Attention 3D Dilated Multi-Fiber Network for Abdominal Organ Segmentation in CT

Jianping Zhang
Xiangtan University
Xiangtan City, Hunan Province, China
jpzhang@xtu.edu.cn

Abstract

Abdominal organ segmentation plays an important role in medical image processing. In this work, our goal is to segment abdominal CT volume. We apply attention block to DMFNet, and propose a new Attention DMFNet for medical imaging, which can automatically learn and focus target structures of different shapes and sizes. DMFNet is a highly efficient 3D CNN, which can realize real-time dense volume segmentation. It uses 3D multi fiber units composed of lightweight 3D convolution network to significantly reduce the computational cost. Models trained with attention blocks implicitly learn to suppress irrelevant regions in an input image while highlighting salient features useful for a specific task. Integrating attention blocks into DMFNet can improve model sensitivity and prediction accuracy with minimal computational overhead. Extensive experimental results on the MICCAI 2021 FLARE challenge dataset show that the proposed architecture greatly reduces computation cost while maintaining high accuracy for abdominal organ segmentation segmentation.

1. Introduction

There are two major challenges in this mission. The first challenge comes from the diversity of the dataset, which including multi-center, multi-phase, multi-vendor, and multi-disease cases. Medical image segmentation has not been solved today. An important reason is the complexity and diversity of medical images. Due to the differences in the imaging principle of medical image and the characteristics of tissue itself, the formation of image is affected by noise, field offset effect, local volume effect and tissue motion. Compared with ordinary image, medical image inevitably has the characteristics of fuzziness and non-uniformity. In addition, the anatomical structure and shape of human body are complex, and there are considerable differences between people. All these bring difficulties to medical image segmentation. Traditional segmentation techniques either fail completely or require some special processing techniques. Therefore, it is necessary to study image segmentation methods in the field of medical application.The second challenge comes from the efficiency requirement for the proposed solutions.

In order to solve the above difficulties, we use DMFNet as the solution, and add attention blocks to improve the segmentation accuracy. The reasons are as follows. First, DMFNet has less parametersh and it is fast. Second, attention blocks can identify salient image regions and prune feature responses to preserve only the activations relevant to the specific task. Third, it shows powerful performance on several segmentation tasks.

2. Method

DMFNet [1] extended the multi-fiber unit design with an adaptive weighted dilated convolution to capture the multi-scale features in brain MR images. However, when facing multiple targets with large volume difference at the same time, DMFNet can better complete the segmentation task of larger targets, but the prediction accuracy of smaller targets is low. Attention mechanism was originally used in machine translation [2], which can tend to pay attention to some information in the image to assist judgment, and ignore irrelevant information. Attention blocks [3] can be easily integrated into standard generation CNN networks. Inspired by this, we integrate the attention block into dmfnet to realize high-precision image segmentation with less parameters and faster speed. Figure 1 illustrates the applied Attention DMFNet (Attention mechanism is added to the original DMFNet structure diagram [1]).

2.1. Preprocessing

The baseline method includes the following preprocessing steps:

- Rotating, flipping, unifying window width and window level.
- Cropping strategy: Cut the original image into small pieces of $128 \times 128 \times 64$ for training.(Note: when the image cannot be completely cut to $128 \times 128 \times 64$, padding the missing part of the picture.)
- Resampling method for anisotropic data: None.
- Intensity normalization method: None.

2.2. Proposed Method

2.2.1 Dilated Multi-Fiber (DMF) Unit

Channel Grouping. The idea of channel grouping is to split the convolutional channels as multiple groups that can reduce the connections between the feature maps and kernels for parameter saving significantly. As examples shown in Figure 2 (a) and (b), the regular residual unit is grouped into g parallel residual units that are called fibers. We assume the kernel size is constant,e.g. kernel $= 3 \times 3 \times 3$ and denote $\text{param}_{(a)}$ and $\text{param}_{(b)}$ as the parameter amounts of Figure 2 (a) and (b), respectively. Thus, we have $\text{param}_{(a)} = \text{kernel} \times (c_{in} \times c_{\text{mid}} + c_{\text{mid}} \times c_{\text{out}})$, where c_* is the number of channel. With the strategy of multiple fibers grouping, the amount of parameter comes to $\text{param}_{(b)} = g \times \text{kernel} \times (c_{\text{in}}/g \times c_{\text{mid}}/g + c_{\text{mid}}/g \times c_{\text{out}}/g) = \text{param}_{(a)}/g$, which is g times less than $\text{param}_{(a)}$.

