Biopsy Needle Segmentation using Deep Networks on inhomogeneous Ultrasound Images*

Yue Zhao, *Member, IEEE*, Yi Lu, Xin Lu, Jing Jin, Lin Tao, and Xi Chen

***Abstract*— In minimally invasive interventional surgery, ultrasound imaging is usually used to provide real-time feedback in order to obtain the best diagnostic results or realize treatment plans, so how to accurately obtain the position of the medical biopsy needle is a problem worthy of study. 2D ultrasound simulation images containing the medical biopsy needle are generated, and our images background is from the real breast ultrasound image. Based on the deep learning network, the images containing the medical biopsy needle are used to analyze the effectiveness of different networks for needle localization for the purpose of returning needle positions in non-uniform ultrasound images. The results show that attention U-Net performed best and can accurately reflect the real position of the medical biopsy needle. The IoU and Precision can reach 90.19% and 96.25%, and the Angular Error is 0.40 °.**

***Clinical Relevance*— Based on the deep network, for 2D ultrasound images containing medical biopsy needle, the localization precision can reach 96.25% and the Angular Error is 0.40 °.**

I. INTRODUCTION

2D ultrasound is usually used to guide minimally invasive interventional procedures and provide a reference for the positioning of medical biopsy needles, such as breast, liver and kidney disease biopsy. Due to the strong speckle noise in ultrasound images, it is difficult to distinguish the position of needle-like interventional instruments within the tissue, and visualization often relies on the subjective judgment of operators. Additionally, anatomical features, artifacts, and needle-like tissue in ultrasound images can cause traditional needle localization algorithms to fail.

At present, researchers have proposed image-based methods for needle intervention guidance, Zhao *et al.*[1,2] compares different biopsy needle localization algorithms in 2D, 3D and 4D situations to evaluate their accuracy and execution time. To satisfy the demand for localization accuracy and calculation time, the line-filter + RANSAC algorithm is the best choice in the 2D situation, and the ROI based RANSAC and Kalman filter [3] was the most stable and time-saving method in both 3D and 4D situations. But when used for more complex clinical images, image-based methods are less robust and take longer to process.

To address these limitations, many recent studies have used deep learning algorithms. The most representative algorithm in deep learning is convolutional neural network (CNN), which is widely used for detection, localization and segmentation tools in medical images. The method proposed by the Pourtaherian *et al.* [4] used patch classification and semantic segmentation techniques with a CNN architecture and tested their method on in vitro datasets of chicken breast and pork leg with data augmentation. The tip positioning error of this method is less than 0.7 mm. Arif *et al.* [5] also proposed a CNN-based method to locate the needle in a 3D ultrasound model and liver images using a V-Net model with an average needle tip error of 1 mm and an angular error of 2 °. Lee *et al.* [6] used a deep learning method based on LinkNet architecture to segment kidney biopsy needles in 2D ultrasound images. Dice similarity coefficient (DSC) is 56.65%, root mean square (RMS) distance error is 9.5 pixels, RMS angle error is 13.3 °.

The application of 3D ultrasound in minimally invasive interventional procedures is still limited, and 2D ultrasound remains the main way to provide real-time guidance for interventional procedures [7,8]. Therefore, this is the focus of this study. Several common structures and improvements of CNN, such as PSPNet [9], U-Net [10], Attention U-Net [11], are implemented for the biopsy needle segmentation using the simulated datasets. The performances of the three network structures are also evaluated.

The paper is organized as follows: part II introduces the network structure used for needle detection; part III presents the data simulated procedure; part IV gives the simulation result, and the conclusion is given in part V.

II. METHODS

In this paper, three kinds of deep networks are chosen. PSPNet with pyramid pooling module is widely used in the field of semantic segmentation. Since medical images usually contain noise and blurred boundaries, it is difficult to detect or identify objects in ultrasound images only using the shallow features of the image. Meanwhile, due to the lack of detailed information of images, it is impossible to obtain accurate boundaries only by relying on image semantic features. While U-Net combines low-resolution feature maps and high-resolution feature maps through skip connections, which

* This research was partially supported by the Natural Science Foundation of China (62173116) the Postdoctoral Research Funds of Heilongjiang Province (Grant No. LBH-TZ13 and No. LBH-Z15068), postdoctoral initial funding of Heilongjiang Province (No. LBH-Q21097) and Degree & Postgraduate Education Reform Project of Harbin Institute of Technology (No. 21MS002).

