CEUSEGNET: A CROSS-MODALITY LESION SEGMENTATION NETWORK FOR CONTRAST-ENHANCED ULTRASOUND

Zheling Meng[1,2†], *Yangyang Zhu*[3,4†], *Xiao Fan*[3,4], *Jie Tian*[1,6⋆], *Fang Nie*[3,4,5⋆], *Kun Wang*[1,2⋆]

[1]Institute of Automation Chinese Academy of Sciences (IACAS), CAS Key Laboratory of Molecular Imaging (⋆Email: kun.wang@ia.ac.cn)
[2]University of Chinese Academy of Sciences, School of Artificial Intelligence
[3]Lanzhou University Second Hospital, Department of Ultrasound (⋆Email: ery_nief@lzu.edu.cn)
[4]Gansu Province Ultrasonic Imaging Clinical Medical Research Center
[5]Gansu Province Medical Engineering Research Center for Intelligence Ultrasound
[6]Beihang University, Beijing Advanced Innovation Center for Big Data-Based Precision Medicine (⋆Email: jie.tian@ia.ac.cn)

ABSTRACT

Contrast-enhanced ultrasound (CEUS) is an effective imaging tool to analyze spatial-temporal characteristics of lesions and diagnose or predict diseases. However, delineating lesions frame by frame is a time-consuming work, which brings challenges to analyzing CEUS videos with deep learning technology. In this paper, we proposed a novel U-net-like network with dual top-down branches and residual connections, named CEUSegNet. CEUSegNet takes US and CEUS part of a dual-amplitude CEUS image as inputs. Cross-modality Segmentation Attention (CSA) and Cross-modality Feature Fusion (CFF) are designed to fuse US and CEUS features on multiple scales. Through our method, lesion position can be determined exactly under the guidance of US and then the region of interest can be delineated in CEUS image. Results show CEUSegNet can achieve a comparable performance with clinicians on metastasis cervical lymph nodes and breast lesion dataset.

***Index Terms*—** Contrast-enhanced ultrasound, cross-modality, lesion segmentation

1. INTRODUCTION

Contrast-enhanced ultrasound (CEUS) provides an economical and non-invasive tool for clinicians to visualize the dynamic enhancement process of tissue blood-flow distribution and perfusion over time[1]. In recent years, more and more researchers tend to use deep learning methods to extract spatial-temporal features in CEUS sequences to diagnose and predict diseases of different organs, such as liver[2], cervical lymph node[3] and breast[4]. However, it is a time and energy-consuming work that delineating the region of interest (RoI) containing lesion and surrounding microvasculature frame by frame. On the one hand, clinicians need to observe the ultrasound (US) part of a CEUS frame image at the same time to determine the lesion position (seen in Figure 1). On the other hand, they also need to observe the brightness change back and forth to determine RoI. Therefore, realizing the automatic lesion segmentation of CEUS is a key part of automatic CEUS analysis.

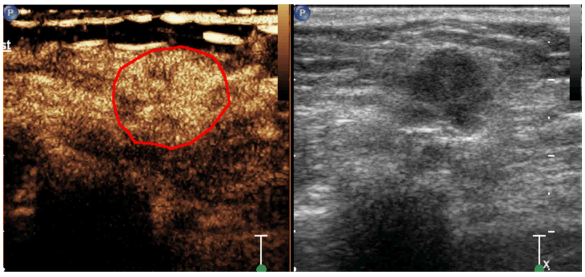


Fig. 1. An example of dual-amplitude CEUS image of breast lesion. Left: CEUS part; right: US part. The red circle outlines lesion region in CEUS part.

An intuitive idea is to imitate the clinician's method to segment the lesion. Wan et al[5] proposed CEUS-Net which learns spatial-temporal features and re-weights them. However, as mentioned above, a CEUS frame contains CEUS part and US part, which constitute a dual-amplitude image but US part is abandoned by CEUS-Net. Actually, US part can provide rough location of lesion and RoI can be determined finally in CEUS part. In addition, the lesion area in CEUS part is usually larger than in US due to the need of including surrounding microvasculature, which can be learned via annotated data by model. We think in this way, it is possible to do segmentation on a single-frame CEUS image with the help

[†]Co-authors.
[⋆]Corresponding authors.

