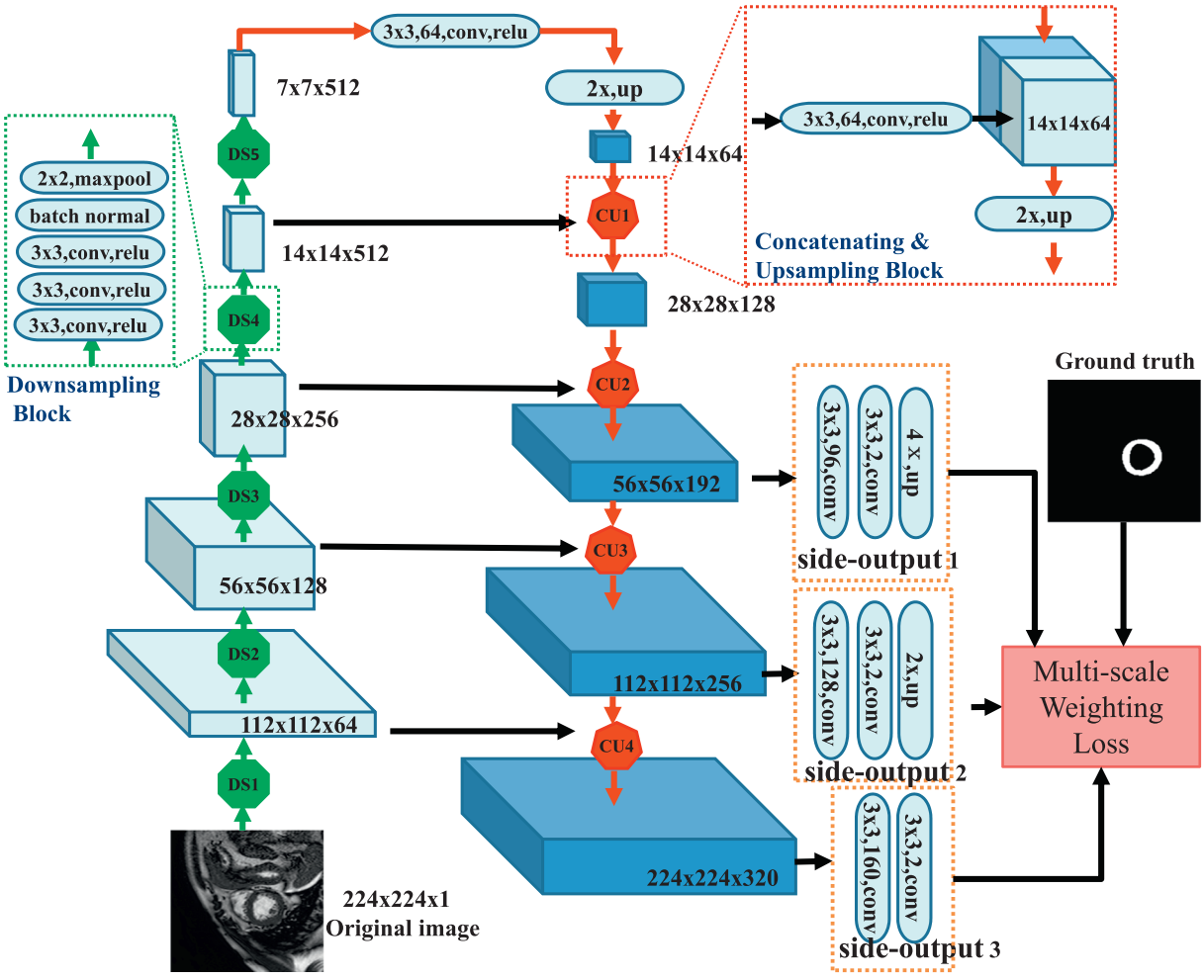


Fig. 3. The proposed network architecture. The green and red arrow indicate the bottom-up and top-down pathway respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

$$P^k(y = l|x) = \frac{\exp(O_x^k(l))}{\sum_{l'=1}^{M} \exp(O_x^k(l'))} \tag{1}$$

Then we predict the segmentation result (denoted by S_x^k) by the Bias decision theory as follows:

$$S_x^k = \text{argmax}_l P^k(y = l|x),\ l = 1...L \tag{2}$$

Algorithm 1. Computing pixel-wise weight maps.

Input: $G,\ S = \{S^k, k = 1...K\}$

Initial: $w_x^1 = \frac{1}{N}, \varepsilon = 10^{-6}, \delta = 1.5$

1: **for** $k \in \{1...K\}$ **do**

2: $a^k = \frac{1}{N} \sum_{x \in \Omega} (S_x^k = G_x)$

3: $w_x^{k+1} = w_x^k \cdot \begin{cases} (a^k + \varepsilon)^{-\delta} & if \quad S_x^k \neq G_x \\ (a^k + \varepsilon)^{\delta} & otherwise \end{cases}$

4: $w_x^{k+1} = \frac{w_x^{k+1}}{\sum_{x \in \Omega} w_x^{k+1}}$

Output: $w = \{w^k, k = 1...K\}$

Supposing there are totally K side-output layers, the pixel-wise weight maps $w = \{w^k, k = 1...K\}$ would be computed by Algorithm 1. We firstly initialize the pixel-wise weight map in the first output layer to $1/N$. Then we calculate the accuracy of classifier (denoted by a^k) in k-th output layer, by comparing the ground truth G with segmentation result S^k. For the case that a pixel x was predicted wrongly by the k-th classifier ($S_x^k \neq G_x$), the weight value of x in the next layer would increase, by multiplying by the factor $(a^k + \varepsilon)^{-\delta}$, which is greater or equal to 1. By contrast, the weight value of correctly predicted pixels would decrease, by multiplying by the factor $(a^k + \varepsilon)^{\delta}$, which is less or equal to 1. The parameter δ at the third line is applied to control the amplitude of variation of weight value, while the constant ε (0.000001) is a bias term for avoiding the zero value of a^k. At the last line, the weight map is normalized for ensuring that the sum of weights equals to 1. Fig. 4 shows three examples of side-output result and weight map generated from Algorithm 1. By comparing the side-output results, we can find that the weight value of well-classified pixels are decreased, and the side-output results are improved step-by-step during training stage.

