Gao J, Liu P, Liu GD *et al.* Robust needle localization and enhancement algorithm for ultrasound by deep learning and beam steering methods. JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY 36(2): 334–346 Mar. 2021. DOI 10.1007/s11390-021-0861-7

Robust Needle Localization and Enhancement Algorithm for Ultrasound by Deep Learning and Beam Steering Methods

Jun Gao[1,2], Paul Liu[2], Guang-Di Liu[3], and Le Zhang[1,4,5,*], *Senior Member*, *CCF*, *Member*, *ACM*

[1] *College of Computer Science, Sichuan University, Chengdu 610065, China*

[2] *Stork Healthcare, Chengdu 610000, China*

[3] *College of Computer and Information Science, Southwest University, Chongqing 400715, China*

[4] *West China Biomedical Big Data Center, West China Hospital, Sichuan University, Chengdu 610065, China*

[5] *Pera Corporation Ltd., Beijing 100025, China*

E-mail: gaojun@stu.scu.edu.cn; paul@storkhealthcare.com; liuguangdi@cdu.edu.cn; zhangle06@scu.edu.cn

Received July 29, 2020; accepted February 25, 2021.

Abstract Ultrasound (US) imaging is clinically used to guide needle insertions because it is safe, real-time, and low cost. The localization of the needle in the ultrasound image, however, remains a challenging problem due to specular reflection off the smooth surface of the needle, speckle noise, and similar line-like anatomical features. This study presents a novel robust needle localization and enhancement algorithm based on deep learning and beam steering methods with three key innovations. First, we employ beam steering to maximize the reflection intensity of the needle, which can help us to detect and locate the needle precisely. Second, we modify the U-Net which is an end-to-end network commonly used in biomedical segmentation by using two branches instead of one in the last up-sampling layer and adding three layers after the last down-sample layer. Thus, the modified U-Net can real-time segment the needle shaft region, detect the needle tip landmark location and determine whether an image frame contains the needle by one shot. Third, we develop a needle fusion framework that employs the outputs of the multi-task deep learning (MTL) framework to precisely locate the needle tip and enhance needle shaft visualization. Thus, the proposed algorithm can not only greatly reduce the processing time, but also significantly increase the needle localization accuracy and enhance the needle visualization for real-time clinical intervention applications.

Keywords ultrasound, deep learning, segmentation, classification, optimization

1 Introduction

Currently, ultrasound (US) is commonly used for interventional operations such as therapeutic injection, biopsy, and anesthesia [1], which usually require high image quality of the needle shaft and tip for precise localization. However, specular reflection of the smooth needle surface often causes returning echoes to reflect to positions outside the probe field of view, which results in very blurred or even invisible needles [2]. Furthermore, high-intensity signals coming from soft tissue and bone along the needle trajectory may interfere with the needle visibility, especially for complicated tissues [2]. Therefore, accurate needle detection and localization remains an ongoing challenge in clinical intervention applications.

At present, needle detection and localization methods consist of hardware-based and software-based methods. The hardware-based methods include electromagnetic tracking [3], optical tracking [4], and embedded sensors at the needle tip [5, 6]. Recently, software-based methods have been commonly used for needle detection and localization, since hardware-based methods [3–6] are relatively expensive, impose higher environmental re-

Regular Paper

Special Section on AI and Big Data Analytics in Biology and Medicine

This work was supported by the National Science and Technology Major Project of China under Grant No. 2018ZX10201002.

*Corresponding Author

quirements, and hinder the standard clinical workflow. Usually, software-based methods include projection-based [7, 8], 3D ultrasound [9, 10] and line detection [11–13] methods, which mainly depend on the intensity of the shaft and tip reflection. Therefore, once the intensities of the shaft and tip are insignificant, these methods become unreliable.

Recently, learning-based methods [14–20] have made great progress in precise localization and robustness. For example, Hatt *et al.* developed a learning-based method with beam steering [20], which obtains in-plane linear probe needle localization with a mean targeting error of 0.48 mm and a 86.2% localization success rate in *ex vivo* datasets. However, these algorithms are so computationally intensive [14–16, 20] or require such large amounts of input data [17, 18] that they cannot real-time locate the needle.