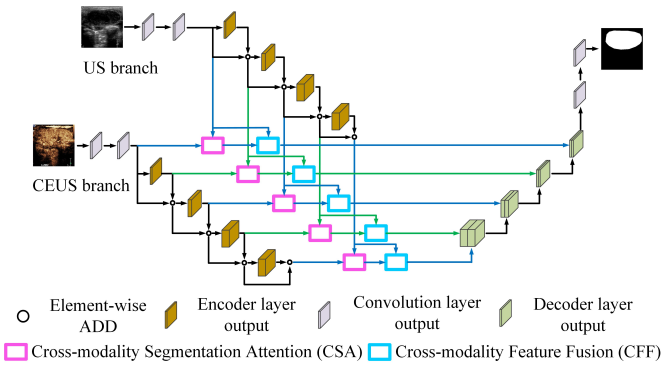


Fig. 2. The structure of our proposed CEUSegNet.

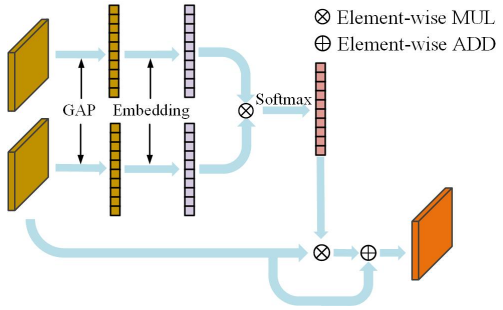


Fig. 3. Cross-modality Segmentation Attention (CSA). GAP: Global Average Pooling.

of dual-modality to reduce the time and space complexity but tedious for humans while the performance of segmentation on CEUS can be further improved.

Based on the above discussion, we proposed a cross-modality lesion segmentation network for contrast-enhanced ultrasound, named CEUSegNet and Figure 2 shows the structure. CEUSegNet is a U-net-like network but has dual top-down branches corresponding to CEUS and US part of a CEUS image respectively. Cross-modality Segmentation Attention (CSA) and Cross-modality Feature Fusion (CFF) are placed in each skip connection. CSA can highlight those important features corresponding to lesion segmentation in CEUS branch, and CFF can perform feature fusion on multiple scales. Under the guidance of US part, CEUSegNet can achieve single-frame lesion segmentation, which decreases the time and space complexity and provides an implicit solution to the CEUS spatial-temporal feature alignment problem. Results show that CEUSegNet achieves 91.05% dice, 80.06% mIoU on the cervical lymph node dataset and 89.97% dice, 75.62% mIoU on breast lesion dataset.

2. METHODS

2.1. CEUSegNet with dual top-down branches

Figure 2 shows the structure of CEUSegNet. Specifically, we followed U-net[6] but modified it into a structure with dual top-down branches. Given a dual-amplitude CEUS frame I, its CEUS part I^{ceus} and US part I^{us} are feed into the branches respectively. Then, Cross-modality Segmentation Attention (CSA) and Cross-modality Feature Fusion (CFF) module are proposed for dual modalities feature fusion. CSA can generate a set of weights under the guidance of US branch to highlight those significant features of CEUS branch while CFF can fuse US features and re-weighted CEUS features on multi scales. Fused dual-modality features are concatenated with higher semantic features and a lesion segmentation score map M is obtained after four decoders and an output convolution layer. The introduction of the dual branches structure allows us to perform lesion segmentation on any frame of any period in a CEUS video.