Let W denotes all learn-able variables in network such as the weight and bias terms of Convolutional layers. The multi-scale segmentation result and weight map altogether consist of the loss function of our network, called Multi-scale weighting loss (MWL), which is defined as Eq. (3):

$$L(I, G; W) = -\sum_{k=1}^{K} \sum_{x \in \Omega} w_x^k \log(P^k(y = G_x|x)) \tag{3}$$

Due to the pixel-wise weighting strategy, the misclassified pixels at low-level layer would make more contribution to MWL compared with the correctly predicted pixels. In order to reduce the loss, the

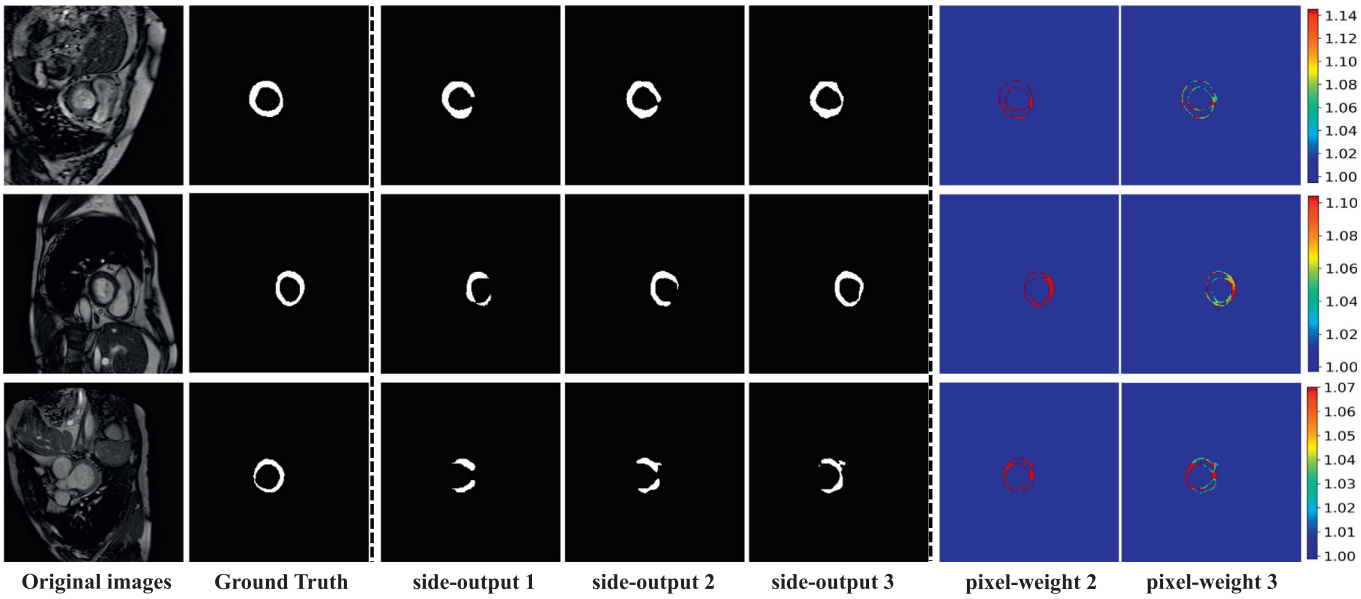


Fig. 4. Examples of side-output result and pixel-wise weight map during training stage.

optimization algorithm would update the unknown variables W in favor to correctly predict the misclassified pixels. In this way, the segmentation result could be improved at next side-output layer. Compared with other classical weighting strategies, there are several obvious advantages of our weighting method as follows: 1) adjusting the weight value dynamically rather than fixing the weight value, 2) fully considering the different importance among pixels rather than categories, 3) integrating the weighting strategy into our network instead of a pre-processing step, 4) strengthening the relationships between the multi-scale side-output layers.

3. Experimental framework

We evaluate the segmentation performance of our method on one public Cardiac MRI dataset. The details about data pre-processing and augmentation methods used in our experiments are illustrated in this section. Beside these, the main experimental procedures, including algorithm implementation, training and inference stage, are also shown. Note that all experimental codes are available at https://github.com/appiek/Left_ventricular_segmentation_Demo.

3.1. Dataset

The LV dataset used in this paper is provided by the Cardiac Atlas Project[1] [27]. This dataset is randomly selected from DETERMINE cohort [28] and made publicly available as a part of MICCAI 2013 SATA Challenge on automated LV myocardium segmentation. The dataset contains 83 "Nii.gz"-formatted files (4D matrix including height, width, position and time dimensions) for training and another 73 files for testing. However, only the training files have corresponding expert-guided segmentation regions of myocardium, in where the foreground and background denoted by 1 and 0, respectively. In our experiments, we divide the training "Nii.gz"-formatted files into two parts: training data with 50 files and testing data with 33 files.

3.2. Data pre-processing and augmentation

Before training, we need to prepare enough training and validation data for training the network. However, the original Cardiac MR image "Nii.gz"-formatted files are not suitable as the training data directly for several reasons: 1) the MR image data is a multi-dimension matrix containing sequence of short-axis slices, while our network is designed for 2D image segmentation; 2) the slice shape of each image data is not uniform; 3) the great diversity of pixel intensities in MR image data has a negative impact on the segmentation performance; 4) the surrounding tissues and myocardium are easily to be confused; etc.