For this reason, Mwikirize *et al.* developed a real-time needle tip localization method to locate the needle tip in 67 fps with a 0.55 mm tip localization error based on a cascade of simple neural networks [19]. Nevertheless, this method is not robust against motion artifacts since it not only heavily relies on the subtraction of consecutive frames [19], but also cannot detect and enhance the needle shaft.

To address these challenges, we propose a robust needle localization and enhancement algorithm (NLEM) with three innovations to overcome the previous problems. First, we employ beam steering to maximize the reflection intensity of the needle, which can help us to detect and locate the needle precisely. Second, we modify the U-Net [21], an end-to-end network commonly used in biomedical segmentation, by using two branches instead of one in the last up-sampling layer and adding three layers after the last down-sample layer. Thus, the modified U-Net can real-time segment the needle shaft region, detect the needle tip landmark location and determine whether an image frame contains the needle by one shot. Third, we develop a needle fusion framework that employs the outputs of the multi-task deep learning framework (MTL) to precisely locate the needle tip and enhance the needle shaft visualization.

Our experimental results show that the tip localization errors for the NLEM method are approximately 0.29±0.02 mm and the overall processing time for the NLEM is 0.014 9 s per frame, which meets the requests of real-time clinical intervention applications. Here, we consider that beam steering innovation increases the reflection signal of the needle to a greater extent as well as the MTL network and the fusion framework greatly reduces the processing time, improves the localization accuracy, and significantly enhances the needle visualization.

2 Proposed NLEM

Fig.1 shows the three steps of the NLEM algorithm. Step 1 illustrates that the ultrasound front-end (FE) alternately emits normal-angled and steered US beams, forming the normal-angled image (NI) and steered images (SI), respectively. Next, we train the MTL framework by deep learning, which is detailed in Fig.A1(a) of supplementary materials of [22]. After the MTL framework (step 2) receives the SI from step 1, it segments the needle shaft region (NS), detects the needle tip landmark (LM) location, and predicts the probability of SI including the needle (P_{needle}). Subsequently, step 3 shows that NLEM employs the outputs (NS, LM, P_{needle}) from the MTL framework to precisely locate the needle tip and enhance the needle shaft visualization. The code for the NLEM algorithm is available online①.

2.1 Datasets and Experimental Setup

We obtain ultrasound images by the iMagoC21 (Saset Healthcare, USA) ultrasound scanner with a L5-12 linear probe and a 2D handheld wireless ultrasound (H35L, Stork Healthcare, China). The insertion angle and the depth of the needle range from (10° and 20 mm) to (45° and 50 mm).

During data acquisition, the radiologist adjusts the steer angle (SA) of the FE to guarantee that the steered beam is perpendicular to the needle. We collect 80 video sequences from porcine, bovine, and chicken tissues using a hollow 18 G cannula needle, which is the same as in previous studies [15, 19, 20].

Each sequence has more than 200 frames. The training image set for our NLEM method consists of 7 500 images from 60 sequences, in which 5 000 images have the needle and 2 500 images do not have the needle. The testing image set includes 1 200 images from 20 sequences, in which 800 have the needle and 400 images do not have the needle. All of the images are scaled to 256×256 as the previous research [19]. The coordinates of the needle tip and tail are annotated in the steered images by two expert radiologists.

① https://github.com/gaojun0821/NLEM, Feb. 2021.