2.2. Cross-modality Segmentation Attention

Let F^{ceus} and F^{us} denote the output features of CEUS and US branch at a certain layer. We apply a global average pooling to F^{ceus} and F^{us} so that two vectors $v^{ceus} \in R^c$ and $v^{us} \in R^c$ are obtained, where c is the number of channels of F^{ceus} and F^{us}. Two learnable matrix $W^{ceus} \in R^{c\times c}$ and $W^{us} \in R^{c\times c}$ are introduced to embed v^{ceus} and v^{us} into the same Euclidean space. Then the two embedded vectors are multiplied and the attention weight w is obtained through Softmax operation. The above process can be formulated as:

$$w = Softmax(W^{ceus}GAP(F^{ceus}) \odot W^{us}GAP(F^{us})) \tag{1}$$

where $\odot$ is element-wise multiplication. Multiply F^{ceus} with w channel by channel and re-weighted $F^{ceus'}$ can be obtained. Experiments show that residual connection is useful to the improvement of performance so $F^{ceus'}$ is (as shown in Figure 3):

$$F^{ceus'} = F^{ceus} \oplus (w \otimes F^{ceus}) \tag{2}$$

where $\oplus$ denotes element-wise addition and $\otimes$ channel-wise multiplication.

CSA is used to stress those significant features and avoid irrelevant to prevent harm to the whole performance. Global average aggregation operation reduces the dimension of input features and generates representation vectors. The embedding matrixes embed the representation vectors from different modalities into the same Euclidean space and makes them comparable. As a consequence, CEUSegNet can pay more attention to the more important CEUS branch features.

2.3. Cross-modality Feature Fusion

Figure 4 shows the structure of Cross-modality Feature Fusion (CFF). In a CFF module, features F^{us} and re-weighted

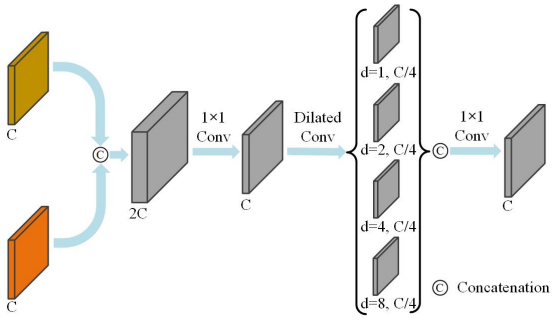


Fig. 4. Cross-modality Feature Fusion (CFF).

$F^{ceus'}$ are concatenated into $F \in R^{2c\times w\times h}$ and c, w and h denote the channel number, width and height of F^{us} and $F^{ceus'}$. Then a 1×1 convolution is applied on F to fuse features in channel dimension. We notice that there are lesions that have an area of different sizes. Therefore, inspired by [7], we use four parallel dilated convolutions to extract fused features on different scales. The number of channels on each scale is $c/4$. Then features on each scale are concatenated and a 1×1 convolution is applied again in order to decrease the gap between different scales and different modalities further. Since we have taken this module on each skip connection, we needn't repeat it on the up-sampling path. In a word, CFF plays a role in making up for the semantic gap caused by skip connection and the modality gap caused by multi-modality inputs.

3. EXPERIMENTS, RESULTS AND DISCUSSIONS

We collected CEUS videos of 199 patients with metastasis cervical lymph nodes and 146 patients with breast lesion from the local hospital. This study was approved by the institutional review board of our institution and all patients signed an informed consent form. Each patient was examined by ultrasound instrument Philips IU 22, and a CEUS clip with both wash-in and wash-out periods was obtained. Both CEUS and US part with a resolution of 375×375 are included in a frame. Then we first extract key frames from each video. Specifically, each video is first down-sampled at a sampling rate of 1 fps into a sequence of images; then select k frames with the largest grayscale change of the adjacent frames in the sampling time as the key frame ($k = 40$ here). We randomly select two key frames of each patient, and repeat it until the grayscales of the datasets are uniformly distributed. Therefore, the metastasis cervical lymph nodes (MCLN) dataset contains 398 dual-amplitude images and the breast lesion (BL) dataset contains 292.

The RoI of CEUS part is delineated by a clinician with more than 5 years experience and verified by another with 20 years. All delineated results were consensus results. Via double verification, the label conforms to clinicians' general understanding of CEUS lesion RoI, so that CEUSegNet can imitate their definition of RoI exactly by learning from the annonated data.

Method	MCLN Dataset		BL Dataset	
	Dice(%)	IoU(%)	Dice(%)	IoU(%)
U-net	74.85±9.47	58.42±10.95	69.27±15.06	53.16±14.64
CEUSegNet	**91.05**±4.06	**80.06**±6.65	**89.97**±5.01	**75.62**±7.85

Table 1. The performance of U-net and our proposed CEUSegNet on metastasis cervical lymph node (MCLN) and breast lesion (BL) dataset.