For reducing the interfering from above factors, we presented some measurements. At first, we extracted image slices from the training image data by sampling. Due to the similarity of adjacent slices, we only remained one of every two consecutive slices in the temporal axis. Then, we randomly selected almost a quarter of the slices as the validation data (1216 slices) for adjusting the hyper-parameters of network, the rest as the final training data (3648 slices) for optimizing the weights of network. For the testing dataset, we retained all data and obtained a total of 5859 slices. Next, the pixel intensity of each slice was normalized by Min-Max Normalization as defined in Eq. (4):

$$I_x^{norm} = \frac{I_x - I_{min}}{I_{max} - I_{min}} \tag{4}$$

where I_{min} and I_{max} denote the minimum and maximum of intensity in one slice. We also subtracted the average of pixel intensity from each patch.

In order to improve the invariance and robustness of convolutional network, we employed some data augmentation methods to process the training data such as the center cropping and random flipping. We utilized center cropping to extract the region of interest (ROI) patches with size 224×224 from the slices. To ensure that the LV cavity locate in the center of each patch, we only take the foreground points as the center point of ROI. By random flipping, the image patches would be vertical or horizontal flipped randomly. All these data augmentation methods were only used at training and validation data. The content of data pre-processing and augmentation are summarized in Table 1.

3.3. Training and testing

Our proposed network was implemented based on the open-source deep learning library TensorFlow [29] and its extended version TensorLayer [30]. The experimental procedure was conducted on a workstation with 4.0 GHz Intel(R) i7-6700K CPU and NVIDIA GeForce GTX TITAN X GPU. In this paper, we compare the proposed method with

[1] http://www.cardiacatlas.org/.

Table 1
Data pre-processing and augmentation methods for each stage.

Stage	Niis	Slices	Data augmentation			Data pre-processing	
			Center crop	Vertical flip	Horizontal flip	Min-max norm	Remove mean
Training	50	3648	Yes	Yes	Yes	Yes	Yes
Validation		1216	Yes	Yes	Yes	Yes	Yes
Testing	33	5859	No	No	No	Yes	Yes

other three known semantic segmentation networks: FCN8s [19], HED [22]and U-net [23]. In addition, to explore the effect of the proposed pixel-wise weighting strategy, we incorporated it into Unet network. This hybrid model can be denoted by Unet + PW. We also implemented a simplified version of our network by removing the side-output layers and the pixel-wise weighting strategy. For distinguishing our networks with other competing CNNs in the experiments, we represent the simplified version by "BIUnet" (Bilinear Interpolation Unet), and the full version by "BIUnet + PW". For the sake of fair comparison, we utilized the same tools to re-implemented the architectures of these networks, and trained all networks using the same hyper-parameters and training data.

At training stage, the weights of convolutional layers in the side-output blocks were initialized by the Xavier [31] method with uniform distributed random initialization. We employed the Adam algorithm to minimize the Multi-scale Weighting Loss function defined as Eq. (3) with setting the learning rate and the exponential decay rate for the first momentum to 0.0001 and 0.9. The dropout rate of all Dropout layers is set to 0.65. According to our experience, setting the batch size to 8 is advantageous for the algorithm converging and achieving good results. The parameter δ in Algorithm 1 is set to 1.5. Training iteratively 400 times takes about 16.5 h.

At testing stage, all method process a total of 33 "Nii.gz"-formatted (including 5859 image slices). The testing data must be processed by Min-Max Normalization and removing mean operation before input the data to network, and each slice needs to be resized to 256 × 256 by interpolation method. For our method, the output results of the finest level classifier were achieved as the final segmentation results, instead of combining outputs from multiple scales. Actually, the feature map and segmentation result from the coarse-level are applied to improve the performance of classifier in the fine-level for two reasons:(1) the feature map in the $(k + 1)$th-level contains all the features from upper layers, (2) the classifier in the $(k + 1)$th-level would draw a lot of lessons from the kth-level to refine the segmentation result.

3.4. Evaluation metrics

We evaluate the performance of all segmentation methods from two sides: pixel classification and object segmentation. Assume that S and G denote the binary segmentation result and ground truth of myocardium respectively, we employed Dice Index (DI) and Jaccard Index (JI) to assess the overall effect of segmentation as defined in Eq. (5).

$$DI = 2\frac{S \cap G}{S + G}, \qquad JI = \frac{S \cap G}{S \cup G} \tag{5}$$

In addition, the common metrics for classification task such as Sensitivity (P), Specificity (Q), Positive predictive value (PPV), and Negative predictive value (NPV) are applied to measure the performance of per-pixel classification as follows:

$$P = \frac{T_1}{N_1}, Q = \frac{T_0}{N_0}, PPV = \frac{T_1}{T_1 + F_1}, NPV = \frac{T_0}{T_0 + F_0} \tag{6}$$

where N_1 and N_0 are the number of pixels in the foreground and background of ground truth, T_1 and T_0 denote the number of correctly classified pixels as the foreground and background in segmentation result, and F_1 and F_0 indicate the misclassified pixels as the foreground and background in segmentation result. Note that the value of all these metrics vary from zero (worst) to one (best).

4. Results and discussion

4.1. Slice-level segmentation

We calculated the metric value for each slice, and then computed the mean value and standard deviation of all metric values as the final quantitative evaluating index for all methods. The comparison of segmentation performance and average processing time between the proposed method and other state-of-the-art CNN-based methods are illustrated in Table 2. Overall, our method outperforms other three networks a lot for most evaluation measures. Especially in terms of Positive predictive value, Dice Index and Jaccard Index, we achieve the best score with almost + 7% higher than the second one. Further more, the corresponding standard deviations are lower than others, that means the segmentation performance of our method is more stable compared with other methods. The performance of FCN8s and HED are roughly the same, but worse than Unet. We can also find that the efficiency of all CNN-based semantic segmentation methods are considerable. The average processing time of our method is 0.044 s, which is very close to the HED network.

