HIFUNet: Multi-class Segmentation of Uterine Regions from MR Images Using Global Convolutional Networks for HIFU Surgery Planning

Chen Zhang, Huazhong Shu, *Senior Member, IEEE,* Guanyu Yang, Faqi Li, Yingang Wen, Qin Zhang, Jean-Louis Dillenseger, and Jean-Louis Coatrieux, *Fellow, IEEE,*

***Abstract*— Accurate segmentation of uterus, uterine fibroids, and spine from MR images is crucial for high intensity focused ultrasound (HIFU) therapy but remains still difficult to achieve because of 1) the large shape and size variations among individuals, 2) the low contrast between adjacent organs and tissues, and 3) the unknown number of uterine fibroids. To tackle this problem, in this paper, we propose a large kernel Encoder-Decoder Network based on a 2D segmentation model. The use of this large kernel can capture multi-scale contexts by enlarging the valid receptive field. In addition, a deep multiple atrous convolution block is also employed to enlarge the receptive field and extract denser feature maps. Our approach is compared to both conventional and other deep learning methods and the experimental results conducted on a large dataset show its effectiveness.**

***Index Terms*— Encoder-Decoder, Global convolutional networks, HIFU, MR images, Segmentation, Uterine fibroids**

Manuscript received October 19, 2019; revised March 12, 2020; accepted April 26, 2020. This research was supported by National Natural Science Foundation under grants (31571001, 61828101, 61876037), the National Key Research and Development Program of China (2017YFC0107903), the Excellence Project Funds of Southeast University and the Science Foundation for The Excellent Youth Scholars of Southeast University. *(Corresponding author: Huazhong Shu.)*

C. Zhang, H. Shu and G. Yang are with the Laboratory of Image Science and Technology, Southeast University, Nanjing 210096, China, also with the Centre de Recherche en Information Biomedicale Sino-Francais, F-35000 Rennes, France, and also with the Key Laboratory of Computer Network and Information Integration, Ministry of Education, Southeast University, Nanjing 210096, China (e-mail: chenzhang@seu.edu.cn; shu.list@seu.edu.cn; yang.list@seu.edu.cn).

F. Li is with the State Key Laboratory of Ultrasound Engineering in Medicine, College of Biomedical Engineering, Chongqing Medical University, Chongqing 400016, China (e-mail: lifq@cqmu.edu.cn).

Y. Wen is with National Engineering Research Center of Ultrasound Medicine, Chongqing 401121, China (e-mail: wenyg@haifu.com.cn).

Q. Zhang is with Chongqing Haifu Medical Technology Co. Ltd. Chongqing 401121, China (e-mail: zhangq@haifu.com.cn).

J.-L Dillenseger and J.-L Coarieux are with the Centre de Recherche en Information Biomedicalee Sino-Francais, 35042 Rennes, France, also with the National Institute for Health and Medical Research, F-35000 Rennes, France, and also with the Laboratoire Traitement du Signal et de l'Image, Université de Rennes 1, F-35000 Rennes, France (e-mail: jean-louis.dillenseger@univ-rennes1.fr; jean-louis.coatrieux@univ-rennes1.fr).

I. Introduction

UTERINE fibroids are benign tumors, common and present in up to 25% of women [1]. High intensity focused ultrasound (HIFU) is a new noninvasive surgery method for treating uterine fibroids. Magnetic Resonance (MR) image is clinically used for their diagnosis and the guidance of the HIFU procedure. The segmentation of uterus and uterine fibroids is a prerequisite step for the planning of HIFU treatment. However, the segmentation of the spine is also important in order to avoid any injury to the spinal cord. Manual delineation of the uterus, fibroids, and spine is a time-consuming, tedious task and subject to intra-expert and inter-expert variability during both pre- and post- treatment. Thus, an automatic and accurate segmentation method capable to extract all these structures is of great importance.

Such an objective is challenging because of **1) large shape and size variations among individuals.** As it is shown in Fig. 1, uterine and fibroids are highly variable in different patients; **2) a low contrast between adjacent organs and tissues.** The contrast among uterus and uterine fibroids is quite low, so the boundaries between organs are difficult to distinguish; **3) the number of uterine fibroids and their shapes are unknown.** These issues are illustrated in Fig. 1. Due to the above reasons, the existing methods dealing with uterine fibroid segmentation are often applied after treatment, while the pre-treatment is still performed manually by an operator to mark uterus, fibroids and surrounding organs. Therefore, in order to facilitate the development of a treatment plan, a preoperative segmentation is required.

In recent years, deep learning (DL) methods have been widely used in medical image segmentation [2]–[4]. However, they have to face the overall complexity of the scenes under study. We propose here to derive comprehensive anatomical information through a global convolutional network (GCN) module based on a large valid receptive field and deep multiple atrous convolutions (DMAC) for hierarchically structuring the information. By doing so, the performance in locating and classifying the structures of interest can be improved.

Such semantic segmentation can be built upon the Encoder-Decoder architecture already widely utilized. Inspired by

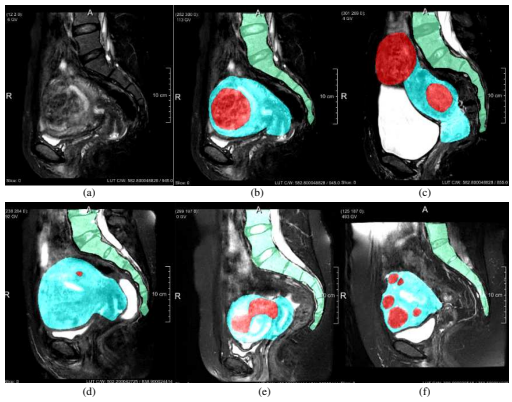

Fig. 1. MR images of uterus regions in different patients. Red denotes the fibroids, blue the uterus, and green the spine. (a) Patient 71 slice14 of raw MR image. (b-f) The labeled images of Patient 71 slice14, Patient 84 slice14, Patient 93 slice12, Patient 26 slice12, Patient 8 slice13. We can observe 1) large shape and size variations among individuals; 2) a low contrast between adjacent organs and tissues; 3) highly variable uterine fibroids numbers and shapes.

Fully Convolutional Network (FCN) [5] which was initially designed for image classification, U-Net was proposed for medical image segmentation by Ronneberger *et al.* [6] where the pooling operators in FCN are replaced by upsampling operators so that the output resolution can be retained at the same size as the input. The state-of-the-art results of U-Net in segmenting medical images, especially with small training dataset, show a promising ability of this Encoder-Decoder architecture. Basically, the Encoder aims to capture features and reduce the spatial dimensions while the Decoder aims to recover the object details and spatial dimension. Therefore, in order to improve the performance of image segmentation, more high-level features need to be automatically captured in the encoder and more spatial information can be saved in the decoder.

The U-Net was later extended in order to tackle different problems. Cicek *et al.* [7] modified the initial U-Net architecture by replacing all 2D operations with their 3D counterparts. Milletari *et al.* [8] presented a novel 3D segmentation approach (called V-Net) that leverages the power of a fully convolutional neural network based on the Dice coefficient for processing volumetric medical images such as MR images. In addition, in contrast with 3D U-Net, the V-net formulates each stage by using a residual function which can accelerate the convergence rate. Many other U-Net based segmentation schemes have been further reported for retinal vessels, liver and tumors in CT scans, ischemic stroke lesion, intervertebral disc and pancreas [9]–[20].

The U-Net shows a good segmentation performance with the usage of skip connections which can concatenate two feature maps of the same size in the corresponding parts of the encoder and decoder. The concatenated feature maps contain the information from both high and low levels, thus achieving feature fusion under different scales to improve the accuracy of model results. Even so, the complex anatomical scene involved in our HIFU therapy application remains a challenge. Large valid receptive fields play an important role in global scene observation. Global convolutional network [21] enables dense connections within a large region by using spatial decomposed convolution with a large kernel. It can capture multi-scale context cues with less computational cost than a general convolution with a large kernel. Therefore, we introduce layer-by-layer the GCN which has an efficient kernel parameter number to enlarge the receptive field in our Encoder-Decoder architecture.

In addition, getting the hierarchical structural information can help to provide more contextual information at various levels by using atrous convolutions. The key element of this method is to insert holes into the convolution kernels, which allows preserving the resolution and enlarging the receptive field. Recently, atrous convolution has been widely used in many deep learning architectures. DeepLab [22], based on FCN and atrous convolutions, maintains the receptive field unchanged. Besides, in order to get a better object segmentation at multiple scales, in DeepLabV2 [23], Chen *et al.* proposed a module called atrous spatial pyramid pooling (ASPP) which uses multiple parallel atrous convolutional layers with different sampling rates. The use of atrous convolutions preserves the spatial resolution of the final map and thus leads to higher performance when compared to most methods in Encoder-Decoder schemes. DeepLabV3+ [24] combines the advantages of Xception [25] and Encoder-Decoder, which employs DeepLabV3 [26] as the encoder.