Multiplexer. To facilitate the information exchange between fibers, the $1\times\ 1 \times 1$ convolutions, dubbed as multiplexer, are utilized for information routing among different fibers. It is comprised of two $1 \times 1 \times 1$ convolution layers, as illustrated in Figure 2. And the input channel c_{in} is squeezed to $c_{in}/4$ and then inflated to c_{in}. By employing two $1 \times 1 \times 1$ convolutions ($\text{params} = c_{in} \times c_{in}/4 + c_{in}/4 \times c_{in} = c_{in}^2/2$), it can reduce half of the parameters as compared to using one $1 \times 1 \times 1$ convolution ($\text{params} = c_{in}^2$) . Besides, the residual shortcuts, which are placed outside the multiplexer and the entire unit, allow the information pass through from lower level to higher level directly, leading to enhanced learning capability without additional parameters.

Dilated Fiber. To enlarge the respective field and capture the multi-scale 3D spatial correlations of the brain tumor lesions, the dilated convolution is employed. As shown in Figure 2 (d), the dilated fiber is comprised of three 3D dilated convolution branches with the dilation rates of $d = 1, 2$ and 3 respectively. We allocate the learnable weights ω_1, ω_2 and ω_3 to each dilated branches, and then sum them up. This weighted sum strategy is conductive to select most valuable information automatically from different field of view. The weight coefficients are one-initialized, which means the branches contribute equally at the beginning of the training process.

2.2.2 Attention Block

Input features (x^l) are scaled with attention coefficients (α) computed in attention block(AB). Spatial regions are selected by analysing both the activations and contextual information provided by the gating signal (g) which is collected from a coarser scale. Grid resampling of attention coefficients is done using trilinear interpolation.Figure 3 illustrates the attention block.

Attention coefficients, $\alpha_i \in [0, 1]$, identify salient image regions and prune feature responses to preserve only the activations relevant to the specific task as shown in Figure 3a. The output of ABs is the element-wise multiplication of input feature-maps and attention coefficients: $\hat{x}_{i,c}^l = x_{i,c}^l \cdot \alpha_i^l$. In a default setting, a single scalar attention value is computed for each pixel vector $x_i^l \in \mathbb{R}^{Fl}$ where F_l corresponds to the number of feature-maps in layer l. As shown in Figure 3 , a gating vector $g_i \in \mathbb{R}^{F_g}$ is used for each pixel i to determine focus regions. The gating vector contains contextual information to prune lower-level feature responses. Using additive attention to obtain the gating coefficient. Additive attention is formulated as follows:

$$q_{att}^l = \psi^T \left(\sigma_1 \left(W_x^T x_i^l + W_g^T g_i + b_g\right)\right) + b_\psi$$
$$\alpha_i^l = \sigma_2 \left(q_{att}^l \left(x_i^l, g_i; \Theta_{att}\right)\right)$$

where $\sigma_2 \left(x_{i,c}\right) = \frac{1}{1+\exp(-x_{i,c})}$ correspond to sigmoid activation function. AB is characterised by a set of parameters Θ_{att} containing: linear transformations $W_x \in \mathbb{R}^{F_l \times F_{\text{int}}}, W_g \in \mathbb{R}^{F_g \times F_{\text{int}}}$, $\psi \in \mathbb{R}^{F_{int} \times 1}$ and bias terms $b_\psi \in \mathbb{R}, b_g \in \mathbb{R}^{F_{int}}$. The linear transformations are computed using channel-wise 1x1x 1 convolutions for the input tensors.

2.2.3 Other information

- Loss function: In general, the cross-entropy loss function is used. When the target volume is small using weighted cross entropy loss function, improve the weight of the small target, in order to