By comparing BIUnet + PW with BIUnet, it's clear to find that the proposed pixel-wise weighting strategy can improve the segmentation performance of BIUnet network a lot in terms of Dice Index (+ 9.9%), Jaccard Index (+ 12.7%) and Positive predictive value (+ 17.7%). Though BIUnet obtains the highest Sensitivity (0.906/0.207), but the Positive predictive value is very low (0.594/0.208). The big gap between Sensitivity and Positive predictive value means that the

Table 2
Comparison of LV myocardium segmentation performance of slices between FCN8s, HED, Unet, proposed-PW and our proposed method. All values are expressed as "mean/standard deviation".

Method	FCN8s	HED	Unet	Unet + PW	BIUnet	BIUnet + PW
Dice	0.719/0.253	0.708/0.296	0.735/0.276	0.705/0.28	0.704/0.217	**0.803/0.204**
Jaccard	0.609/0.246	0.611/0.278	0.639/0.267	0.602/0.272	0.579/0.211	**0.706/0.214**
Sensitivity	0.824/0.255	0.77/0.323	0.788/0.29	0.725/0.289	**0.906/0.207**	0.859/0.2
Specificity	0.996/0.002	0.997/0.002	0.997/0.002	0.997/0.002	0.993/0.004	**0.998/0.002**
PPV	0.665/0.248	0.682/0.279	0.714/0.26	0.716/0.275	0.594/0.208	**0.771/0.206**
NPV	0.999/0.001	0.998/0.002	0.998/0.002	0.998/0.002	0.999/0.001	**0.999/0.001**
Time (s)	0.129	**0.032**	0.046	0.047	0.42	0.044

appreciable amounts of background pixels were wrongly classified as foreground by BIUnet network. With the help of pixel-wise weighting strategy, the BIUnet network can pay more attention to the misclassified pixels in background, that leads to the substantial reduction of F_1. But this can also slightly impair the ability of correctly classifying the foreground for BIUnet network, that causes the decline of T_1. Therefore, the Positive predictive value of BIUnet + PW increases a lot, while the Sensitivity decreases a bit. This phenomenon proves that the proposed pixel-wise weighting strategy can efficiently reduce the risk of misclassifying background pixels. Besides this, we can also observe that the performance of pixel-level classification in the background is far better than in the foreground for all methods. It means that these methods tend to predict the foreground as the background. The main reason is the class imbalance problem existed in Cardic MR images. Compared with other methods, the gap of pixel classification accuracy between foreground and background for our method is narrow. It suggests that our pixel-wise weighting strategy (see Section 2.3) have the ability to solve class imbalance problem. However, the comparison between Unet and Unet + PW shows that the proposed pixel-wise weighting strategy makes a negative effect on Unet network. The analysis of this phenomenon will be presented in Section 4.4.

To analyze the performance of myocardial segmentation in details, four representative examples are shown in Fig. 5. The red and green curves represent the automatic and manual segmentation results, respectively. The columns from left to right indicate the segmentation results of FCN8s, HED, Unet, BIUnet, and BIUnet + PW. Obviously, the contour of segmentation results of our method are more close to the ground truth compared with other methods. And the HED and BIUnet + PW network are likely to predict the background regions near the myocardial region as foreground.

4.2. Object-level segmentation

Due to that the left ventricular in MRI is a 3D object rather than 2D object, the analysis of slice-level segmentation performance is not comprehensive enough. So we evaluated the segmentation performance of each “Nii.gz”-formatted file, which is a patient's 3D scanning result of heart object. To do this, we firstly employed the CNN-based image segmentation methods to process the slices in each “Nii.gz”-formatted file. Then we obtain the 3D LV segmentation result by stacking the segmentation result of slices. Next, we computed the Sensitivity, Positive predictive value, Dice Index and Jaccard Index of each 3D LV segmentation result. Fig. 6 shows the box plots of 3D segmentation metric values for different methods.

Through analyzing these experimental data, we can observe that BIUnet network obtains the best performance in the term of Sensitivity, while the other evaluation values are worse than other methods. Our proposed BIUnet + PW exhibits apparent advantages in the term of Positive predictive value. Though the Dice Index of all methods range approximately from 0.6 to 0.9, our method achieves the highest minimum, first quartile, median, third quartile, and maximum than other methods. In addition, our method obtains the highest first quartile, median, third quartile, and maximum in the terms of Jaccard Index. These findings further validate the advantages of our method and have a high level of agreement with the conclusions made from Table 2.

4.3. Segmentation results at various positions

As stated in article [4], the segmentation complexity is closely related with the position of slice. Specially, the apical and basal slices of left ventricle are more difficult to segment than the mid-slices. Because the closer to the apex, the smaller the structure of LV is, resulting in that

