MULTI-ENCODER PARSE-DECODER NETWORK FOR SEQUENTIAL MEDICAL IMAGE SEGMENTATION

Paper ID: 2018

Dachuan Shi[1] Ruiyang Liu[1] Linmi Tao[1] Zuoxiang He[2] Li Huo[3]

[1]Department of Computer Science and Technology, Tsinghua University, Beijing, China

[2]Beijing Tsinghua Changgung Hospital, School of Clinical Medicine, Tsinghua University, China

[3]Nuclear Medicine Department, Peking Union Medical College Hospital, Beijing, China

ACKNOWLEDGEMENT

This work is supported by the National Science Foundation of China at the Project 61672017.

INTRODUCTION

Deep learning models, especially U-Net and its derivate models, have been widely used in medical image segmentation. These approaches have achieved promising results in many medical image segmentation tasks with a limited number of training samples.

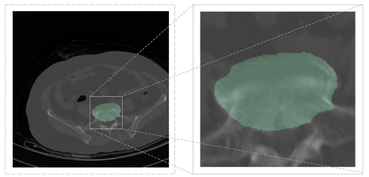

Figure 1. Medical image segmentation is aimed to separate organs or lesions from different kinds of medical images.

In general, medical images are captured sequentially with the help of medical instruments such as Computer Tomography and Magnetic Resonance Imaging. The spatial continuity information contained between adjacent medical image frames can be used to improve the segmentation accuracy of the intermediate frame.

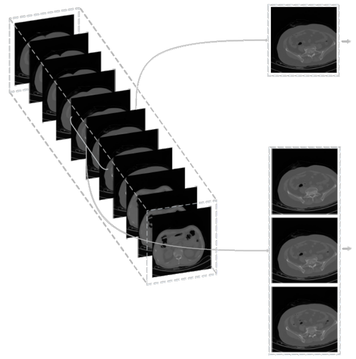

Figure 2. Sequentially captured medical images.

However, most existing 2D medical image segmentation models, such as U-Net and Attention U-Net fail to exploit the spatial continuity information. Some 3D medical image segmentation models, such as 3D U-Net and V-Net, can implicitly exploit continuity information, but the huge computational overhead of the 3D convolution kernel limits the practical application of these models. In this manuscript, we propose the multi-encoder parse-decoder network (MEPDNet) for sequential medical image segmentation, which can fully utilize the spatial continuity information contained between adjacent medical image frames without using resource-consumed 3D convolution kernels.

METHODOLOGY

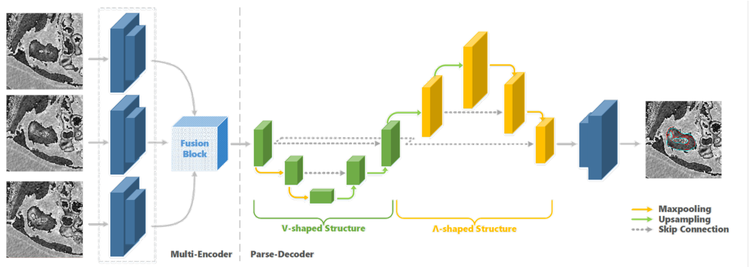

Figure 3. Overall architecture. MEPDNet contains a Multi-Encoder module and a Parse-Decoder module. The Multi-Encoder contains a parameter-shared downsampling encoder and a fusion block, while the Parse-Decoder contains a V-block and an up-sampling decoder.

METHODOLOGY Cont.

As illustrated in the figure 3, Sequential images are input into parameter-shared encoders for getting feature maps, which are then fused by a fusion block. A VΛ -Block is structured to parse and reconstruct the fused feature map. The output is fed into a decoder for generating segmentation masks.

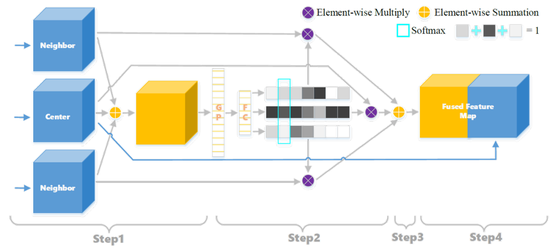

Figure 4. Diagram of fusion block.

As illustrated in the figure 4, The fusion block is designed to fuse feature maps:

1. The fusion block takes multi-path inputs and integrates them with an element-wise summation.
2. A global average pooling layer and a fully connected layer are employed to calculate weights for feature maps. And a softmax operation is then applied on the channel-wise to produce weights for different downsampling paths.
3. The integrated feature map is calculated by weighted summation of the input feature maps.
4. The output fused feature map is generated by concatenating the integrated feature map and the feature map from the center path of the inputs.

As illustrated in the figure 3, the VΛ Block is constituted by a V-shaped structure and a Λ-shaped structure. The V-shaped structure is similar to the U-Net. In the Λ-shaped structure, feature map output from the V-shaped structure is upsampled beyond the origin map size to expand the information representation with a larger scale. Benefited from this structure, subsequent convolution operations are conducted on expanded high-resolution feature maps, and thus provide more detailed local context information. There are also skip connections transport feature maps between the same stages of the Λ-shaped structure.