However, the uncertainty regarding the location, the numbers and the sizes of uterine fibroids leads to an increase of complexity for segmentation and many existing deep learning segmentation models lack using features from different levels efficiently. Subsequently, in some cases, the targets can be segmented incorrectly. More effective feature extraction approaches are required for uterine fibroid segmentation.

Motivated by the above discussions and ResNet [27] structures, we propose a novel network named HIFUNet to segment uterus, uterine fibroids and spine automatically. The main contributions of the paper can be summarized as follows:

1) To address the segmentation errors (*i.e., classifying uterine neck as uterine fibroid because of insufficient receptive field*), we introduce a GCN module able to enlarge the receptive field effectively.

2) We integrate the GCN and DMAC to further extract context-based semantic information and generate more abstract features for large scaled uterine fibroids.

3) The proposed HIFUNet behaves similarly to clinical experts and, as it will be shown through a large number of experiments, performs better than many existing semantic segmentation networks.

4) The segmentation of the uterus and uterine fibroids is, to the best of our knowledge, the first methodological attempt using convolutional neural networks in HIFU therapy. The inclusion of the spine segmentation, a critical organ in HIFU therapy, is another major feature of our approach.

The structure of this paper is as follows: In Section II, we describe up-to-date related work. Our solution is then introduced in section III. Our experiments are reported in section IV, including performance comparisons with conventional and other deep learning methods. In section V, we draw some conclusions and perspectives.

II. Related Work

We sketch here the conventional methods proposed so far for segmenting the uterus and uterine fibroids and we review the state-of-the-art MR image segmentation methods based on CNN architectures.

A. Conventional Methods of Uterus and Uterine Fibroid Segmentation

Very few contributions have been reported for segmenting uterus and uterine fibroids from MR images. The main methods are summarized below:

Approaches based on level-set: Ben-Zadok *et al.* [28] presented an interactive level set segmentation framework that allows user feedback. It is a semi-automatic method where the users have to select seed-points. Khotanlou *et al.* [29] proposed a two-stage method combining the region-based level set [30] and the hybrid Bresson methods [31]. Yao *et al.* [32] employed a method based on a combination of fast marching level-set and Laplacian level set.

Approaches based on Fuzzy C-Means (FCM): Fallahi *et al.* [33] segmented the uterine fibroids by combining a fuzzy C-means method with some morphological operations. Later, on the basis of [33], a two-step method [34] was proposed by employing a Modified Possibilistic Fuzzy C-Means (MPFCM) [35] in a second step.

Approaches based on region-growing: Militello *et al.* [36] used a semi-automatic approach based on region-growing and reported a quantitative and qualitative evaluation of the HIFU treatment by providing the 3D model of the fibroid area. Rundo *et al.* [37] presented a two-phase method where the first phase is an automatic seed-region selection and region detection while the second one is aimed at uterine fibroid segmentation.

Other mixed methods: Antila *et al.* [38] designed an automatic segmentation pipeline without user input. They applied the active shape model (ASM) to get the deformed surface, and classified PV (perfused volume: the untreated tissue) and NPV (nonperfused volume: the treated tissue) by an expectation maximization (EM) algorithm. Militello *et al.* [39] proposed a novel fully automatic method based on the unsupervised Fuzzy C-Means clustering and iterative optimal threshold selection algorithms for uterus and fibroid segmentation.

Recently, Rundo *et al.* [40] evaluated the above mentioned two computer-assisted segmentation methods [37], [39] and provided a quantitative comparison on segmentation accuracy in terms of area-based and distance-based metrics. Their results show that both methods remarkably outperform the other ones.

However, there are still some limitations and drawbacks in the conventional methods and a fully-automatic and accurate method, able to reduce or even to remove pre-processing/post-processing procedures as well as the interventions of the medical physicists, is still expected. For this purpose, a detailed comparison between the methods reported in [37] and [39] and our method will be shown in Section IV.

B. Deep Learning Methods of MR Image Segmentation

Only a few attempts have been reported for the uterus segmentation using CNN-based methods. Kurata *et al.* [41], [42] evaluated the clinical feasibility of fully automatic uterine segmentation on T2-weighted MR images based on an optimized U-Net. The segmentation of uterus in this research was focused on the staging of uterine endometrial cancer and on estimating the extent of tumor invasion to the uterine myometrium. To the best of our knowledge, there is no literature published on the uterine fibroid segmentation using CNN-based methods. Even so, it is important to highlight that many innovative deep learning methods have been proposed for MR image processing [43], [44]. The most common applications concern segmentation of organs, substructures, or lesions, often as a preprocessing step for feature extraction and classification. Deep learning methods for MR image segmentation can be divided into two different categories.

DL based on image patches: Features are extracted from a local patch for every voxel using convolutional layers. These features are then classified with a fully connected neural network to obtain a label for every voxel. This method is for instance widely used in brain tumor [45], white matter segmentation in multiple sclerosis patients [46], normal components of brain anatomy [47] and rectal cancer segmentation [48]. However, such methods have some disadvantages. The main problem is that their computational efficiency is very low because they have to process overlapping parts of the image. Another disadvantage is that each voxel is segmented based on a finite size context window, ignoring the broader context. In some cases, more global information may be needed to properly assign these labels to pixels or voxels.

Fully convolutional neural network (FCNN): In this case, the entire image or a large portion is processed, the output being a segmentation result instead of a label of a single pixel or voxel. Such an approach solves the shortcomings of the former method and improves the efficiency of the algorithm. Many architectures can be considered for segmentation among which, as mentioned in Section I, encoder-decoder ones such as U-Net and its modified versions [9]–[20]. For MR images, we refer to [43] for a full survey. Zhang *et al.* [49] used CNN for segmenting the infant brain tissues by combining T1, T2, and FA images into white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF). Brain tumor segmentation was addressed in [50]. Avendi *et al.* [51] associated DL algorithms with deformable models for the left ventricle segmentation of the heart. Milletari *et al.* [8] proposed a 3D image segmentation based on a volumetric, fully convolutional, neural network. Their CNN was trained end-to-end on MR image volumes depicting the prostate and learned to predict segmentation for the whole volume at once. Some universal architectures were also proposed (for instance CE-Net by Gu *et al.* [52]) to address different clinical applications.

However, our target presents significant differences with these examples (i.e. brain, prostate, and heart). The deformation of the uterus shape is very large among the patients. The uterus position is also varying a lot. The high number of surrounding organs together with their similarity in tissue

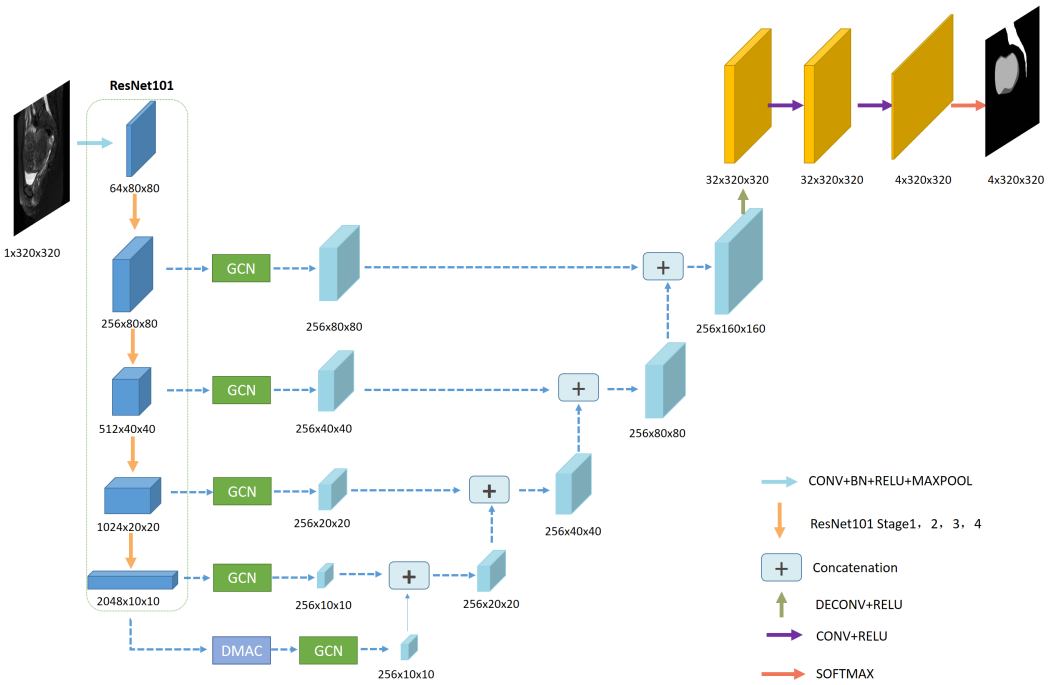

Fig. 2. The architecture of our proposal network (HIFUNet). The network consists of Resnet101 backbone, GCN module, DMAC module, upsampling layers, concatenation layers, and an output layer. The parameters and sizes of output features in different layers are presented in different colors.

features makes more challenging the segmentation. In addition, different kinds of uterine fibroids (such as subseries fibroids, submucosal fibroids, intramural uterine fibroid tumors, pedunculated leiomyomas, and parasitic uterine fibroids) may be located in different regions of the uterus, and the gray level of these fibroids are affected by the signal intensity and other experimental factors. All these considerations have guided the design of our approach.