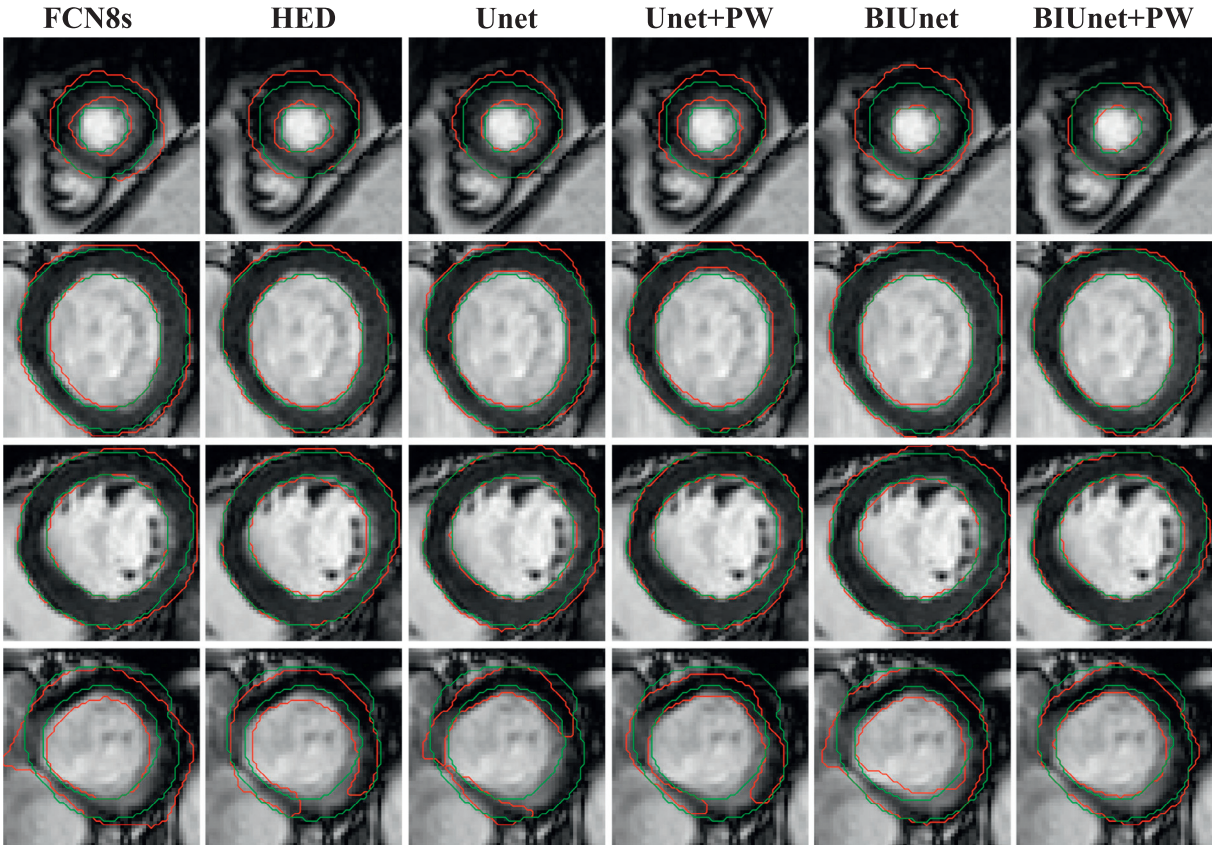


Fig. 5. Examples of left ventricular segmentation on CAP dataset. Red color: the contour of automatic segmentation results, green color: the contour of expert-guided manual segmentation results. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

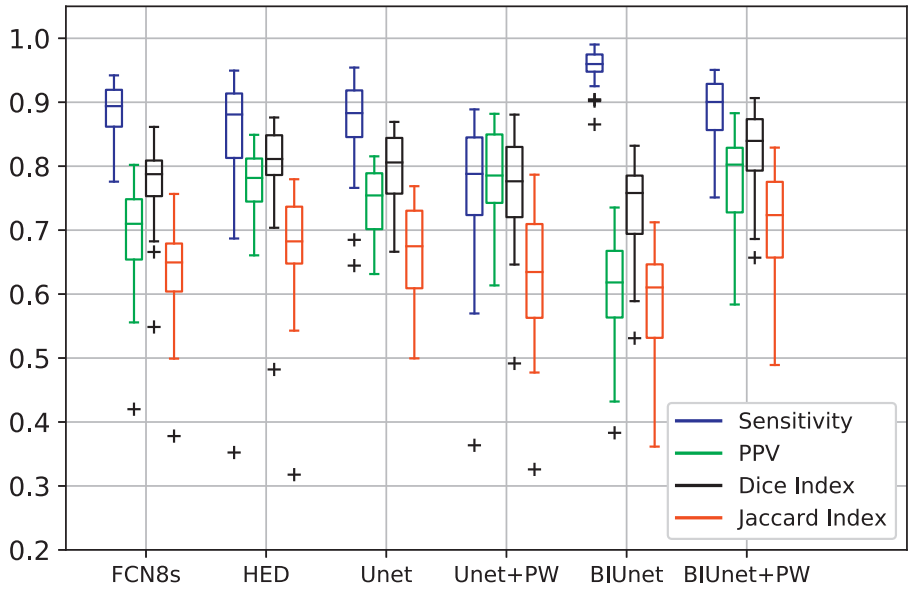


Fig. 6. Boxplots of the sensitivity (blue), positive predictive value (green), Dice Index (black) and Jaccard Index (red) of 3D segmentation result for different methods. The black symbol "+" denote the outlier. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

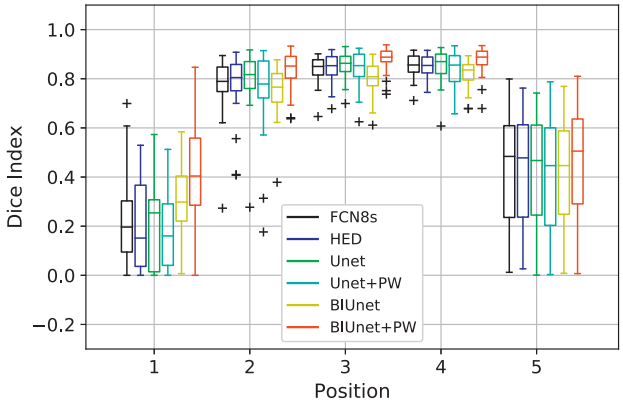
the foreground pixels in apical slices are often neglected by the automatic segmentation methods. Compared with the mid-slices, the shape of LV in basal slices is often irregular, which may confuse the pixel-level classifier.