Yue Zhao is with Control Theory and Engineering, School of Astronautics, Harbin Institute of Technology, Harbin, China. (Corresponding author, 0086-451-86413341, yue.zhao@hit.edu.cn)

Yi Lu and Jing Jin are with Control Theory and Engineering, School of Astronautics, Harbin Institute of Technology, Harbin, China.(hit_luyi@163.com, jinjinghit@hit.edu.cn).

Xin Lu is with School of Computer Science and Informatics Faculty of Computing, Engineering and Media (CEM), De Montfort University, Leicester, United Kingdom. (Co-Corresponding author, +44(0)1162577201, xin.lu@ dmu.ac.uk)

Lin Tao and Xi Chen are with Department of Mammary Surgery, The Second Affiliated Hospital of Harbin Medical University, Harbin, China (13804517666@163.com).

effectively fuses shallow and deep image features, and it is widely used in the field of medical image segmentation. Due to the problem of a large semantic gap between low-resolution and high-resolution features in the U-Net structure, attention U-Net with attention module is the improvements to this problem.

A. *PSPNet (Pyramid Scene Parsing Network)*

Most of the existing semantic segmentation neural networks are improved from FCN (Fully Convolutional Network), but do not fully consider the connection between pixels and surrounding pixels, which makes the network lose consistency. To solve this problem, PSPNet is proposed. When the receptive field is large, the probability of mis-segmentation will be low. Based on this idea, PSPNet is mainly based on FCN combined with the pyramid pooling structure. The features at different scales are synthesized to obtain more global information and provide more effective features for pixel classification. Fig. 1 shows the structure of the PSPNet.

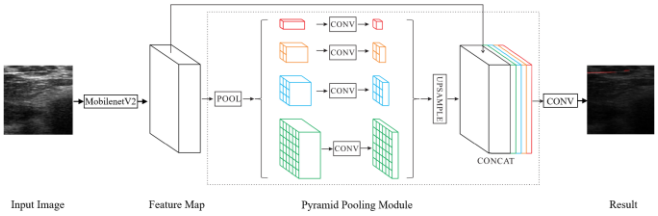

Fig. 1 PSPNet Structure Diagram.

The leftmost is the input image. Before down-sampling, according to the characteristics of the input containing biopsy needle images, an undistorted resize is first performed on the image, so that the image is resized to a size of 473×473 pixels. Then pass it into the backbone feature extraction network. MobileNetV2 is used in this article. The feature map can be obtained.

The feature map is concatenate with the part processed by the pyramid pooling module. For the module, the input biopsy needle feature map will be divided into 6×6 area, 3×3 area, 2×2 area and 1×1 area, and for each area average pooling is done. Corresponding to the green, blue, orange and red areas in Fig. 1, respectively.

B. *U-Net*

U-Net is a network commonly used for medical image segmentation with small amount of data. It is improved on the basis of FCN. Jump connections and symmetry are typical characteristics of U-Net. The network structure of U-Net is divided into two parts. The left side is the down-sampling part, and the right side is the up-sampling part, as shown in Fig. 2.

The size of the 2D simulated ultrasound image containing the medical biopsy needle obtained is 256×256×3. When this is used as the input in the experiment, the specific implementation structure is shown in Fig. 2. After down-sampling on the left, five preliminary effective feature layers are obtained, and then up-sampling and feature fusion on the right are performed. The final feature layer extracted by the network is 256×256×64.In the final segmentation part, 1×1 convolution is used to adjust the number of channels, and the finally obtained 256×256×64 feature layer is adjusted to 256×256×2.

C. *Attention U-Net*

U-Net uses the idea of skip connection to stack the feature map obtained by down-sampling with the up-sampling result.

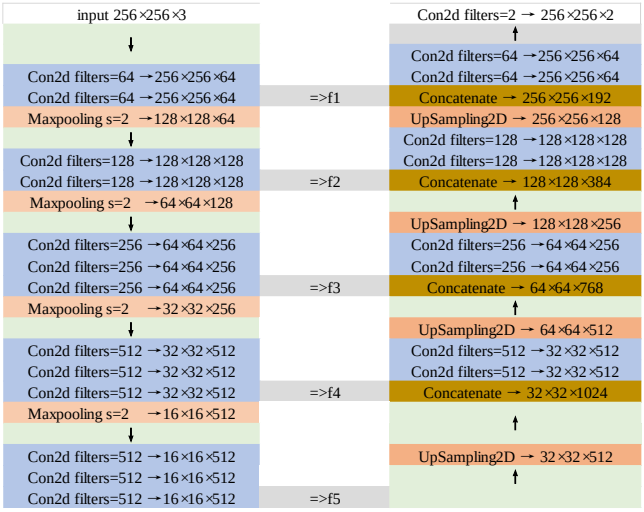

Fig. 2 Schematic diagram of the numbers of feature fusion channels of the U-Net.