III. Method

To accurately segment the uterus, uterine fibroids and spine from the raw MR images, we propose an Encoder-Decoder global convolutional network. The whole pipeline is illustrated in Fig. 2. This network (called HIFUNet) consists of three major parts: the feature encoder module, the feature extractor part (with the global convolution network and deep multiple atrous convolutions) and the feature decoder module.

A. Encoder Module

The encoder part uses pre-trained ResNet101 [27]. In [53], the authors demonstrated that the use of residual connections promotes information propagation both forward and backward, so it helps to improve significantly both the training speed and the performance. Because we have only one channel in our raw 2D input image, we change the original first portion which forms three input channels to one channel and we obtain 64 channels after the first Conv1. Then, four feature extracting blocks are employed. The first, second, third, and fourth stages contain 3, 4, 23, and 3 bottlenecks respectively and each block has no average pooling layer or fully connected layers.

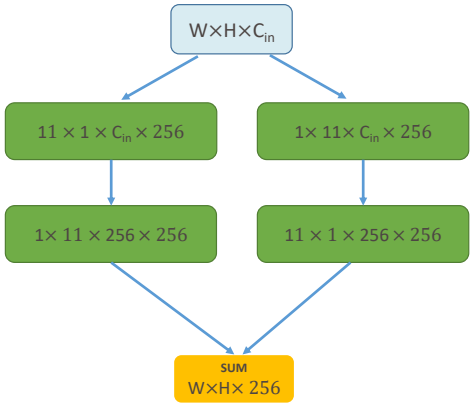

Fig. 3. Global Convolutional Network

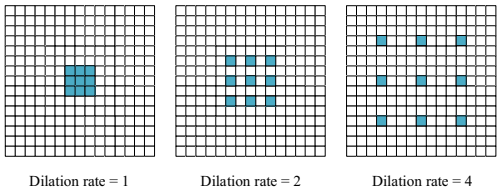

Fig. 4. Atrous convolutions with $\mathbf{3} \times \mathbf{3}$ kernel (blue blocks) and rates 1, 2 or 4.

B. Global Convolution Network

The current trend in architecture design goes toward stacking small convolution kernels because this option is more

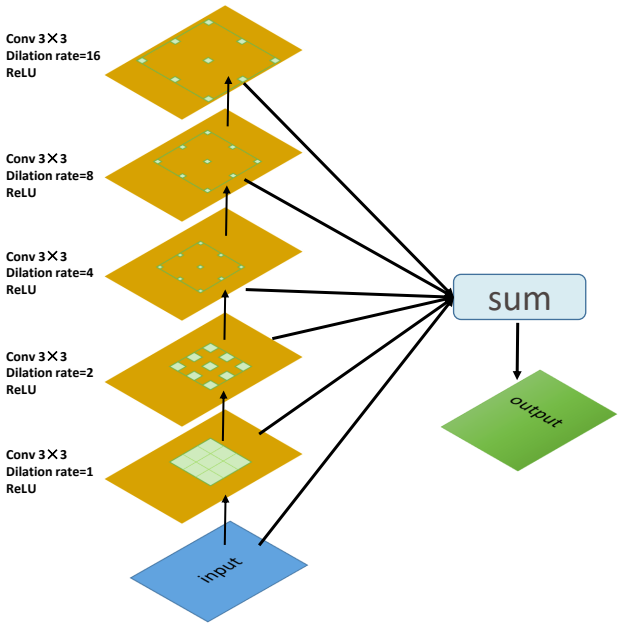

Fig. 5. Deep multiple atrous convolutions (DMAC) consist of five atrous convolutional layers.

efficient than using a large convolution kernel with the same amount of computation. However, considering that semantic segmentation tasks require pixel-by-pixel segmentation prediction, Peng *et al.* [21] proposed a global convolutional network to improve the accuracy of classification and localization simultaneously. In GCN, a fully-convolutional layer is adopted to replace the global pooling layer in order to keep the localization information. Besides, large kernels are introduced to increase the valid receptive field (VRF). However, using a large kernel or a global convolution directly is inefficient. To further improve the computational efficiency, GCN uses a combination of two large 1D convolutional kernels to replace a single 2D kernel for the skip-connector layer. The architecture of GCN is shown in Fig. 3. The kernel size we use in our segmentation approach is 11×11.

C. Deep Multiple Atrous Convolutions

Atrous convolutions solve the problem of reduced resolution caused by the Deep Convolutional Neural Networks (DCNNs) while adjusting the receptive field of the filter. Fig. 4 illustrates the atrous convolution. The main idea of atrous dilation rate convolution is to insert "holes" (zeros) between pixels in convolutional kernels to increase the image resolution, enabling thus dense feature extraction in DCNNs. The atrous convolution was initially proposed to efficiently compute the undecimated wavelet transform [54] and the wavelet decomposition [55] in the atrous scheme. In recent years, atrous convolution has been widely used in tasks such as semantic segmentation and object detection. The Deeplab series [22]–[24], [26] and dense upsampling convolution (DUC) [56] made thorough studies of atrous convolution. Fig. 5 shows our proposed deep multiple atrous convolution scheme to achieve multi-scale representations. We implement five convolutional layers with 3×3 kernels with different sampling rates to extract the different features. Finally, we fuse all features with the input image to generate the final result.

When compared to the conventional network structure, our deep multiple atrous convolutions can extract multiple features and provide receptive fields of multiple sizes. It can be noticed that the architecture of our atrous convolution scheme adopts a serial frame instead of a parallel structure such as Inception and Atrous Spatial Pyramid Pooling (ASPP). We employ the DMAC block in the final layer of the encoder and this way more abstract information can be exploited. Within the DMAC block, as the layer is deeper, the dilation rate is getting larger. Because of the kernel discontinuity, not all pixels are used for calculation, so more atrous rate convolutions can compensate for the uncalculated information in the serial structure, which can increase the receptive field effectively. Besides, different sizes of atrous rates can help to extract different sized targets (from small fibroids to large organs like uterus or spine). The serial structure can get global distribution information from various scales of atrous convolution. The final step sums up as the output the abstract information extracted from the multiple layers. This output is then sent to the decoder phase in order to recover the object details and spatial dimensions. Therefore, in order to improve the performance of image segmentation, more low and high-level features are automatically captured in the encoder.

D. Decoder Module

The decoder module mainly uses the concatenation operation to fuse the multi-scale features. U-Net concatenates the downsampling feature maps with the corresponding upsampling feature maps. Here, this concatenation is performed between two neighboring feature maps after the GCN modules and this from the bottom to the top. After four concatenation operations, the image scale increases from 1/32 to 1/2 of the input image size. Then, we use a deconvolution operation to enlarge the image scale to the initial size and to restore features with more detailed information. Finally, the output mask is obtained after applying two convolution operations and softmax. As illustrated in Fig. 2, the decoder module mainly includes four concatenation operations (a 1×1 convolution, a 4×4 transposed convolution, and two 3×3 convolutions consecutively). Then, the feature decoder module outputs a mask with the same size as the original input.

E. Loss Function

The HIFUNet can be trained by minimizing the cross-entropy error between its prediction result and the ground-truth. The loss function is defined as

$$L = \sum_{i \in \Omega} y_{c_i} \log(p_{c_i}) + (1 - y_{c_i}) \log(1 - (p_{c_i})) \qquad (1)$$

where p_{c_i} denotes the predicted probability of c-th class for pixel i in the predicted result p, $y_{c_i} \in \{0, 1\}$ is the corresponding ground-truth value. If $y_{c_i} = 1$, it means that

pixel i belongs to the c-th class. If $y_{c_i} = 0$, it means that pixel i does not belong to the c-th class. $c = 0$ denotes the background, $c = 1$ denotes the uterus, $c = 2$ denotes the uterine fibroids while $c = 3$ denotes the spine. Ω denotes the space of the predicted result of p and the ground-truth y. By minimizing the loss function on a training database, the parameters of HIFUNet can be optimized. Then the trained HIFUNet can be applied for automated uterus, uterine fibroids and spine segmentation on different datasets.