EXPERIMENTS

To verify the effectiveness of the proposed MEPDNet, we conduct experiments on three different medical image segmentation datasets. These include:

1. A new proposed Lumbar-CT Dataset.
2. A subset of MSD-Colon Cancer Dataset.
3. A subset of Pancreas-CT Dataset.

Method	MSD-Colon Cancer			Pancreas-CT			Lumbar-CT			Parms
	Dice	Recall	Precision	Dice	Recall	Precision	Dice	Recall	Precision	
SegNet [14]	29.14	44.89	26.33	49.81	57.93	80.79	92.23	92.01	93.31	112.32M
Deeplabv3+ [16]	38.01	43.36	41.87	62.29	66.65	79.68	94.55	93.91	96.36	208.67M
U-Net [3]	40.48	55.14	38.64	65.76	64.09	**90.54**	95.32	94.72	96.29	65.87M
Attention U-Net [4]	39.16	52.29	38.52	70.80	74.52	77.60	95.94	95.43	96.70	133.05M
R2U-Net [6]	45.40	56.31	43.25	68.45	70.20	87.31	95.90	94.21	97.86	508.57M
Attention R2U-Net [6]	48.42	56.91	**49.52**	71.70	**75.15**	87.57	96.11	94.55	**97.92**	509.91M
ScSE U-Net [5]	45.46	62.15	41.79	58.85	64.52	83.99	95.85	94.98	96.93	33.16M
CE-Net [11]	47.57	**68.46**	43.12	66.58	67.43	86.68	95.99	**96.78**	95.42	148.66M
UNet++ [7]	42.57	51.51	42.56	58.28	60.62	84.66	95.72	95.57	96.00	139.73M
MEPDNet (ours)	**52.99**	65.33	49.16	**75.24**	74.63	88.66	**96.39**	95.83	97.09	**19.37M**

Table 1. Experimental results of MEPDNet and comparison against other deep learning-based segmentation models.

Experiments on three datasets show MEPDNet outperforms other segmentation models.

EXPERIMENTS Cont.

We conduct ablation studies on the Lumbar-CT Dataset. The results prove the effectiveness of the proposed VΛ -Block architecture.

Block	No VΛ	Λ	V	ΛV	VΛ	No fusion
Dice	94.57	95.55	95.15	96.06	**96.39**	95.92

Table 2. Ablation studies

We compare our model with transfer learning and multitask learning that can also introduce spatial information into the training. The experimental results illustrate that by MEPDNet direct introduction of spatial continuity information is superior to that of introducing coarse spatial continuity information through multi-task learning and transfer learning.

Models	Top	Middle	Bottom	Mean
U-Net (Baseline)	95.93	94.73	94.56	95.07
Transfer U-Net	96.41	94.57	94.62	95.20
Multi U-Net (1:1:1)	96.16	95.27	93.81	95.08
Multi U-Net (1:8:10)	95.69	95.96	95.34	95.66
MEPDNet	**96.59**	**96.54**	**95.45**	**96.19**

Table 2. Comparison against transfer learning and multi-task learning.

The Figure 5 and Figure 6 demonstrates the visual comparison for qualitative analysis. The visualization figure indicates that spatial continuity information allows MEPDNet to segment more accurately.

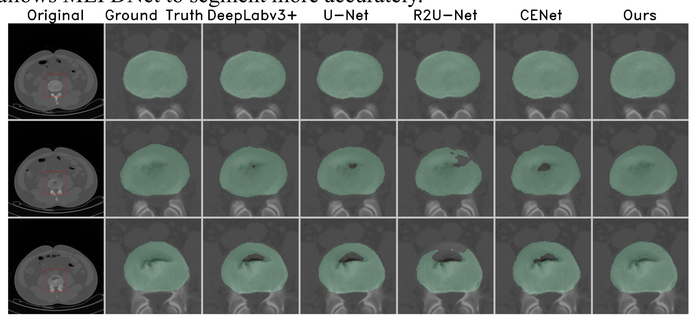

Figure 5. Visualization of segmentation results on Lumbar-CT Dataset. Segmentation results of three sequential slices with spatial continuity are displayed from top to bottom. The dark black area in the target lumbar results in an extra prediction curve in the middle slice given by other models, while MEPDNet is not affected by this due to spatial continuity information provided by the previous and next slices.

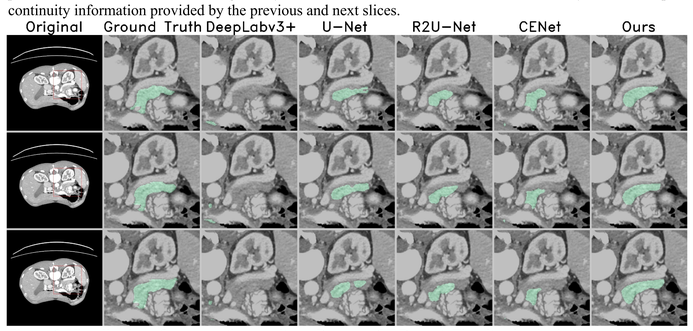

Figure 6. Visualization of segmentation results on Pancreas-CT Dataset. The situation is similar to the Figure 5.

CONCLUSION

We propose the MEPDNet for **sequential medical image segmentation**, which can fully **utilize the spatial continuity information** contained between adjacent frames without using resource-consumed 3D convolution kernels. Experimental results on several datasets show that the MEPDNet model **outperforms other segmentation models**. It is also proved that MEPDNet which directly introduces **continuous spatial constraints is superior to** multi-task and transfer learning methods that introduce **coarse spatial continuity information**.