This operation avoids losing more spatially accurate details during the down-sampling process, but extracts the low-dimensional feature map contains more redundant information. For the biopsy needle segmentation task, it has a small foreground and a large background. It is necessary to add an attention module, as shown in Fig. 3. It weights each pixel of the feature map. This assigns larger weights to biopsy needle regions, suppressing activation in low-correlation regions, and thus reducing redundant information.

The attention module mainly adjusts the weight of the biopsy needle in the input feature map. Apply it to the U-Net structure as shown in Fig. 3. Before splicing the feature map obtained by down-sampling and the feature map obtained by up-sampling, the attention module is used to achieve the goal of eliminating irrelevant feature content generated by direct skip connections and merging related activation features.

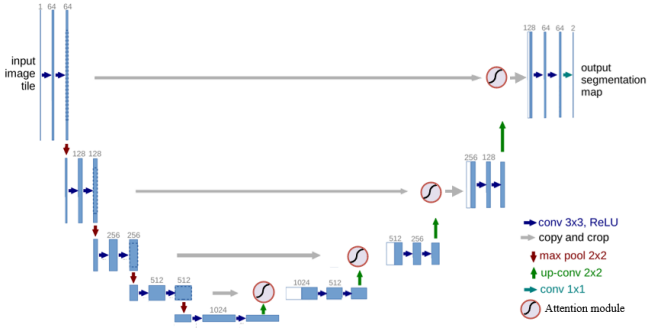

Fig. 3 Schematic diagram of Attention U-Net structure.

III. Data Generation

A. *Experiment Data*

The Field-II Matlab toolbox is used to simulate the sound field produced by a medical ultrasound imaging system. We mainly use the Field-II [12, 13] toolbox to generate the medical biopsy needles, and then use the real breast ultrasound images as the background. The two are superimposed to

obtain the experimental ultrasound images. Finally, perform labeling and data set division.

1) Generation of simulation images

The background of the 2D ultrasound image containing the medical biopsy needle is derived from real breast ultrasound data. In a 50*40*10 mm cube space, 100,000 stutterers that obey a uniform distribution are generated, and the reflection coefficients of these scatters obey a Gaussian distribution.

The diameter of the commonly used needle in breast biopsy is 14 Ga., so the radius of the simulated medical biopsy needle is specified as 2.11mm. Two points are randomly generated in the space within the specified range of the cube, and the circle with the two points as the center is the upper and lower bottom surfaces of the cylindrical biopsy needle. The biopsy needle is generated by drawing a cylinder using the line connecting the two points as the normal vector. Superimposed with the real breast ultrasound background to obtain simulated images.

2) Dataset visualization

According to the real situations, 2D simulated ultrasound images each containing one medical biological needle with different positions, different lengths, different insertion angles, and different backgrounds are generated. There are a total of 999 needle pictures, of which 279 pictures have the needle running through the whole picture, and 720 pictures have the short length and the needle inserted at one end, as shown in Fig.4.

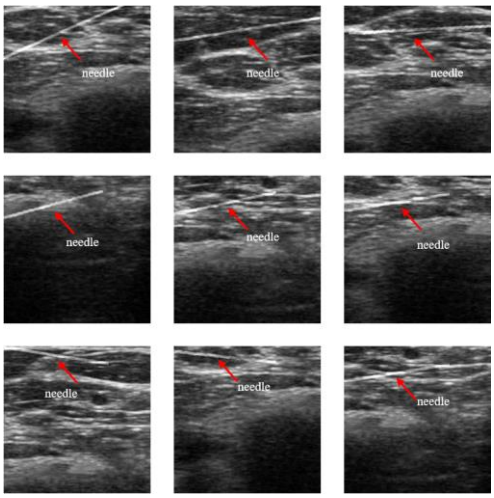

Fig. 4. 2D simulated ultrasound images with different needle positions, angles and lengths.

IV. RESULTS

A. Evaluation Criteria

1) Precision

Among the evaluation criteria of semantic segmentation, the most basic and most widely used is Pixel Accuracy (PA). The pixel accuracy of each category is also called Precision.

For the segmentation task of biopsy needles in this paper, this category is only the foreground — medical biopsy needles, and the proportion of biopsy needles in the entire image is only 1.8% - 2.7%. It is better to use the accuracy of biopsy needles as the evaluation criteria.

$$Precision_{needle} = \frac{TP}{TP + TP} \tag{1}$$

here *TP* and *FP* denote true positive, and false positive, respectively.