F. Further Discussion

The main difference between our HIFUNet and other state-of-the-art deep learning networks including GCN, HRNet, U-Net, CE-Net, AttentionUNet, and LEDNet is summarized as follows:

• GCN uses large kernels to enlarge the effective receptive field which can help classify different objects.

Different from GCN, in order to exploit more abstract information, HIFUNet adds an original DMAC block which improves the accuracy of segmentation of key parts such as the cervix and minor fibroids.

• HRNet relies on a parallel structure enabling the model to connect multi-resolution subnetworks in a novel and effective way. It starts from a high-resolution subnetwork as the first stage and gradually adds high-to-low resolution subnetworks one by one to form more stages, the multiresolution subnetworks being connected in parallel.

The main difference is that HIFUNet and HRNet use different ways for computing high-resolution representation. Our HIFUNet employs the way of recovering high-resolution representations from low-resolution representations outputted by a network (eg., ResNet). While in HRNet, the authors propose another way that maintaining high-resolution representations through high-resolution convolutions and strengthening the representations with parallel low-resolution convolutions.

• U-Net uses a simple downsampling way to extract features while HIFUNet uses ResNet101 as the backbone to extract more features. We add large kernels in the skip-connections to increase the valid receptive field (VRF).

• CE-Net uses the Dense Atrous Convolution (DAC) module with multi-scale convolution and the Residual Multi-kernel Pooling (RMP) with multi-scale pooling at the bottom to extract and decode multi-scale features in parallel, as well directly integrate them. It ignores the global scene content at each level which further enhance the localization effect of the skip connection, as well as the progressivity and the correlativity among the multi-scale structure.

Especially different from the CE-Net, the proposed HIFUNet adopts GCN in each skip connection between the encoder and the decoder. So that it is able to embed global scene information in the decoder, avoiding the global scene information loss in the dimension reduction during encoding. Besides, the HIFUNet also employs DMAC with the series structure and hierarchical fusion at the bottom of the encoder to progressively and correlatively extract multi-scale structure for the semantic objects.

• AttentionUNet proposes a novel Attention Gate (AG) model for medical imaging that automatically learns to focus on target structures of varying shapes and sizes, which brings a risk of transmitting multiplicative error along with the network.

CE-Net and AttentionUNet are both based on the U-Net and keep the way of extracting features in the encoder of U-Net. Differently, we choose to use a ResNet-101 pre-trained on Imagenet as our backbone, which can be easier to train Resnet than training simple deep convolutional neural networks and resolove the problem of accuracy degradation.

• LEDNet aims at real-time semantic image segmentation. It employs an asymmetric encoder-decoder architecture. The encoder adopts a ResNet as the backbone network, where two new operations, channel split and shuffle, are utilized in each residual block to greatly reduce the computational cost while maintaining a higher segmentation accuracy. On the other hand, an Attention Pyramid Network (APN) is employed in the decoder to further decrease the entire network complexity.

In our task, we pay more attention to the segmentation accuracy than to the efficiency of training. In the decoder part, LEDNet focuses on the last feature map from the encoder network, while some low-level features can be let out, which is not conducive to recovering detailed information . Therefore, we choose to recover the high-resolution information by concatenating low- and high-level features, which can help to identify the objects of all sizes and the details in complex medical images.

IV. Experiment and Discussions

A. Datasets

The preoperative fat-suppressed T2-weighted MR images in the sagittal direction from 297 patients were used in this work. These images were collected from the First Affiliated Hospital of Chongqing Medical University. Sagittal T2-weighted fast spin-echo images were performed using a 3.0T MR unit (Signa HD Excite, GE Healthcare, Marlborough, MA) with an eight-channel phased-array coil. The scan parameters and characteristics of MR images are shown in TABLE I.

TABLE I
The Scan Parameters and Characteristics of MR Datasets

Variable	Value
Repetition time (TR)	3040 ms
echo time (TE)	107.5 ms
field of view (FOV)	28×22.4 cm
slice thickness	6 mm
slice gap	1 mm
matrix	304×304
age (years)	$40.8 \pm 6.6^*$

* Age is $Meanvalue \pm S.D.$

Each MR volume consists of 25 slices of 304×304 pixels. The ground truth has been generated through a proper annotation process. To ensure an objective and consistent clinical reference, two radiologists were solicited for consensus agreement. This procedure included three steps:

1) Annotations through discussions: The discussion between

two radiologists A (7-year experience) and B (15-year experience) was held in a face-to-face mode to set the annotation rules and identify special and complicated cases. It appeared, in this application, that the variability of the annotations mainly exists on the contour of the cervix and some minor fibroids.

2) The radiologist A took 2 months in annotating (no more than 5 volumes per day). After annotating 10 volumes, a second face-to-face discussion was held to analyze the first-round annotation, and improve the annotation rule further.

3) Then the radiologist A processed all cases (297 patients). Radiologist B checked all results and marked the cases which have some divergent views. Then, they held a face-to-face discussion and solved these situations.

After the above three steps, a full agreement between the two radiologists was obtained.

The research associated with the treatment of uterine fibroids was approved by the ethics committee and has no implication on patient treatment.

B. Experimental Setup

1) Training and testing phase: We used for training and testing MR images from 260 and 37 patients, respectively. The number of images in the testing set is 925. The use of a small amount of training data can result in overfitting. To prevent overfitting due to the limited number of images, the training data were augmented by image manipulation [57]. We applied the random shifting and scaling strategies (zoom range of 0.1, the shift of 0.5mm).

2) Parameter settings and platform: For the optimization of our network, we use the Adam optimizer and set the initial learning rate to 2e-4. After each epoch, if we observe that the validation loss does not decrease for three consecutive times, the learning rate is reduced to 1/5 of its current value until it stops at 5e-7. Therefore, the number of training epochs is determined by the decreasing learning rate. The batch size is set to 8. All the comparative experiments adopt the same strategy for updating the hyperparameters. Besides, in the ablation study, the hyperparameters are fixed when removing parts of the network.

Our proposed network is based on the pretrained ResNet101 model on ImageNet. Notice that we adapt the first convolution operation because, as mentioned in III.A, we have a single channel input image instead of RGB channels like in natural images. The implementation is carried out on the PyTorch platform. The training and testing bed are ubuntu 16.04 system with NVIDIA Titan XP GPU (12 GB memory) and CUDA 9.0.

C. Evaluation Metrics

Different quantitative measures are used to comprehensively evaluate and compare the segmentation performance with other methods. We use the area-based indexes to compare the predicted segmentation results with the ground-truth manually labeled by radiologists. These indexes include the Dice coefficient (DSC) [58], Precision [59], Sensitivity (SE) [60], Specificity (SP) [60], Jaccard index (JI) [61], False Positive Ratio (FPR), False Negative Ratio (FNR) and False Region Ratio (FRR) [40]. We also use the distance- based indexes to evaluate the segmentation in terms of the location and shape accuracy of the extracted region boundaries such as the Mean Absolute Distance (MAD) [40], Maximum Distance (MAXD) [40] and Hausdorff Distance (HD) [62].

D. Comparison with Conventional Methods and Discussion

As mentioned in Section II.A, Rundo *et al.* [37] and Militello *et al.* [39] proposed to segment uterine fibroids after treatment and evaluated them in [40]. We compare their methods with our method on the same dataset (fat-suppressed T2-weighted MR images composed of 375 slices issued from 15 patients).

It can be noticed that the above two methods are based on the fact that ablated fibroids appear as homogenous hypo-intense regions with respect to the rest of the uterus (after contrast medium injection). Before the treatment, all kinds of fibroids appear as different states, which makes the segmentation task harder. For all patients, area-based and distance-based indexes were computed based on a slice-by-slice comparison and were performed on each slice having a fibroid area. The results are displayed in TABLE II. They show the superiority of the proposed method over the other two approaches and demonstrate its ability for uterine fibroid segmentation.

Some visual results are depicted in Fig. 6. It can be seen that, in the Patient 4, the gray level values around the area outlined by the circles have little difference from adjacent tissue. While in the post-treatment MR images the ablated tissue does not absorb the contrast medium and is hypo-intense with respect to the uterus, the use of simple adaptive global thresholding and region growing methods remains possible. However, the quality of the MR images is affected by noise which may lead to gray values in the regions of uterine fibroids similar to those of the surrounding tissues. As it is shown in Patient 7, there are two fibroids that appear with different signal strengths because of the different moisture contents: one is dark and the other one is bright. Thus, it is difficult for IOTS to distinguish the two different grayscale distributions of fibroids. SM&RG fails to identify the contour of fibroids and assimilates the uterus to fibroids. The segmentation provided by our DL method is close to the ground-truth segmented by the clinical experts.