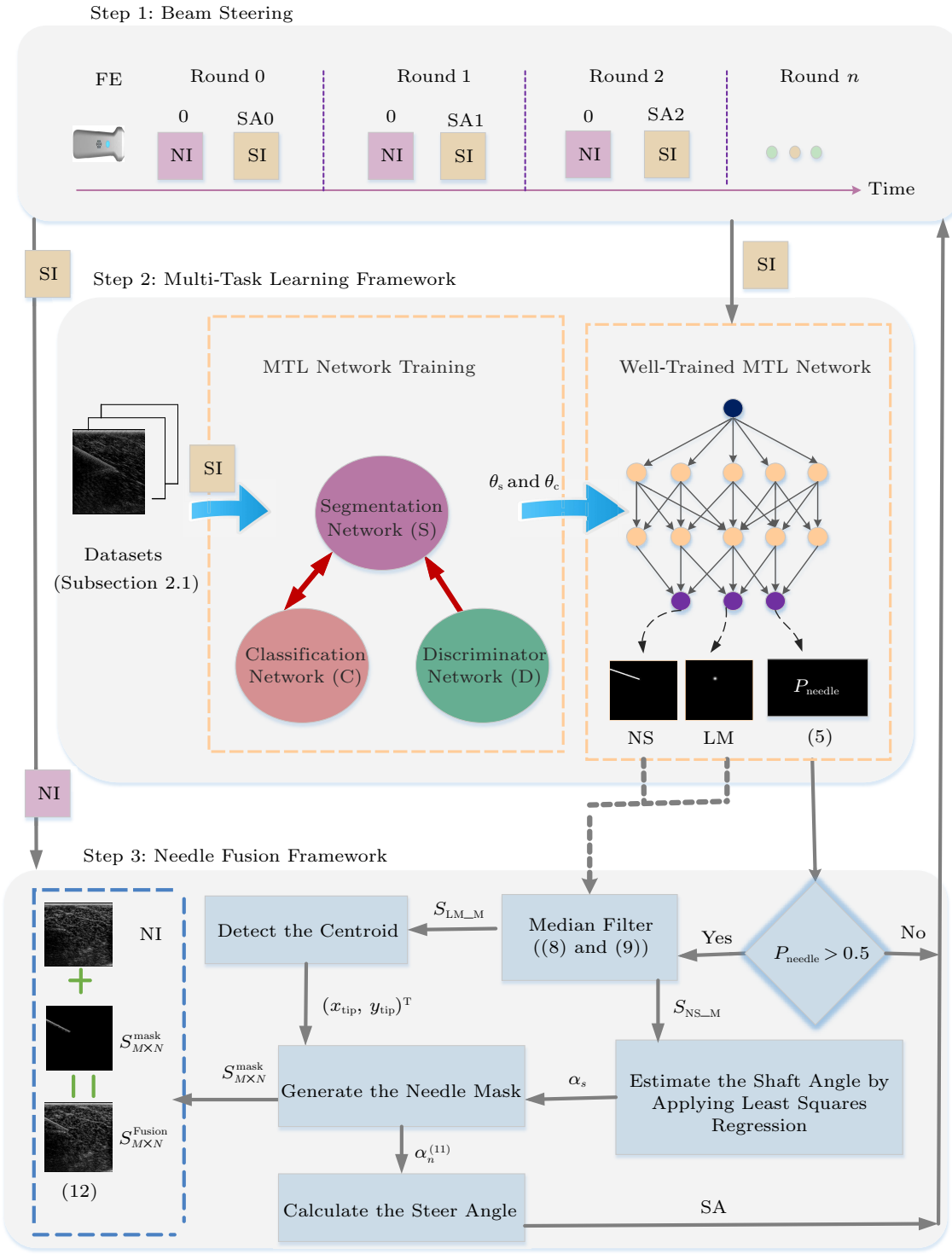

Fig.1. Workflow of NLEM. Details of the centroid detection, the needle mask generation and the steer angle calculation in step 3 are described by Fig.A1(b), Fig.A1(c) and Fig.A1(d) of the supplementary materials of the paper [22], respectively.

According to the coordinate information, the computer program automatically generates the shaft mask annotations and tip landmark annotations. The landmark annotation is as the same size as the steered image and produces a distance-based 2D Gaussian map centered at the tip coordinate ($\sigma = 3.0$ pixels).

The shaft mask annotation is a line segment that denotes the needle tail as the starting point and the

needle tip as the end point, and the width of a line is ω pixels. The pixel values on the line are set to 255 and the other pixels are set to 0. The value of ω depends on the type of needles and the resolution of the US image. For example, when the scanning depth is 40 mm, the number of longitudinal sampling points of the image is 512, and the type of needle is 18 G; then, ω is set to 8 pixels.