For quantitative analysis of the segmentation results at various positions, we divide the position axis into five equal regions from apex (1) to base (5). The slices at position from 2 to 4 could be viewed as the mid-slices. Then we calculated the mean value of Dice index and Jaccard index of the slices at these positions. As illustrated in Fig. 7, we can find that the segmentation performance at mid-positions is far better than the apical and basal positions for all methods. For instance, the median of Dice index of our method from position 2 to 4 is about 0.85, while the median is 0.4 at position 1 and 0.5 at position 5. It is clear to see that our method outperforms other competing networks at all positions especially at apex. At the same position, most indicators of our method such as the first quartile, median, third quartile, and maximum, are higher than other networks. Fig. 8 shows some representative segmentation results of slice at various positions, which is consistent with Fig. 7.

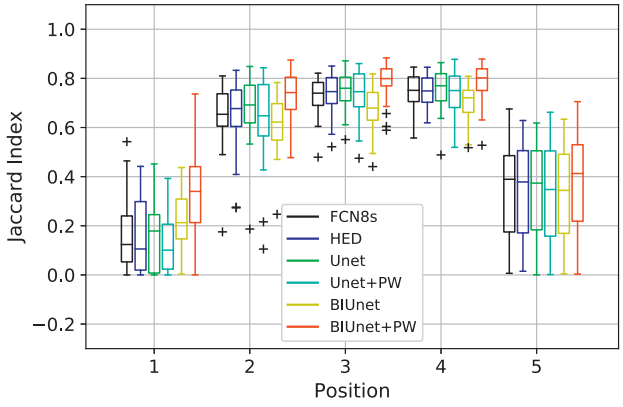
4.4. Analysis of pixel-wise weighting strategy

As shown above, the proposed pixel-wise weighting strategy has obvious positive effect on the proposed BIUnet network, but makes the segmentation performance of Unet worse. The main reason is that the pixel-wise weighting strategy contains a lot of model parameters, which increases the freedom degree of CNNs. Compared with our BIUnet network, the model complexity of Unet network is higher. Therefore, the over-fitting problem in Unet + PW network is more serious than in our BIUnet + PW network. This leads to the large decline of myocardial segmentation performance on testing set for Unet network. To prove the rationality of this explanation, we show the iteration process of Unet, Unet + PW, BIUnet, and BIUnet + PW networks in Fig. 9. At every 20 iteration times, we utilized the four networks to segment the training and validation data, and calculated the pixel-wise F1-score.

It is clear to see that the pixel-wise F1-score of Unet + PW network on training set is far higher than on validation set. This gap is greater than other networks. By contrast, for the proposed BIUnet + PW network, the pixel-wise classification performance on validation set is gradually approach to the training set with the increasing of iteration



(a) Dice index



(b) Jaccard index

Fig. 7. Boxplots of Dice Index (a) and Jaccard Index (b) at various positions from apex (1) to base (5). The black symbol "+" denote the outlier.

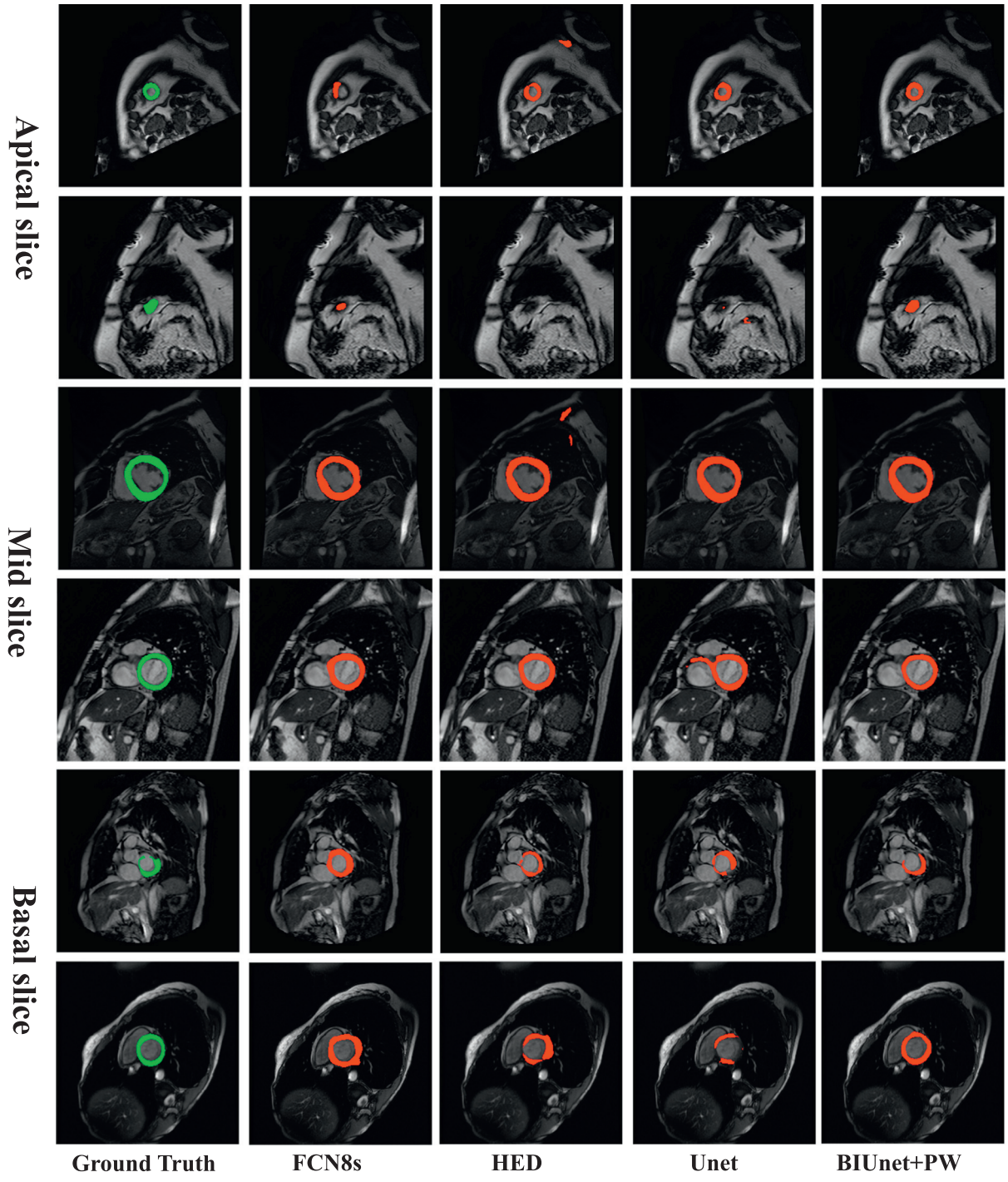


Fig. 8. Examples of segmentation result at various positions.