Additional comments on the two methods used here for comparison deserve to be made. The uterus ROI segmentation is a preliminary step for a robust fibroid detection in [40]. This task can be accomplished manually by the user to remove parts outside the uterus which are present in sagittal sections [37] or can rely on the Fuzzy C-Means (FCM) [39], which is an automatic method but where the number of clusters is set according to a visual inspection (i.e., anatomical properties of the analyzed pelvic images by considering image features) and experimental evidence (by means of segmentation trials). It means that the intervention of the experts is indispensable and that a complex and time-consuming preprocessing is needed before applying the intensity-based clustering technique. In conclusion, although these conventional methods have some

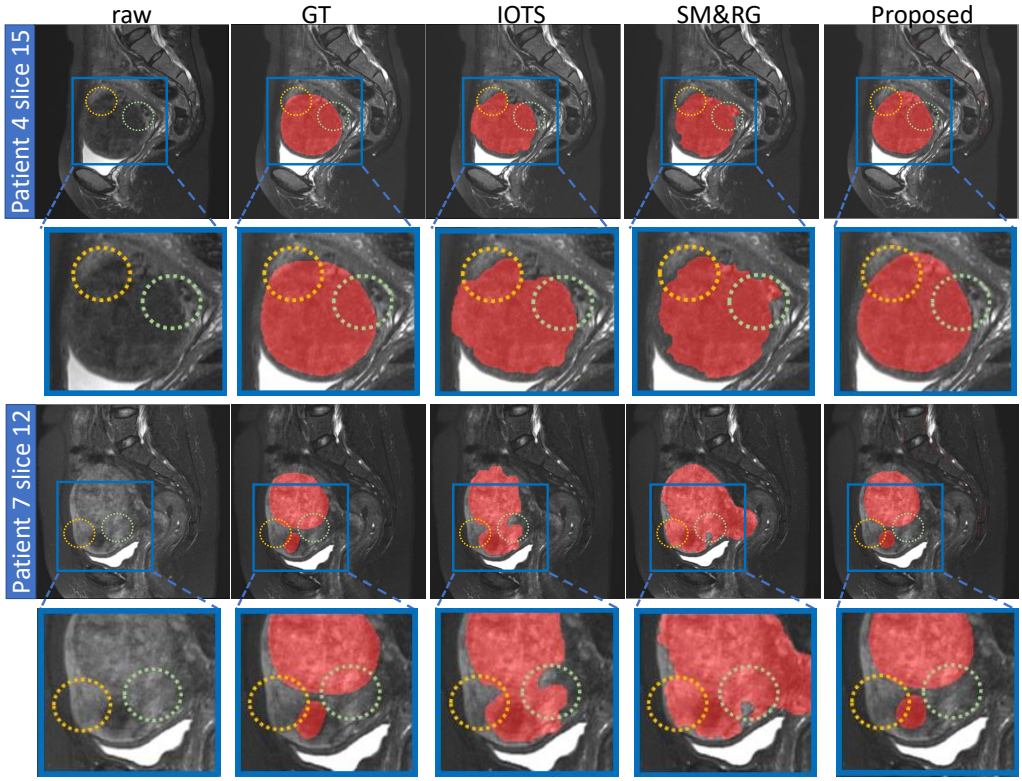

Fig. 6. Visualization of the segmentation results of uterine fibroids by using the proposed method and methods in [37][39] for two patients. Red denotes the fibroids, and the yellow and green circles point out incorrect segmentation of uterine fibroids due to the little gray value difference with surrounding tissues.

TABLE II
VALUES OF AREA-BASED AND DISTANCE-BASED FOR SEGMENTING UTERINE FIBROIDS USING DIFFERENT METHODS ON T2-WEIGHTED MR IMAGES

Method	area-based								distance-based		
	DSC(%)	Precision(%)	SE(%)	SP(%)	JI(%)	FPR	FNR	FRR	MAD	MAXD	HD
ITOS [39]	80.50	76.83	89.03	98.22	69.34	0.018	0.110	0.540	2.432	7.893	8.893
SM&RG [37]	81.15	77.74	89.47	98.33	72.13	0.017	0.105	0.429	3.422	11.536	12.935
Proposed	86.58	88.17	88.45	99.53	78.45	0.005	0.116	0.709	2.955	9.365	16.372

merits in terms of performance, they show some practical limits in the clinical setting.

E. Comparison with Other Deep Learning Methods

We compare our method with six state-of-the-art (SOTA) algorithms, including U-Net [6], Attention-UNet [20], GCN [21], CE-Net [52], HRNet [63], LEDNet [64]. Their original implementations were kept and the same experimental conditions were used.

We select four of these competitive methods (U-Net, GCN, HRNet and CE-Net) to visually compare our method in Fig. 7 where the segmentation results are overlaid on the raw images. Different colors denote different classes (red denotes the fibroids, blue the uterus and green the spine). The images show that our method provides more accurate results. The performance of the six selected methods is presented in TABLE III for quantitative comparison. Among them, HRNet is the best method for segmenting uterus and fibroids. Besides, for the spine which has a high contrast with adjacent tissues, the introduction of the attention mechanism (i.e. AttentionUNet) performs quite well. However, overall the best results are obtained by our method. Regarding the computation cost, we choose the GPU memory requirements and the test time for evaluating each slice to evaluate. Because of using ResNet as our backbone, our HIFUNet has a larger number of parameters. However, in clinical applications, the accuracy of the segmentation is much more important than the computation cost. From TABLE III, we can see that the performance of HIFUNet is significantly better in comparison to the other methods. We found it acceptable that the increases in computational costs are negligible for the improvement in accuracy. The computational cost of our method at test time can be borne by a standard GPU.

As can be seen from Fig. 7, the fibroids are more difficult to segment than the uterus, due to their unclear boundaries and undefined shapes. For patient 9, GCN and HRNet fail to

TABLE III
QUANTITATIVE COMPARISON OF THREE EVALUATION INDEXES OF DIFFERENT SEGMENTATION METHODS ON TESTING DATASET (THE BEST RESULTS ARE INDICATED IN BOLD)

Method	Uterus			Fibroids			Spine			Memory	Test time
	DSC	Precision	Recall	DSC	Precision	Recall	DSC	Precision	Recall		
GCN [21]	79.44%	79.27%	80.37%	80.43%	82.88%	80.04%	80.50%	85.14%	77.74%	464.96M	108.25ms
HRNet [63]	80.43%	78.29%	83.45%	80.88%	**85.39%**	80.76%	**85.45%**	83.77%	86.50%	561.88M	165.55ms
U-Net [6]	75.34%	76.97%	74.81%	77.58%	78.39%	79.23%	78.15%	**89.10%**	71.46%	317.97M	**14.56ms**
CE-Net [52]	74.69%	75.42%	74.99%	76.38%	75.05%	80.66%	82.48%	86.99%	79.15%	123.22M	105.77ms
AttentionUNet [20]	74.79%	76.08%	74.56%	76.24%	74.97%	81.18%	83.28%	88.54%	79.25%	927.34M	159.12ms
LEDNet [64]	77.87%	77.10%	79.46%	78.92%	83.71%	76.12%	79.02%	87.19%	74.19%	**121.37M**	73.84ms
Proposed	**82.37%**	**79.45%**	**86.00%**	**83.51%**	84.48%	**83.70%**	85.01%	82.51%	**88.69%**	503.71M	109.83ms

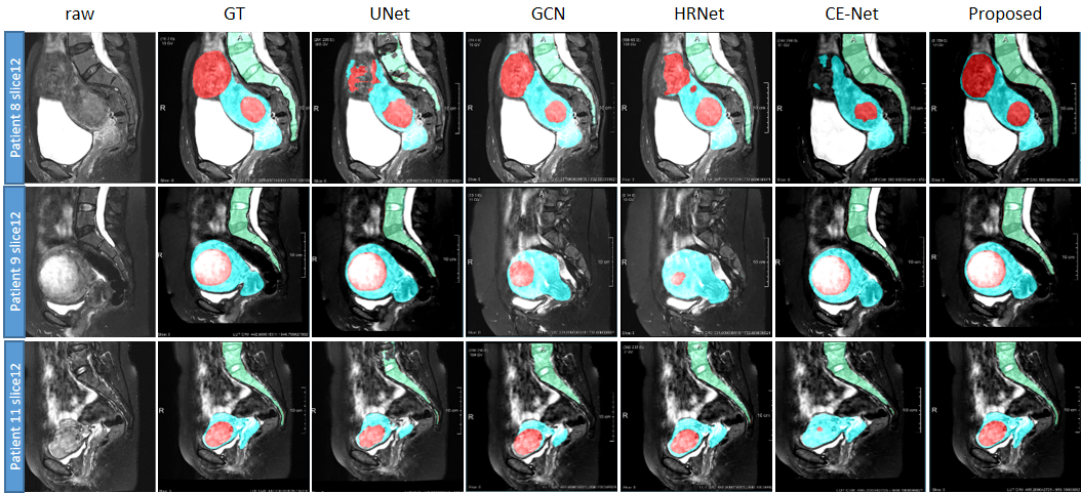

Fig. 7. Visualization of the segmentation results of uterus, fibroids and spine by using the proposed method and other four SOTA methods. From top to bottom are three different patients. Red denotes the fibroids, blue denotes uterus, and green denotes spine.