For negative examples, we generate an empty needle shaft mask and an empty tip landmark in which all of pixels are set to 0. We implement the proposed NLEM method on an NVIDIA GeForce GTX 1050Ti 4 GB GPU, 3.70 GHz Intel® Core™ i7-8700k 16 GB CPU Windows PC [23–25]. The needle fusion framework is implemented in MATLAB [19]. The MTL framework is implemented in Keras 2.2.4 on the Tensorflow 2.0.0 backend [15].

2.2 MTL Framework

Fig.2 and Fig.A2 in the supplementary materials of [22] show that the MTL framework consists of a segmentation network (S), a discriminator network (D) and a classification network (C). Here, the segmentation network is used to estimate the needle shaft (NS) region and the tip landmark (LM); the discriminator network is used to improve the training effect for the segmentation network; the classification network is used to determine if an SI frame has a needle or not.

2.2.1 Segmentation Network

The segmentation network is based on the U-Net [21] architecture, which is a U shape network. We develop this modified U-Net (segmentation network) by using two branches instead of one in the last up-sampling

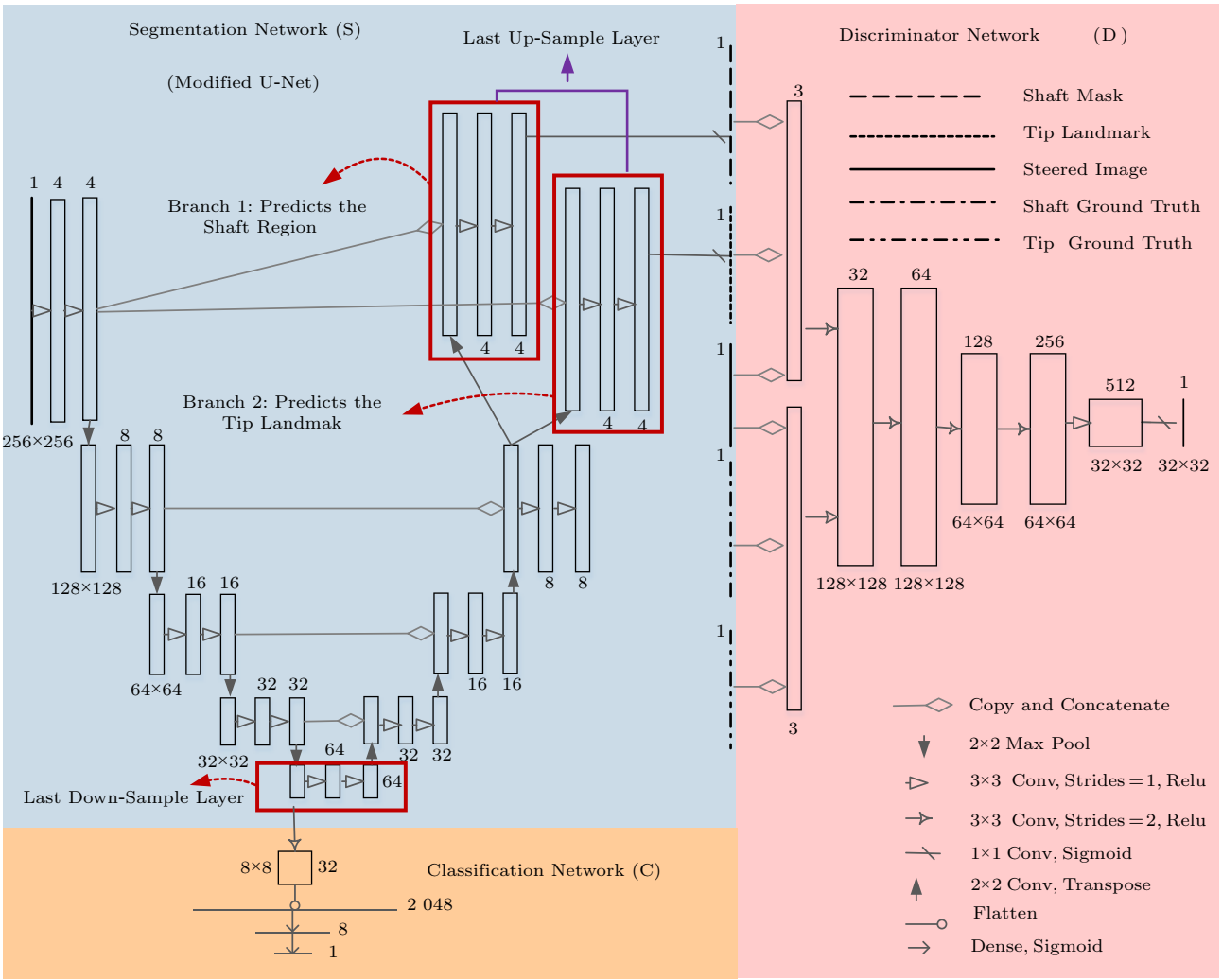

Fig. 2. Structure of segmentation network (S), classification network (C) and discriminator network (D) for the proposed MTL framework. The discriminator network (D) is only used in the training phase.

layer, as shown in Fig.2. One branch predicts the shaft region, and the other branch detects the landmark location of the tip.

We denote $S_{\mathrm{NS}}(SI, \theta_s)$ and $S_{\mathrm{LM}}(SI, \theta_s)$ as the functions to estimate the needle shaft (NS) region and tip landmark (LM) from the input steered image (SI), respectively. θ_s is the parameter for the segmentation network (S). To train the segmentation network, we use a weighted binary cross-entropy L_{NS} to compare the shaft prediction p_i with the ground truth $\hat{p}_i$ by (1).

$$L_{\mathrm{NS}} = w_{\mathrm{ns}} \sum_i \left(-\left(\hat{p}_i \log p_i + (1 - \hat{p}_i) \log (1 - p_i) \right) \right), \quad (1)$$

where i represents the index of a pixel. The weight of $w_{\mathrm{ns}} = T/(2T_{c1} + 1)$ is given to class $c1$, where $c1 \in \{\text{shaft mask, background}\}$. T is the total number of pixels in a training sample, and T_{c1} denotes the number of pixels in class $c1$.