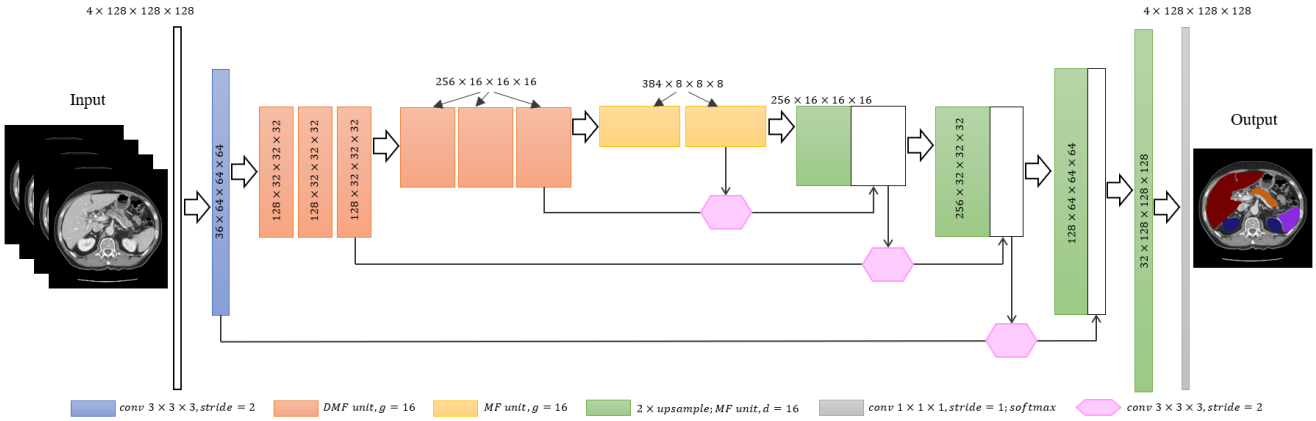

Figure 1. Attention DMFNet architecture. Where g is referred to the number of groups, e.g. $g = 16$ used in this work. We use 2-stride convolution to downsample the feature maps.

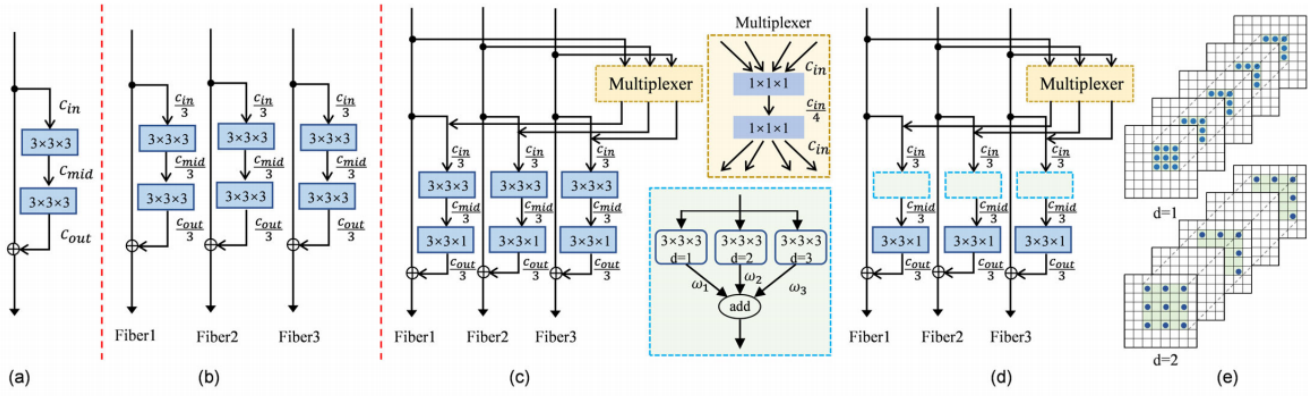

Figure 2. From [1]. (a) A residual unit with two regular convolution layers. (b) The multi-fiber design consisting of multiple separated residual units, called fibers. (c) The multi-fiber (MF) [1] unit takes advantage of a multiplexer for information routing. (d) The dilated multi-fiber (DMF) unit with an adaptive weighting scheme for different dilation rates. (e) The schematic diagram of the 3D dilated convolution operation. d is the dilation rate. $d = 1$ indicates the regular convolution.

Table 1. Data splits of FLARE2021.

Data Split	Center	Phase	# Num.
Training (361 cases)	The National Institutes of Health Clinical Center	portal venous phase	80
	Memorial Sloan Kettering Cancer Center	portal venous phase	281
Validation (50 cases)	Memorial Sloan Kettering Cancer Center	portal venous phase	5
	University of Minnesota	late arterial phase	25
	7 Medical Centers	various phases	20
Testing (100 cases)	Memorial Sloan Kettering Cancer Center	portal venous phase	5
	University of Minnesota	late arterial phase	25
	7 Medical Centers	various phases	20
	Nanjing University	various phases	50

improve the prediction precision of the small target.

- Number of model parameters: 4.724M.
- Number of flops: 483.469M .

2.3. Post-processing

Use the label and region props functions in the measure sub module in the skimage package to post process the data, to mark and sort multiple segmentation results.

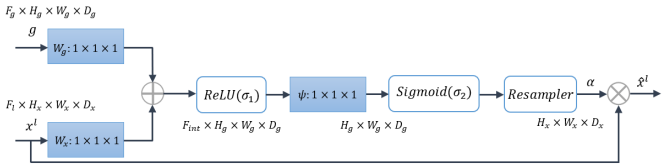

Figure 3. From [3]. Schematic of the proposed additive attention block.