segment the spine. For patient 8, U-Net, HRNet and CE-Net lead to incomplete segmentations. We can also observe the crucial role of the large receptive field used in our approach. Fig. 8 and Fig. 9 show the DSC of uterus and fibroid segmentation results in the form of box plots. Our method provides the best and steadiest performance in segmenting both uterus and fibroids while the performance of HRNet is slightly weaker.

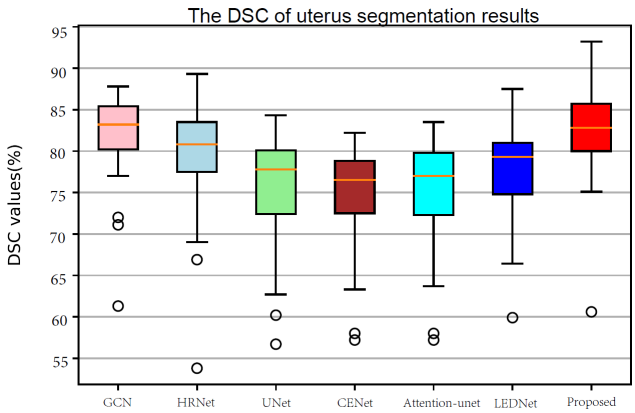

Fig. 8. The qualitative uterus segmentation performance is presented as boxplots. The y axis indicates the DSC values, while the x axis corresponds to the different methods (Unfilled circles denote the suspected outliers).

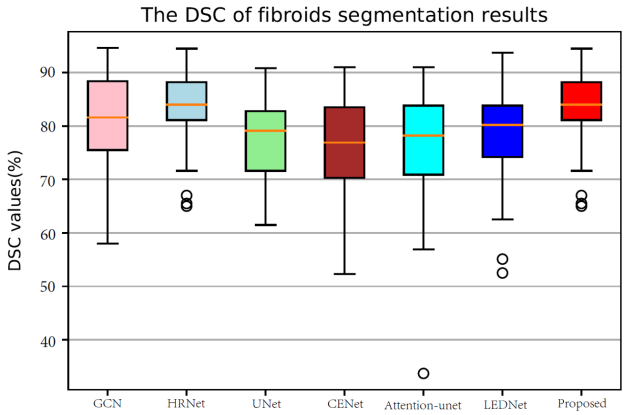

Fig. 9. The qualitative fibroid segmentation performance is presented as boxplots. The y axis indicates the DSC values, while the x axis specifies the different methods (Unfilled circles denote the suspected outliers).

F. Ablation Study

DMAC block We first conducted ablation studies and validated the effectiveness of our DMAC block using the same training strategy and datasets. The original GCN (GCN-no DMAC [21]) was compared with the modified GCN (GCN-DMAC) with a DMAC block added in the last layer. In the proposed HIFUNet (Proposed-DMAC), the DMAC block was

TABLE V
THE PERFORMANCE ON TESTING DATASET BY USING DIFFERENT DNCODER METHODS
(THE BEST RESULTS ARE INDICATED IN BOLD)

Method	Uterus			Fibroids			Spine		
	DSC	Precision	Recall	DSC	Precision	Recall	DSC	Precision	Recall
Backbone+JPU	66.26%	70.44%	63.37%	67.36%	70.20%	66.74%	66.07%	76.90%	58.79%
Backbone+CS	76.37%	**80.77%**	73.27%	80.06%	79.55%	83.31%	83.88%	**87.45%**	81.37%
Backbone+CD(Proposed)	**82.37%**	79.45%	**86.00%**	**83.51%**	**84.48%**	**83.70%**	**85.01%**	82.51%	**88.69%**

TABLE IV
THE MEAN DSC AND COMPUTATION TIME OF DIFFERENT SEGMENTATION METHODS USING DMAC BLOCK
(THE BEST RESULTS ARE INDICATED IN BOLD)

Methods	DSC			Time(s)
	Uterus	Fibroids	Spine	
GCN-no DMAC	79.44%	80.43%	80.50%	164
GCN-DMAC	80.15%	81.08%	80.01%	**161**
Proposed -no DMAC	76.87%	78.84%	84.28%	479
Proposed -DMAC behind	77.72%	77.47%	80.89%	441
Proposed-DMAC	**82.37%**	**83.51%**	**85.01%**	1094

put in the last layer and before the operation of global convolution. Comparisons were performed between the Proposed-DMAC, removal of DMAC block (Proposed-no DMAC) and insertion of the DMAC after the global convolutional operation (Proposed- DMAC behind). TABLE IV shows the results of this study together with the time needed for each training epoch. They point out that the segmentation results are not significantly improved for GCN-DMAC. Concerning the D-MAC position in our method, the computation time is strongly reduced when it is behind but the performance is worse than DMAC in-front (i.e. Proposed-DMAC).

Some images are shown in Fig. 10 for visual inspection. GCN leads to a relatively good segmentation of uterus and spine but the boundary of the fibroids is clearly inaccurate, and most parts of the fibroids fail to be labeled out. Adding the DMAC helps to refine the inaccurate boundary of the uterus and correct to some extent the wrong segmentation of fibroids. When replacing GCN by our proposed main structure, two fibroids are labeled out successfully with accurate boundaries (see Patient 20 slice 13) which shows the advantage of our main structure. In the same slice, by comparing GCN and the Proposed-no DMAC, the boundary of the spine is corrected, which confirms the previous observation. A slightly better result can be achieved with DMAC. In all cases, our method labels both the uterus and the inside fibroids accurately which shows the effectiveness of the proposed DMAC. In particular, by comparing the last two columns, we can conclude that DMAC can extract the features of a large receptive field in a multi-scale context from multi-level feature maps.

Decoder method In our approach, we replace the summation operation in GCN by a concatenation operation in U-Net. Besides, in the procedure of upsampling, the deconvolution operation is employed to recover the original image size and to get the output mask. Recent contributions focus on the use of an upsampling module to upsample a low-resolution feature map given high-resolution feature maps as guidance. For instance, Joint Pyramid Upsampling (JPU) [65] aims at generating a high-resolution target image by transferring details and structures from the guidance image. DUpsampling (DUP) [66] was also proposed to replace the standard bilinear upsampling to recover the final pixel-wise prediction. The DUP takes advantage of the redundancy in the label space of semantic segmentation and is able to recover the pixel-wise prediction from low-resolution outputs of CNNs.

We report here the experiments made in order to compare different ways of decoding. Inspired by Octave Convolution [67], in which Chen proposed to store and process low-frequency and high-frequency characteristics respectively, we plan to deal with low and high channels separately. Also motivated by the Inception module [68] which employed a split-transform-merge strategy, we design a Channel-Split module that splits channels of each feature map after the GCN module into high and low channels and then we use concatenation and summation operations to integrate features of different layers in a continuous way. Different from Octave Convolution in [67] which is an operation as a direct replacement of vanilla convolutions, Channel-Split (CS) is a decoder strategy to change the way of merging different channels from different layers. Another decoding method is shown in Fig. 2. It removes the operations of summation in each layer and mainly uses deconvolution and concatenation. We name it Concatenation-Decoding (CD).

We train the three networks, with JPU or CS or CD as decoder respectively. The backbone here is the encoder of ResNet101 with GCN block and DMAC block. DUP is not trained because there is no formal code implementation of it. The comparison experiments are based on the same training parameter settings over the same training and validation dataset. The quantitative assessment is performed on the same testing dataset. The implementation of the JPU refers to the official PyTorch version on https://github.com/wuhuikai/FastFCN.