For the landmark detection, (2) uses a weighted mean square error L_{LM} as the loss function.

$$L_{\mathrm{LM}} = w_{\mathrm{lm}} \sum_{i=1}^{N} (g_i - \hat{g}_i)^2, \quad (2)$$

where $\hat{g}_i$ and g_i are the ground truth and the prediction of the tip landmark respectively, and i represents the index of a pixel. To balance against the number of samples during the training, a weight of $w_{\mathrm{lm}} = N \times (N + N_{c2}) / N_{c2}$ is given to class $c2 \in \{\text{tip landmark, background}\}$, where N is the total number of pixels and N_{c2} denotes the number of pixels in class $c2$.

2.2.2 Discriminator Network

The predicted shaft region and the tip landmark are concatenated in the last channel and then passed to the discriminator network (D) with adversarial training [26]. The concatenated predictions from the segmentation network (S) to identify true or false outputs on the basis of N_m patches. The discriminator network (D) regularizes the output with binary cross-entropy. The discriminator loss L_{AD} is described by (3).

$$L_{\mathrm{AD}} = \frac{1}{N_m} \sum_m \left(\sum\nolimits_{i \in m} -(\hat{y}_i \log y_i + (1 - \hat{y}_i) \log (1 - y_i)) \right), \quad (3)$$

where m indicates a single patch, $\hat{y}_i$ is the real true/false, and y_i is the probability of true/false prediction. Here, i represents the index of m. The discriminator network learns to discriminate the distributions of the ground truth annotations against the outputs of the segmentation model.

2.2.3 Classification Network

Here, we develop the classification network to determine if a needle is in the SI image or not. Described in Fig.2, this network shares the down-sampling layers of modified U-Net with the segmentation network.

To train the classification network, we use traditional binary cross-entropy to compare the needle prediction with the label of ground truth $\hat{c}_i$ by (4).

$$L(\theta_c) = \sum_i \left(-\left(\hat{c}_i \log c_i + (1 - \hat{c}_i) \log (1 - c_i) \right) \right), \quad (4)$$

where c_i is the predictive probability of needle, i represents the index of the training samples, and θ_c is the parameter for classification network (C).