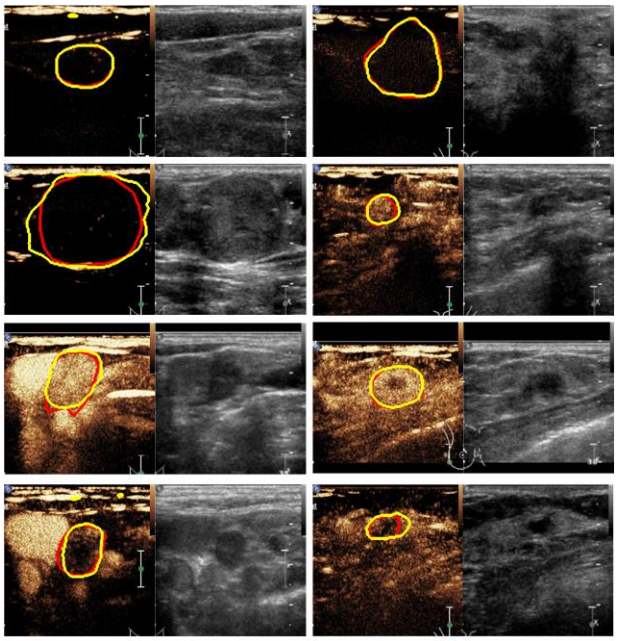
Input Modality		Network Module		Evaluation Index	
CEUS	US	CSA	CFF	Dice(%)	IoU(%)
✓				78.04 (+0.00)	61.66 (+0.00)
	✓			86.20 (+8.16)	73.15 (+11.49)
✓	✓			87.32 (+9.28)	74.53 (+12.87)
✓	✓	✓		86.64 (+8.60)	73.38 (+11.72)
✓	✓		✓	89.76 (+11.72)	78.03 (+16.37)
✓	✓	✓	✓	**91.05** (+13.01)	**80.06** (+18.40)

Table 2. The ablation experiment results of CEUSegNet on MCLN dataset.

We divide the datasets into training set and testing set in a ratio of 2:1, respectively. Under PyTorch framework, we train CEUSegNet with a batch size of 16 on GPU. The learning rate is set to 0.001 and the number of training epoch is 150. Adam optimizer is used to optimize the network. The input of CEUS branch and US branch is resized to 128×128. 0.5 is the threshold to assign a pixel to foreground or background. IoU and dice index are metrics of the performance of methods.

3.1. Evaluation and comparative experiment

Table 1 shows the results of CEUSegNet on CMLN and BC dataset. For comparison, we adopt U-net[6] to experiment under the same conditions, where U-net is modified into a structure with two top-down branches and a 1×1 convolution layer on each skip connection but without residual connection between down-sampling layers. On CMLN dataset, CEUSegNet achieved 91.05% dice and 80.06% IoU, while U-net only achieved 74.85% and 58.42% respectively. On BL dataset, CEUSegNet achieved 89.97% dice and 75.62% IoU, while U-net only achieved 69.27% and 53.16% respectively. Figure 5 shows some examples of segmentation results on two datasets. As can be seen, the segmentation results of CEUSegNet (yellow) are highly consistent with annotated RoI (red). Both the lesion and the surrounding blood vessel area are well segmented together. This demonstrates that through the annotated data, CEUSegNet learns to determine the position of lesion area based on US image and then determines the exact boundary of RoI based on CEUS image to cover both the lesion and peripheral blood vessel area.



(a) MCLN Dataset (b) BL Dataset

Fig. 5. Examples of segmentation results on metastasis cervical lymph nodes (MCLN) and breast lesion (BL) dataset at different periods. The red circle outlines the RoI annotated by the clinician and the yellow circle is segmentation results given by CEUSegNet. Each row represents a case.