As it is shown in TABLE V, CD is much better than JPU and CS methods, with a benefit in DSC varying between 6% and 16% for uterus. It can be concluded that concatenation helps to recover the features especially in complex contexts and multiple targets. The summation is applied in the shortcuts (skip connections) in ResNet. It can help the network to speed up the training process and improve the gradient flow since the shortcuts are taken from previous convolution operations. Therefore, it is effective for the backpropagation to transfer error corrections to earlier layers, which can address the

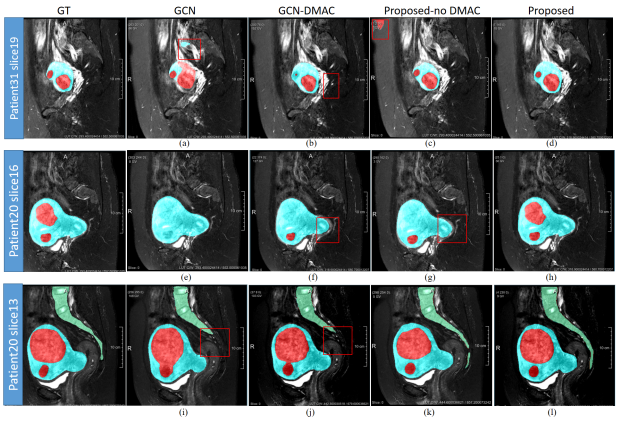

Fig. 10. Visualization of the segmentation results of uterus, fibroids and spine from two patients by using different methods which are mentioned in TABLE III. From left to right: ground-truth, GCN [21], GCN with DMAC, our proposed method without/with DMAC. Red denotes the fibroids, blue denotes uterus, and green denotes spine. Details are drawn by red box to see apparently.

problem of vanishing gradient. However, due to the summation of the different channels or feature maps in CS, it may be difficult for the networks to distinguish different targets or recover the object details in the decoder. In contrast, the concatenation in CD operates on the feature maps generated by different filter sizes and keeps the information of different resolution feature maps since the information of features is not lost by summing up. JPU mainly uses the last three layers in the encoder. Therefore, the features of multiple objects in our complicated context may not be fully exploited by employing JPU.

V. Conclusions

In this paper, we have proposed a global convolutional network with deep multiple atrous convolutions to segment uterus, uterine fibroids and spine automatically. The employment of the DMAC block allows capturing effectively more low and high-level features.

Experimental results on the same datasets and platform demonstrated (i) the accuracy and robustness of the proposed method, (ii) a significant improvement when compared to state-of-the-art segmentation methods and (iii) the performance could be close to radiologist level.

Although the proposed method shows promising results, some boundary inaccuracies may still be present in patients depicting multiple fibroids (see the left fibroid in the first row of Fig. 10). We plan to improve our approach by working directly in 3D (i.e. 3D convolutional filters) instead of dealing with 2D slices. This will make the training issues (improving efficiency and reducing training time) more critical. Other ideas should also be explored such as the use of prior anatomical and pathological knowledge on the uterus and spine. Coupling our approach with other techniques (active contour models, for instance) to refine the boundaries of the uterus and spine may also offer a sound way to correct the remaining errors mentioned above.

References

[1] E. A. Stewart, "Uterine fibroids," *The Lancet*, vol. 357, no. 9252, pp. 293–298, 2001.
[2] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, *et al.*, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, 2017.
[3] D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis," *Annu. Rev. Biomed. Eng.*, vol. 19, pp. 221–248, 2017.
[4] R. Ge, G. Yang, Y. Chen, L. Luo, C. Feng, H. Ma *et al.*, "K-Net: Integrate Left Ventricle Segmentation and Direct Quantification of Paired Echo Sequence," *IEEE Trans. Med. Imag.*
[5] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *CVPR*, 2015, pp. 3431–3440.
[6] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *MICCAI*. Springer, 2015, pp. 234–241.
[7] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3d u-net: learning dense volumetric segmentation from sparse annotation," in *MICCAI*. Springer, 2016, pp. 424–432.
[8] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *IC3DV*. IEEE, 2016, pp. 565–571.
[9] X. Li, H. Chen, X. Qi, Q. Dou, C.-W. Fu, and P.-A. Heng, "H-denseunet: hybrid densely connected unet for liver and tumor segmentation from ct volumes," *IEEE Trans. Med. Imag.*, vol. 37, no. 12, pp. 2663–2674, 2018.
[10] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: A nested u-net architecture for medical image segmentation," in *Int. Workshop Deep Learn. Med. Image Anal. Int. Workshop Multimodal Learn. Clin. Decis. Support*. Springer, 2018, pp. 3–11.
[11] J. Zhang, Y. Jin, J. Xu, X. Xu, and Y. Zhang, "Mdu-net: Multi-scale densely connected u-net for biomedical image segmentation," *arXiv preprint arXiv:1812.00352*, 2018.
[12] Q. Jin, Z. Meng, T. D. Pham, Q. Chen, L. Wei, and R. Su, "Dunet: A deformable network for retinal vessel segmentation," *Knowledge-Based Syst.*, vol. 178, pp. 149–162, 2019.
[13] Q. Jin, Z. Meng, C. Sun, L. Wei, and R. Su, "Ra-unet: A hybrid deep attention-aware network to extract liver and tumor in ct scans," *arXiv preprint arXiv:1811.01328*, 2018.
[14] J. Dolz, I. B. Ayed, and C. Desrosiers, "Dense multi-path u-net for ischemic stroke lesion segmentation in multiple image modalities," in *Int. MICCAI Brainlesion Workshop*. Springer, 2018, pp. 271–282.
[15] J. Guo, J. Deng, N. Xue, and S. Zafeiriou, "Stacked dense u-nets with dual transformers for robust face alignment," *arXiv preprint arXiv:1812.01936*, 2018.
[16] F. Isensee, J. Petersen, A. Klein, D. Zimmerer, P. F. Jaeger, S. Kohl, *et al.*, "nnu-net: Self-adapting framework for u-net-based medical image segmentation," *arXiv preprint arXiv:1809.10486*, 2018.
[17] A. Clèrigues, S. Valverde, J. Bernal, J. Freixenet, A. Oliver, and X. Lladó, "Sunet: a deep learning architecture for acute stroke lesion segmentation and outcome prediction in multimodal mri," *arXiv preprint arXiv:1810.13304*, 2018.
[18] J. Dolz, C. Desrosiers, and I. B. Ayed, "Ivd-net: Intervertebral disc localization and segmentation in mri with a multi-modal unet," in *Int. Workshop and Challenge on Comput. Methods and Clin. Appl. for Spine Imaging*. Springer, 2018, pp. 130–143.
[19] J. Zhuang, "Laddernet: Multi-path networks based on u-net for medical image segmentation," *arXiv preprint arXiv:1810.07810*, 2018.
[20] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, *et al.*, "Attention u-net: Learning where to look for the pancreas," *arXiv preprint arXiv:1804.03999*, 2018.
[21] C. Peng, X. Zhang, G. Yu, G. Luo, and J. Sun, "Large kernel matters–improve semantic segmentation by global convolutional network," in *CVPR*, 2017, pp. 4353–4361.
[22] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Semantic image segmentation with deep convolutional nets and fully connected crfs," *arXiv preprint arXiv:1412.7062*, 2014.
[23] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, 2017.
[24] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *ECCV*, 2018, pp. 801–818.
[25] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *CVPR*, 2017, pp. 1251–1258.

[26] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," *arXiv preprint arXiv:1706.05587*, 2017.

[27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *CVPR*, 2016, pp. 770–778.

[28] N. Ben-Zadok, T. Riklin-Raviv, and N. Kiryati, "Interactive level set segmentation for image-guided therapy," in *IEEE Int. Symp. Biomed. Imag:From Nano to Macro*. IEEE, 2009, pp. 1079–1082.

[29] H. Khotanlou, A. Fallahi, M. A. Oghabian, and M. Pooyan, "Segmentation of uterine fibroid on mr images based on chan–vese level set method and shape prior model," *Biomed. Eng.: Appl., Basis. and Commun.*, vol. 26, no. 02, pp. 1450030, 2014.

[30] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Trans. Image Processing*, vol. 10, no. 2, pp. 266–277, 2001.

[31] X. Bresson, P. Vandergheynst, and J.-P. Thiran, "A variational model for object segmentation using boundary information and shape prior driven by the mumford-shah functional," *IJCV*, vol. 68, no. 2, pp. 145–162, 2006.

[32] J. Yao, D. Chen, W. Lu, and A. Premkumar, "Uterine fibroid segmentation and volume measurement on mri," in *SPIE*, vol. 6143. International Society for Optics and Photonics, 2006, pp. 614322.

[33] A. Fallahi, M. Pooyan, H. Ghanaati, M. A. Oghabian, H. Khotanlou, M. Shakiba, *et al.*, "Uterine segmentation and volume measurement in uterine fibroid patients mri using fuzzy c-mean algorithm and morphological operations," *Iran. J. Radiol.*, vol. 8, no. 3, pp. 150, 2011.