3. Dataset and Evaluation Metrics

3.1. Dataset

- A short description of the dataset used:
  The dataset used of FLARE2021 is adapted from MSD [4] (Liver [5], Spleen, Pancreas), NIH Pancreas [6, 7, 8], KiTS [9, 10], and Nanjing University under the license permission. For more detail information of the dataset, please refer to the challenge website and [11].

- Details of training / validation / testing splits:
  The total number of cases is 511. An approximate 70%/10%/20% train/validation/testing split is employed resulting in 361 training cases, 50 validation cases, and 100 testing cases. The detail information is presented in Table 1.

3.2. Evaluation Metrics

- Dice Similarity Coefficient (DSC)

- Normalized Surface Distance (NSD)

- Running time

- Maximum used GPU memory (when the inference is stable)

4. Implementation Details

4.1. Environments and requirements

The environments and requirements of the baseline method is shown in Table 2.

4.2. Training protocols

The training protocols of the baseline method is shown in Table 3.

4.3. Testing protocols

Pre-processing steps of the network inputs and outputs are the same as the strategy is applied in training.

Table 2. Environments and requirements.

Windows/Ubuntu version	Ubuntu 18.04.5 LTS
CPU	Intel(R) Xeon(R) Silver 4114 CPU@2.20GHz
RAM	125×2GB; 2950MT/s
GPU	GEFORCE RTX 2080 Ti
CUDA version	11.4
Programming language	Python3.6
Deep learning framework	Pytorch (Torch 1.9.0, torchvision 0.4.0+cu92)
Specification of dependencies	Attention DMFNet
(Optional) code is publicly available at	DMFNet

Table 3. Training protocols.

Data augmentation methods	Rotations, scaling and mirroring.
Initialization of the network	"he" normal initialization
Patch sampling strategy	None.
Batch size	2
Patch size	128×128×64
Total epochs	300
Optimizer	Stochastic gradient descent with nesterov momentum ($\mu = 0.99$)
Initial learning rate	0.01
Learning rate decay schedule	learning rate policy: $0.01 \times 0.5^{[epoch/500]}$
Stopping criteria, and optimal model selection criteria	Stopping criterion is reaching the maximum number of epoch (300).
Training time	92.5 hours
CO_2eq[1]	9,927.3276g

Table 4. Quantitative results on validation set.

Organ	DSC (%)	NSD (%)
Liver	71.7±4.7	60.6±5.2
Kidney	44.2±12.5	45.3±20.1
Spleen	53.6±15.1	55.0±15.0
Pancreas	9.7±3.1	11.6±4.2

5. Results

5.1. Quantitative results for 5-fold cross validation.

5.2. Quantitative results on validation set.

Table 4 illustrates the results on validation cases. Figure 4 is the corresponding violin plots of the organ segmentation performance. For DSC, the DSC values of liver,kidney and spleen are obviously higher than the values of pancreas. And the NSD values of liver,kidney and spleen are also higher than the values of pancreas. That is to say, Whether based on region or boundary, the segmentation performance of liver, kidney and spleen is better than that of pancreas. This situation is probably caused by the high difficulty of segmentation of small objects with irregular edges. The segmentation ability of the model for such objects needs to be improved.And the high dispersed distributions from the violin plots of all organs indicate that the robustness of the model is poor, which need further improvements.

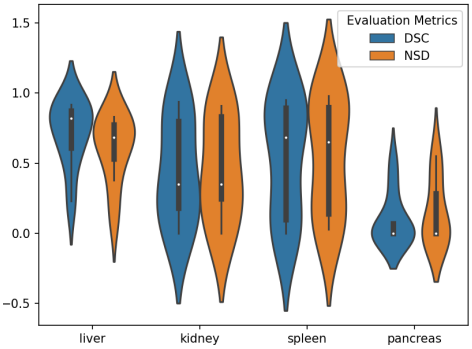

Figure 4. Violin plots of the organ segmentation results (DSC and NSD) on validation set.

5.3. Qualitative results

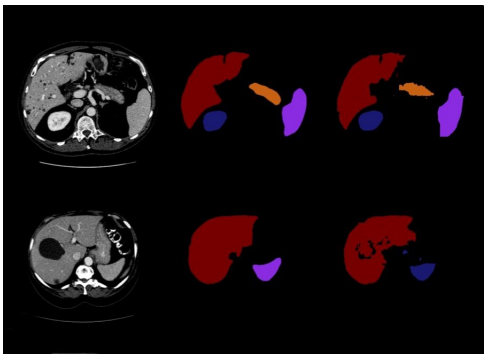

Figure 5. Challenging examples.First column is the image, second column is the ground truth, and third column is the predicted results by our method.