times. Moreover, the BIUnet + PW network obtains a considerable improvement both on training set and validation set compared with BIUnet network. Therefore, we can make conclusion that our pixel-wise weighting strategy can aggravate the model over-fitting problem for Unet network, but makes positive contribution to our BIUnet network. The over-fitting problem also exists in BIUnet + PW network, but the its influence on our method is lower than Unet network. Though our pixel-wise weighting strategy is not a generalized technique for improving the performance of CNNs, but it successfully help the proposed BIUnet network to solving the class imbalance problem in myocardial segmentation task and improve the segmentation performance for apical and basal slices.

5. Conclusion and future work

In this paper, we have proposed an effective fully automatic segmentation neural network architecture for Left Ventricle segmentation in Short-axis MRI, called Dynamic Pixel-wise Weighting-based Fully Convolutional Neural Networks. Our network employs the feature pyramid with scale-invariant characteristic to predict the per pixel label for MR images. In order to improve the segmentation result, we design a pixel-wise weighting strategy to adjust the importance of each pixel. The experiments on the public LV dataset provided by the Cardiac Atlas Project confirmed the effectiveness and advantages of our method.

Though the CNN-based semantic segmentation methods spend less

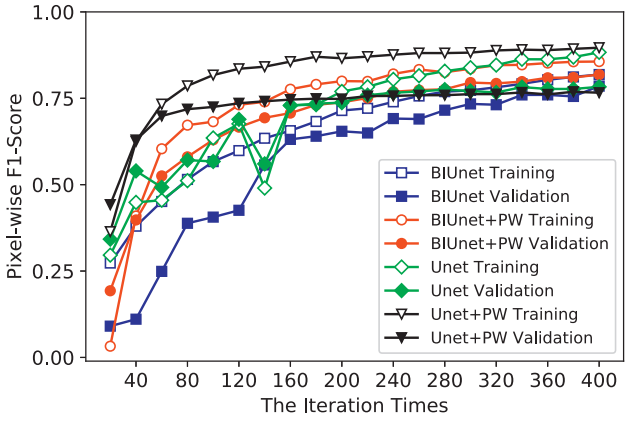


Fig. 9. Pixel-wise F1-score versus iteration times on training and validation set for four different networks: 1) Unet (green line), 2) Unet + PW (black line), 3) BIUnet (blue line), and 4) BIUnet + PW (red line). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

time achieving considerable performance in LV myocardium segmentation, there are two limitations that reduces the practicality of these algorithms. The first one is that the serious class-imbalance problem in LV segmentation task, which increases the False negative rate of pixel-level classification. Another obvious limitation is that these methods have poor effect on segmentation of image slice at apex and basal, though the apical slices are often considered to have little effect on computing clinical indices. In the future, we will focus on the segmentation of apical and basal slices apex for improving the performance of fully automatic LV segmentation methods. We will also attempt to combine other advanced image segmentation methods, such as level sets and active contour methods, with our method as a more robust computing framework for LV segmentation.

Acknowledgments

This work is supported by NSFC (Grant numbers G0561671135, G0591630311, M0501020111531005). The authors thank the anonymous reviewers for their valuable comments.

References

[1] Tan L K, Liew Y M, Lim E, McLaughlin R A. Convolutional neural network regression for short-axis left ventricle segmentation in cardiac cine MR sequences. Med Image Anal 2017;39:78–86. https://doi.org/10.1016/j.media.2017.04.002.

[2] Suinesiaputra A, Cowan B R, Al-Agamy A O, Elattar M A, Ayache N, Fahmy A S, et al. A collaborative resource to build consensus for automated left ventricular segmentation of cardiac MR images. Med Image Anal 2014;18(1):50–62. https://doi.org/10.1016/j.media.2013.09.001.

[3] Avendi M, Kheradvar A, Jafarkhani H. A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac MRI. Med Image Anal 2016;30:108–19. https://doi.org/10.1016/j.media.2016.01.005.

[4] Petitjean C, Dacher J-N. A review of segmentation methods in short axis cardiac MR images. Med Image Anal 2011;15(2):169–84. https://doi.org/10.1016/j.media.2010.12.004.

[5] Krasnobaev A, Sozykin A. An overview of techniques for cardiac left ventricle segmentation on short-axis MRI. vol. 8. EDP Sciences; 2016. p. 01003. https://doi.org/10.1051/itmconf/20160801003.

[6] Mitchell SC, Lelieveldt BPF, van der Geest R J, Bosch HG, Reiver JHC, Sonka M. Multistage hybrid active appearance model matching: segmentation of left and right ventricles in cardiac MR images. IEEE Trans Med Imaging 2001;20(5):415–23. https://doi.org/10.1109/42.925294.

[7] Katouzian A, Prakash A, Konofagou E. A new automated technique for left- and right-ventricular segmentation in magnetic resonance imaging. 2006 International Conference of the IEEE Engineering in Medicine and Biology Society IEEE; 2006. https://doi.org/10.1109/iembs.2006.260405.