[34] A. Fallahi, M. Pooyan, H. Khotanlou, H. Hashemi, K. Firouznia, and M. A. Oghabian, "Uterine fibroid segmentation on multiplan mri using fcm, mpfcm and morphological operations," in *in proc. IEEE 2 nd Comput. Engin. Tech.*, vol. 7. IEEE, 2010, pp. V7–1.

[35] L. Ma and R. C. Staunton, "A modified fuzzy c-means image segmentation algorithm for use with uneven illumination patterns," *Pattern recogn.*, vol. 40, no. 11, pp. 3005–3011, 2007.

[36] C. Militello, S. Vitabile, G. Russo, G. Candiano, C. Gagliardo, M. Midiri, *et al.*, "A semi-automatic multi-seed region-growing approach for uterine fibroids segmentation in mrgfus treatment," in *IEEE 7th Int. Conf. Complex, Intelligent, and Software Intensive Syst.* IEEE, 2013, pp. 176–182.

[37] L. Rundo, C. Militello, S. Vitabile, C. Casarino, G. Russo, M. Midiri, *et al.*, "Combining split-and-merge and multi-seed region growing algorithms for uterine fibroid segmentation in mrgfus treatments," *Med. Biol. Eng. Comput.*, vol. 54, no. 7, pp. 1071–1084, 2016.

[38] K. Antila, H. J. Nieminen, R. B. Sequeiros, and G. Ehnholm, "Automatic segmentation for detecting uterine fibroid regions treated with mr-guided high intensity focused ultrasound (mr-hifu)," *Med. Phys.*, vol. 41, no. 7, pp. 073502, 2014.

[39] C. Militello, S. Vitabile, L. Rundo, G. Russo, M. Midiri, and M. C. Gilardi, "A fully automatic 2d segmentation method for uterine fibroid in mrgfus treatment evaluation," *Comput. Biol. Med.*, vol. 62, pp. 277–292, 2015.

[40] L. Rundo, C. Militello, A. Tangherloni, G. Russo, R. Lagalla, G. Mauri, *et al.*, "Computer-assisted approaches for uterine fibroid segmentation in mrgfus treatments: Quantitative evaluation and clinical feasibility analysis," in *Italian Workshop on Neural Nets*. Springer, 2017, pp. 229–241.

[41] Y. Kurata, M. Nishio, K. Fujimoto, M. Yakami, A. Kido, H. Isoda,*et al.*, "Automatic segmentation of uterus with malignant tumor on MRI using U-net," *32nd IJCARS*, 2018.

[42] Y. Kurata, M. Nishio, K. Fujimoto, M. Yakami, A. Kido, H. Isoda, et. al., "Automatic segmentation of the uterus on MRI using a convolutional neural network," *Comput. Biol. Med.*,vol. 114, no. 13438, 2019.

[43] J. Liu, Y. Pan, M. Li, Z. Chen, L. Tang, C. Lu, *et al.*, "Applications of deep learning to mri images: A survey," *Big Data Mining and Analytics*, vol. 1, no. 1, pp. 1–18, 2018.

[44] S. M. Anwar, M. Majid, A. Qayyum, M. Awais, M. Alnowami, and M. K. Khan, "Medical image analysis using convolutional neural networks: a review," *J. Med. Syst.*, vol. 42, no. 11, pp. 226, 2018.

[45] F. Milletari, S.-A. Ahmadi, C. Kroll, A. Plate, V. Rozanski, J. Maiostre, *et al.*, "Hough-cnn: deep learning for segmentation of deep brain regions in mri and ultrasound," *Comput. Vis. Image. Und.*, vol. 164, pp. 92–102, 2017.

[46] S. Valverde, M. Cabezas, E. Roura, S. González-Villà, D. Pareto, J. C. Vilanova, *et al.*, "Improving automated multiple sclerosis lesion segmentation with a cascaded 3d convolutional neural network approach," *NeuroImage*, vol. 155, pp. 159–168, 2017.

[47] C. Wachinger, M. Reuter, and T. Klein, "Deepnat: Deep convolutional neural network for segmenting neuroanatomy," *NeuroImage*, vol. 170, pp. 434–445, 2018.

[48] S. Trebeschi, J. J. van Griethuysen, D. M. Lambregts, M. J. Lahaye, C. Parmar, F. C. Bakers, *et al.*, "Deep learning for fully-automated localization and segmentation of rectal cancer on multiparametric mr," *Sci. Rep.*, vol. 7, no. 1, pp. 5301, 2017.

[49] W. Zhang, R. Li, H. Deng, L. Wang, W. Lin, S. Ji, *et al.*, "Deep convolutional neural networks for multi-modality isointense infant brain image segmentation," *NeuroImage*, vol. 108, pp. 214–224, 2015.

[50] J. Bernal, K. Kushibar, D. S. Asfaw, S. Valverde, A. Oliver, R. Martí, *et al.*, "Deep convolutional neural networks for brain image analysis on magnetic resonance imaging: a review," *Arti. Intell. Med.*, vol. 95, pp. 64–81, 2019.

[51] M. Avendi, A. Kheradvar, and H. Jafarkhani, "A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac mri," *Med. Image Anal.*, vol. 30, pp. 108–119, 2016.

[52] Z. Gu, J. Cheng, H. Fu, K. Zhou, and J. Liu, "Ce-net: Context encoder network for 2d medical image segmentation," *IEEE Transactions on Medical Imaging*, vol. PP, no. 99, pp. 1–1, 2019.

[53] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *AAAI*, 2017.

[54] M. Holschneider, R. Kronland-Martinet, J. Morlet, and P. Tchamitchian, "A real-time algorithm for signal analysis with the help of the wavelet transform," *Wavelets*. Springer, 1990, pp. 286–297.

[55] J.-M. Combes, A. Grossmann, and P. Tchamitchian, "Wavelets: Time-Frequency Methods and Phase Space," Proceedings of the International Conference, Marseille, France, December 14–18, 1987,". Springer Science & Business Media, 2012.

[56] P. Wang, P. Chen, Y. Yuan, D. Liu, Z. Huang, X. Hou, *et al.*, "Understanding convolution for semantic segmentation," in *WACV*. IEEE, 2018, pp. 1451–1460.

[57] Z. Hussain, F. Gimenez, D. Yi, and D. Rubin, "Differential data augmentation techniques for medical imaging classification tasks," in *AMIA Annu. Symp.*, vol. 2017. American Medical Informatics Association, 2017, pp. 979.

[58] L. R. Dice, "Measures of the amount of ecologic association between species,"*Ecology*,vol. 26, no. 3, pp. 297–302, 1945.

[59] F. C. Monteiro, and A. C.Campilho, "Performance evaluation of image segmentation," in *ICIAR*. Springer, 2006, pp. 248-259.

[60] R. Trevethan, "Sensitivity, specificity, and predictive values: foundations, pliabilities, and pitfalls in research and practice," *Front. Public Health*, vol. 5, pp. 307, 2017.

[61] P. Jaccard, "The distribution of the flora in the alpine zone," *New Phytol.*, vol. 11, no. 2, pp. 37-50, 1912.

[62] J. Henrikson, "Completeness and total boundedness of the Hausdorff metric," *MURJ*, vol. 1, pp. 69-80, 1999.

[63] K. Sun, Y. Zhao, B. Jiang, T. Cheng, B. Xiao, D. Liu, *et al.*, "High-resolution representations for labeling pixels and regions," *arXiv preprint arXiv:1904.04514*, 2019.

[64] Y. Wang, Q. Zhou, J. Liu, J. Xiong, G. Gao, X. Wu, *et al.*, "Lednet: A lightweight encoder-decoder network for real-time semantic segmentation," *arXiv preprint arXiv:1905.02423*, 2019.

[65] H. Wu, J. Zhang, K. Huang, K. Liang, and Y. Yu, "Fastfcn: Rethinking dilated convolution in the backbone for semantic segmentation," *arXiv preprint arXiv:1903.11816*, 2019.

[66] Z. Tian, T. He, C. Shen, and Y. Yan, "Decoders matter for semantic segmentation: Data-dependent decoding enables flexible feature aggregation," in *CVPR*, 2019, pp. 3126–3135.

[67] Y. Chen, H. Fang, B. Xu, Z. Yan, Y. Kalantidis, M. Rohrbach, *et al.*, "Drop an octave: Reducing spatial redundancy in convolutional neural networks with octave convolution," *arXiv preprint arXiv:1904.05049*, 2019.

[68] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *CVPR*, 2016, pp. 2818–2826.