Figure 5 presents two challenging examples. It can be found that the method we proposed can segment the lesion affected organs in some extent. The first row of Figure 5 shows an example with liver (red) tumor which bring correct segmentation.Maybe it is because that this type of tumor has small shape , which has little impact on segmentation. Second row of Figure 5 illustrates a fatty liver case where some of the liver is darker than healthy ones. In this case, our method fails to segment the liver (red) completely, and makes wrong segmentation for the spleen.

6. Discussion and Conclusion

The method we proposed can work well on cases where no diseases exist and some cases where the lesion area is very small . Besides, the DSC and NSD scores of liver segmentation,spleen segmentation and kidney segmentation are higher than pancreas segmentation , indicating these three organs maybe comparable easier tasks as a result of their bigger size and consistent shape. Bad performance is obtained for pancreas segmentation as a result of the interpatient anatomical variability of volume and shape.

Obviously, the presence of a large number of lesion areas in organs is the main factor affecting the segmentation effect. How to segment those cases precisely is important. Besides, obtaining an accurate boundary segmentation need further investigate. Moreover, the pancreas's inter-patient anatomical variability of volume and shape make the segmentation of pancreas a problem.

Acknowledgment

The authors of this paper declare that the segmentation method they implemented for participation in the FLARE challenge has not used any pre-trained models nor additional datasets other than those provided by the organizers.

References

[1] C. Chen, X. Liu, M. Ding, J. Zheng, and J. Li, "3d dilated multi-fiber network for real-time brain tumor segmentation in mri," Medical Image Computing and Computer Assisted Intervention – MICCAI 2019. MICCAI 2019. Lecture Notes in Computer Science, vol. 11766, 2019. 1, 3

[2] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," arXiv preprint arXiv:1409.0473, 2014. 1

[3] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz et al., "Attention u-net: Learn-

ing where to look for the pancreas," arXiv preprint arXiv:1804.03999, 2018. 1, 4

[4] A. L. Simpson, M. Antonelli, S. Bakas, M. Bilello, K. Farahani, B. Van Ginneken, A. Kopp-Schneider, B. A. Landman, G. Litjens, B. Menze et al., "A large annotated medical image dataset for the development and evaluation of segmentation algorithms," arXiv preprint arXiv:1902.09063, 2019. 4

[5] P. Bilic, P. F. Christ, E. Vorontsov, G. Chlebus, H. Chen, Q. Dou, C.-W. Fu, X. Han, P.-A. Heng, J. Hesser et al., "The liver tumor segmentation benchmark (lits)," arXiv preprint arXiv:1901.04056, 2019. 4

[6] H. Roth, A. Farag, E. Turkbey, L. Lu, J. Liu, and R. Summers, "Data from pancreas-ct. the cancer imaging archive (2016)." 4

[7] H. R. Roth, L. Lu, A. Farag, H.-C. Shin, J. Liu, E. B. Turkbey, and R. M. Summers, "Deeporgan: Multi-level deep convolutional networks for automated pancreas segmentation," in International conference on medical image computing and computer-assisted intervention. Springer, 2015, pp. 556–564. 4

[8] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle et al., "The cancer imaging archive (tcia): maintaining and operating a public information repository," Journal of digital imaging, vol. 26, no. 6, pp. 1045–1057, 2013. 4

[9] N. Heller, F. Isensee, K. H. Maier-Hein, X. Hou, C. Xie, F. Li, Y. Nan, G. Mu, Z. Lin, M. Han et al., "The state of the art in kidney and kidney tumor segmentation in contrast-enhanced ct imaging: Results of the kits19 challenge," Medical Image Analysis, vol. 67, p. 101821, 2021. 4

[10] N. Heller, S. McSweeney, M. T. Peterson, S. Peterson, J. Rickman, B. Stai, R. Tejpaul, M. Oestreich, P. Blake, J. Rosenberg et al., "An international challenge to use artificial intelligence to define the state-of-the-art in kidney and kidney tumor segmentation in ct imaging." American Society of Clinical Oncology, vol. 38, no. 6, pp. 626–626, 2020. 4

[11] J. Ma, Y. Zhang, S. Gu, C. Zhu, C. Ge, Y. Zhang, X. An, C. Wang, Q. Wang, X. Liu, S. Cao, Q. Zhang, S. Liu, Y. Wang, Y. Li, J. He, and X. Yang, "Abdomenct-1k: Is abdominal organ segmentation a solved problem?" IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021. 4