3.2. Ablation experiment

In this subsection, we use US branch or CEUS branch respectively only, and both branches at the same time (using CSA, CFF or not respectively) to verify the performance of each part of CEUSegNet on CMLN dataset. Table 2 shows the results of the ablation experiment. On the one hand, compared with CEUSegNet, IoU with only CEUS or US branch achieved 61.66% and 73.15% only, and IoU with both branches increased by 12.87% but still less than CEUSegNet. This shows that only single-frame CEUS image can hardly complete the segmentation task (that is, time scale information is required), and the segmentation effect achieved by only US input is poor. On the other hand, we find that using CSA alone can harm the performance lightly. Independently, CFF is more important than CSA but when CSA is used with CFF, the performance is improved by 2.03% and 5.53% in total, indicating that through cross-modality attention mechanism, CFF can better integrate CEUS and US features. Also, we can see that residual connections between encoder layers makes the model easier to train and get a better result.

3.3. Some discussion about single-frame segmentation

CEUSegNet does lesion segmentation on only single frame although CEUS is a dynamic video modality. We choose this method based on the following considerations. First, as mentioned in Section 1, US part of a CEUS frame can help to locate the lesion, providing a good benchmark for segmentation (as seen in Table 2). Second, the size of the dataset can be increased by selecting multiple frames from each video and as a result, it can explore the characteristics of RoI delineation from the massive data. Besides, CEUSegNet only has 9.28M parameters and 12.74G MACs, and the inference time is 20.82($\pm$2.62) ms in our experiment setup, indicating that it is expected to be used for real-time segmentation on CEUS equipment.

4. CONCLUSION

In this paper, CEUSegNet is proposed to use only one frame dual-amplitude CEUS image to segment lesion area on CEUS image. Specifically, we use U-net-like structure but with dual-modality top-down branches and residual connections between down-sampling layers. Cross-modality Segmentation Attention (CSA) and Cross-modality Feature Fusion (CFF) module are also proposed to fuse features better. Experiments shows CEUSegNet can achieve a segmentation performance comparable to clinicians on cervical metastasis lymph node and breast cancer dataset.

5. ACKNOWLEDGEMENT

This work was supported by the Ministry of Science and Technology of China (Grant Nos. 2017YFA0205200), the National Natural Science Foundation of China (Grants Nos. 82027803, 62027901, 81930053, 81227901), the Chinese Academy of Sciences (Grants Nos. YJKYYQ20180048 and QYZDJ-SSW-JSC005) and Gansu Province Science and Technology Plan Project (Grants Nos. 21YF5FA122, 20JR10FA664).

6. REFERENCES

[1] Yong Eun Chung and Ki Whang Kim, "Contrast-enhanced ultrasonography: advance and current status in abdominal imaging," *Ultrasonography*, vol. 34, no. 1, pp. 3, 2015.

[2] Dan Liu, Fei Liu, Xiaoyan Xie, Liya Su, Ming Liu, Xiaohua Xie, Ming Kuang, Guangliang Huang, Yuqi Wang, Hui Zhou, et al., "Accurate prediction of responses to transarterial chemoembolization for patients with hepatocellular carcinoma by using artificial intelligence in contrast-enhanced ultrasound," *European radiology*, vol. 30, no. 4, pp. 2365–2376, 2020.

[3] Qi Zhang, Yue Liu, Hong Han, Jun Shi, and Wenping Wang, "Artificial intelligence based diagnosis for cervical lymph node malignancy using the point-wise gated boltzmann machine," *IEEE Access*, vol. 6, pp. 60605–60612, 2018.

[4] Chen Chen, Yong Wang, Jianwei Niu, Xuefeng Liu, Qingfeng Li, and Xuantong Gong, "Domain knowledge powered deep learning for breast cancer diagnosis based on contrast-enhanced ultrasound videos," *IEEE Transactions on Medical Imaging*, 2021.

[5] Peng Wan, Fang Chen, Xiaowei Zhu, Chunrui Liu, Yidan Zhang, Wentao Kong, and Daoqiang Zhang, "Ceus-net: Lesion segmentation in dynamic contrast-enhanced ultrasound with feature-reweighted attention mechanism," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2020, pp. 1816–1819.

[6] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[7] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam, "Rethinking atrous convolution for semantic image segmentation," *arXiv preprint arXiv:1706.05587*, 2017.