[8] Pednekar A, Kurkure U, Muthupillai R, Flamm S, Kakadiaris IA. Automated left ventricular segmentation in cardiac MRI. IEEE Trans Biomed Eng 2006;53(7):1425–8. https://doi.org/10.1109/tbme.2006.873684.

[9] Feng C, Li C, Zhao D, Davatzikos C, Litt H. Segmentation of the left ventricle using distance regularized two-layer level set approach. Advanced Information Systems Engineering Springer Berlin Heidelberg; 2013. p. 477–84. https://doi.org/10.1007/978-3-642-40811-3_60.

[10] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. vol. 60. Association for Computing Machinery (ACM); 2017. p. 84–90. https://doi.org/10.1145/3065386.

[11] Garcia-Garcia A, Orts-Escolano S, Oprea S, Villena-Martinez V, Garcia-Rodriguez J. A review on deep learning techniques applied to semantic segmentation arXiv preprint https://arxiv.org/abs/1704.06857; 2017. arXiv:1704.06857.

[12] Poudel RPK, Lamata P, Montana G. Recurrent fully convolutional neural networks for multi-slice MRI cardiac segmentation. Reconstruction, segmentation, and analysis of medical images Springer International Publishing; 2017. p. 83–94. https://doi.org/10.1007/978-3-319-52280-7_8.

[13] Tran PV. A fully convolutional neural network for cardiac segmentation in short-axis MRI arXiv preprint arXiv:1604.00494 : https://arxiv.org/abs/1604.00494; 2016.

[14] Dong S, Luo G, Wang K, Cao S, Mercado A, Shmuilovich O, et al. VoxelAtlasGAN: 3D left ventricle segmentation on echocardiography with atlas guided generation and voxel-to-voxel discrimination. Medical Image Computing and Computer Assisted Intervention - MICCAI 2018 Springer International Publishing; 2018. p. 622–9. https://doi.org/10.1007/978-3-030-00937-3_71.

[15] Yan W, Wang Y, Li Z, van der Geest RJ, Tao Q. Left ventricle segmentation via optical-flow-net from short-axis cine MRI: preserving the temporal coherence of cardiac motion. Medical image computing and computer assisted intervention – MICCAI 2018 Springer International Publishing; 2018. p. 613–21. https://doi.org/10.1007/978-3-030-00937-3_70.

[16] Zhang D, Icke I, Dogdas B, Parimal S, Sampath S, Forbes J, et al. A multi-level convolutional LSTM model for the segmentation of left ventricle myocardium in infarcted porcine cine MR images. 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018) IEEE; 2018. https://doi.org/10.1109/isbi.2018.8363618.

[17] Khened M, Kollerathu VA, Krishnamurthi G. Fully convolutional multi-scale residual DenseNets for cardiac segmentation and automated cardiac diagnosis using ensemble of classifiers. Med Image Anal 2019;51:21–45. https://doi.org/10.1016/j.media.2018.10.004.

[18] Huaifei H, Pan N, Wang J, Yin T, Ye R. Automatic segmentation of left ventricle from cardiac MRI via deep learning and region constrained dynamic programming. Neurocomputing 2019. https://doi.org/10.1016/j.neucom.2019.02.008.

[19] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) IEEE; 2015. https://doi.org/10.1109/cvpr.2015.7298965.

[20] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition arXiv preprint arXiv:1409.1556 : https://arxiv.org/abs/1409.1556; 2014.

[21] Szegedy C, Liu W, Jia Y, Sermanet P, Scott Reed, Dragomir Anguelov, et al. Going deeper with convolutions. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) IEEE; 2015. https://doi.org/10.1109/cvpr.2015.7298594.

[22] Xie S, Tu Z. Holistically-nested edge detection. 2015 IEEE International Conference on Computer Vision (ICCV) IEEE; 2015. https://doi.org/10.1109/iccv.2015.164.

[23] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. Lecture notes in computer science Springer International Publishing; 2015. p. 234–41. https://doi.org/10.1007/978-3-319-24574-4_28.

[24] Pinheiro PO, Lin TY, Collobert R, Piotr Dollár. Learning to refine object segments. Computer vision - ECCV 2016 Springer International Publishing; 2016. p. 75–91. https://doi.org/10.1007/978-3-319-46448-0_5.

[25] Lin TY, Dollar P, Girshick R, He K, Hariharan B, Belongie S. Feature pyramid networks for object detection. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) IEEE; 2017. https://doi.org/10.1109/cvpr.2017.106.

[26] Freund Y, Schapire RE. A decision-theoretic generalization of on-line learning and an application to boosting. J Comput Syst Sci 1997;55(1):119–39. https://doi.org/10.1006/jcss.1997.1504.

[27] Fonseca CG, Backhaus M, Bluemke DA, Britten RD, Chung JD, Cowan BR, et al. The cardiac atlas project - an imaging database for computational modeling and statistical atlases of the heart. Bioinformatics 2011;27(16):2288–95. https://doi.org/10.1093/bioinformatics/btr360.

[28] Kadish AH, David Bello, Finn JP, Bonow RO, Schaechter A, Subacius H, et al. Rationale and design for the defibrillators to reduce risk by magnetic resonance imaging evaluation (determine) trial. J Cardiovasc Electrophysiol 2009;20(9):982–7. https://doi.org/10.1111/j.1540-8167.2009.01503.x.

[29] Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, et al. Tensorflow: a system for large-scale machine learning. 2016.: https://arxiv.org/abs/1605.08695. arXiv:1605.08695.

[30] Dong H, Supratak A, Mai L, Liu F, Oehmichen A, Yu S, et al. TensorLayer. 2017. https://doi.org/10.1145/3123266.3129391.

[31] Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. In: Teh YW, Titterington M, editors. Proceedings of the thirteenth international conference on artificial intelligence and statistics. vol.9. Chia Laguna Resort, Sardinia, Italy: PMLR; 2010. p. 249–56: http://proceedings.mlr.press/v9/glorot10a.html.