2) Intersection over Union (IoU)

Another representative evaluation criterion is the Intersection over Union (IoU). Essentially a way to quantify the percentage of overlap between the target true label and our predicted output:

$$IoU = \frac{A_p \cap A_g}{A_p \cup A_g} \tag{2}$$

here, A_g and A_p represent the ground truth and the network predictions, respectively.

3) Angular Error

The angular error is used to characterize the angular deviation between the actual position and the predicted position of the medical biopsy needle. Calculate the straight line equation between the predicted position and the real position based on the least squares method, and then calculate the angle. As follows:

$$\theta = \arctan \frac{|k_2 - k_1|}{1 + k_1 \times k_2} \tag{3}$$

here, k_1 and k_2 represent the biopsy needle label slope and segmentation slope obtained by the least squares method.

B. Model training

999 2D simulated ultrasound images are divided into training set, validation set, and testing set, which are 809 images, 90 images, and 100 images, respectively. The network models are built using the Pytorch framework. The system environment used is Ubuntu 18.04, the graphics card is NVIDIA GeForce RTX 3090, and the memory is 24G, which meets the training requirements. A hybrid loss function consisting of Cross Entropy loss and Dice loss is used in network training. As follows:

$$\mathcal{L}(Y,P) = -\frac{1}{N}\sum_{c=1}^{C}\sum_{n=1}^{N}\left(y_{n,c}\log p_{n,c} + \frac{2y_{n,c}p_{n,c}}{y_{n,c}^2 + p_{n,c}^2} \right) \tag{4}$$

here $y_{n,c} \in Y$ and $p_{n,c} \in P$ denote the target labels and predicted probabilities for class c and n^{th} pixel in the batch. Y and P are the ground-truth and prediction results. C and N represent the number of dataset classes and pixels in the batch.

During model training, the idea based on transfer learning uses the pre-trained weights on ImageNet, and freezes the training to speed up the training and prevent the weights from being destroyed in the early stage of training. The initial learning rate is $lr = 1\times10^{-4}$, and the freezing epoch is 50, batch size is set as 8. The learning rate after unfreezing is $lr = 1\times10^{-5}$, and the unfreezing epoch is 50, batch size equals to 4.

C. Results and analysis

After training and validation, the three well trained deep networks are implemented to the task of needle detection and localization on the testing set. The segmentation results of the needle positions are shown in Fig. 5.

From Fig. 5, it can be seen from the original images that the biopsy needles are unclearly with the inhomogeneous

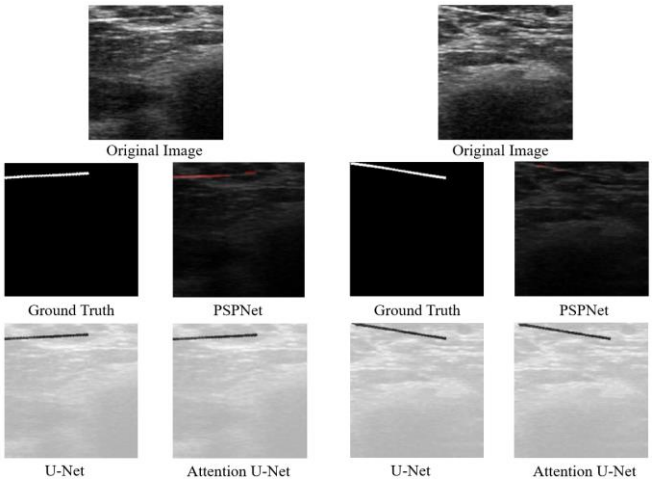

Fig. 5 Comparison of segmentation results on two different test set images

TABLE I. THE RESULTS OF THE THREE NETWORK STRUCTURES USING THE TESTING SET

Network	Evaluation criteria		
	IoU	***Precision***	***Angular Error***
PSPNet	59.11%	78.34%	$3.20^\circ \pm 1.28^\circ$
U-Net	89.99%	95.55%	$\mathbf{0.17}^\circ \pm 0.07^\circ$
Attention U-Net	**90.19%**	**96.25%**	$0.40^\circ \pm 0.03^\circ$

ultrasound background. It is undoubted that the needle localization would fail using the traditional needle localization algorithms. However, using the deep network structures, the needles are successfully segmented from the inhomogeneous background. The results of the needle segmented using PSPNet is worse than the U-Net and attention U-Net. To make the visualization clearer, the brightness flip of the result of U-Net and Attention U-Net has been done. To further evaluate the performance of the different structures, the evaluation criteria in IV. A are calculated and shown in TABLE I.

Since PSPNet designs four pooling kernels of different sizes to encode global contextual information, but cannot well recover detailed information loss caused by merging biopsy needle features during up-sampling. The segmentation results of needles are worse than the other two networks.

Attention U-Net uses the attention module to give a larger weight to the biopsy needle part, so both IoU (90.19 %) and precision (96.25 %) are better than those of the U-Net (89.99 % and 95.55 %, respectively). This indicates that the segmentation boundary is also clearer. However, this also leads to the fact that when performing least squares fitting, the ordinate of the pixels of the segmentation result using attention U-Net has a larger error than the straight line fitting using that of the U-Net. Thus, the calculated angular error of the attention U-Net is 0.40˚, larger than that calculated using the U-Net's segmentation result.

V. CONCLUSION

2D ultrasound is often used for the real-time localization and tracking of the medical biopsy needles in minimally invasive surgery. Three deep network structures are implemented for the needle localization task in this paper. The result shows that Attention U-Net has the best performance among the three networks and can accurately detect the real position of the medical biopsy needle under the inhomogeneous background. Further research will concentrate on the lightweight of the deep network model and implemented the needle localization deep network on the portable devices.

REFERENCES

[1] Y. Zhao, H. Liebgott, C. Cachard, Comparison of the existing tool localization methods on two-dimensional ultrasound images and their tracking results, Control Theory & Applications, IET, Vol 9, Issue 7, pp. 1124-1134.

[2] Y. Zhao, Y. Shen, A. Bernard, C. Cachard, H. Liebgott, Evaluation and Comparison of Current Biopsy Needle Localization and Tracking Methods Using 3D Ultrasound, Ultrasonics, Elsevier, Vol 73, pp. 206-220.

[3] Y. Zhao, C. Cachard, H. Liebgott, Automatic needle detection and tracking in 3D ultrasound using an ROI-based RANSAC and Kalman method, Ultrasonics. Imag.35 (4) (2013) pp. 283–306.

[4] A. Pourtaherian, F.G. Zanjani, S. Zinger, N. Mihajlovic, G. Ng, H. Korsten, P. D. With. Robust and semantic needle detection in 3D ultrasound using orthogonal-plane con-volutional neural networks. Int J Comput Assist Radiol Surg. 2018; 13:1321–1333.

[5] M. Arif, A. Moelker, T. Van Walsum. Automatic needle detection and real-time bi-planar needle visualization during 3D ultrasound scanning of the liver. Med Image Anal. 2019; 53:104–110.

[6] Y. L. Jia, M. Islam, R.W. Jing, T. S. M. Washeem, H. Ren. Ultrasound needle segmentation and trajectory prediction using excitation network. Int J Comput Assist Radiol Surg. 2020; 15:437–443.

[7] R. W. Prager, U. Z. Ijaz, A. H. Gee, G. M. Treece. Three-dimensional ultrasound imaging. Proc Inst Mech Eng Part H J Eng Med. 2010;224:193–223.

[8] L. J. Brattain, N. V. Vasilyev, R. D. Howe. Enabling 3D Ultrasound Procedure Guidance through Enhanced Visualization. In: Lecture Notes in Com-puter Science (Including Subseries Lecture Notes in Artificial Intelli-gence and Lecture Notes in Bioinformatics). Vol 7330 LNCS; 2012:115–124.

[9] H. Zhao, J. Shi, X. Qi, X. Wang, J. JiaZhao, "Pyramid Scene Parsing Network," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, pp. 6230–6239, 2017.

[10] O. Ronneberger, P. Fischer, B. Brox, "U-Net: Convolutional networks for biomedical image segmentation," Int'l. Conf. on Medical Image Computing & Computer-assisted Intervention (Springer, Cham, Switzerland, 2015), pp. 234–241.

[11] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. Mcdonagh, N. Y. Hammerla, B. Kainz, "Attention Unet: Learning where to look for the pancreas", MIDL. 2018.

[12] J.A. Jensen, "Field: A Program for Simulating Ultrasound Systems", Paper presented at the 10th Nordic-Baltic Conference on Biomedical Imaging Published in Medical & Biological Engineering & Computing, pp. 351-353, Volume 34, Supplement 1, Part 1, 1996.

[13] J.A. Jensen, N. B. Svendsen "Calculation of pressure fields from arbitrarily shaped, apodized, and excited ultrasound transducers", IEEE Trans. Ultrason., Ferroelec., Freq. Contr., 39, pp. 262-267, 1992.