Here, we define the probability P_{needle} as the output of (5) and set SI and θ_c as the input of (5) for the classification network (C).

$$P_{\mathrm{needle}} = Classification\_network(\theta_c, SI). \quad (5)$$

2.2.4 Training for the Network

Fig.2 and Fig.A1(a) in the supplementary materials of [22] detail the network structure and training process, respectively.

Given the set of predictions $\{S_{\mathrm{NS}}(SI; \theta_s), S_{\mathrm{LM}}(SI; \theta_s)\}$ and $D(U; \theta_d)$, we define the optimized function for the segmentation network as in (6).

$$\begin{aligned} L(\theta_s) = {} & \mu_1 L_{\mathrm{NS}}(S_{\mathrm{NS}}(SI; \theta_s), \hat{p}) + \\ & \mu_2 L_{\mathrm{LM}}(S_{\mathrm{LM}}(SI; \theta_s), \hat{q}) + \\ & \mu_3 L_{\mathrm{AD}}(D(m; \theta_d), \hat{t}), \end{aligned} \quad (6)$$

where $\hat{p}$ and $\hat{q}$ are the ground truth for the shaft segmentation and tip landmark localization respectively, μ_1, μ_2, and μ_3 are the weighting parameters of the individual loss terms, and $\hat{t}$ is a 32×32 patch with the value of 1. The discriminator network (D) is additionally trained with (7) to classify the distribution of ground truth annotations against the distribution of predicted masks by the segmentation networks (S).

$$\begin{aligned} L(\theta_d) = {} & L_{\mathrm{AD}}(D(\text{Concatenate}(\hat{p}, \hat{q}, SI); \theta_d), \hat{t}) + \\ & L_{\mathrm{AD}}(D(U; \theta_d), \hat{z}), \end{aligned} \quad (7)$$

where $\hat{z}$ is a 32×32 patch with the value of 0, $U = \text{Concatenate}(S_{\mathrm{NS}}(SI; \theta_s), S_{\mathrm{LM}}(SI; \theta_s), SI)$, D is the function representing the discriminator network, and θ_d

is the parameter for the discriminator network. Here, Concatenate() is a commonly used function in deep learning, described in [18].

Table A1 in the supplementary materials of [22] lists the hyper parameters for the MTL network.

2.3 Needle Fusion Framework

From the segmentation network (S), we obtain the needle shaft segmentation S_{NS} and the needle tip landmark S_{LM}. Then, we employ (8) and (9) to process S_{NS} and S_{LM} with a commonly used 3×3 median filter [14] in image processing to yield $S_{\mathrm{NS\_M}}$ and $S_{\mathrm{LM\_M}}$, respectively.

$$S_{\mathrm{NS\_M}} = \mathrm{medfilt2}\left(S_{\mathrm{NS}},\ [3,3]\right), \tag{8}$$

$$S_{\mathrm{LM\_M}} = \mathrm{medfilt2}\left(S_{\mathrm{LM}},\ [3,3]\right). \tag{9}$$

Subsequently, we detect the centroid $(x_{\mathrm{tip}}, y_{\mathrm{tip}})^{\mathrm{T}}$ in the $S_{\mathrm{LM\_M}}$ image and estimate the shaft angle α_s by applying least-squares regression [27] to the $S_{\mathrm{NS\_M}}$ image data. We locate the needle tail coordinates $(x_{\mathrm{tail}}, y_{\mathrm{tail}})^{\mathrm{T}}$ by the tip centroid $(x_{\mathrm{tip}}, y_{\mathrm{tip}})^{\mathrm{T}}$ and the shaft angle (α_s). Here, we use $(x_{\mathrm{nor\_cd}}, y_{\mathrm{nor\_cd}})^{\mathrm{T}}$ to represent the coordinates for the normal-angled image (NI) as well as $(x_{\mathrm{st\_cd}}, y_{\mathrm{st\_cd}})^{\mathrm{T}}$ to represent the coordinates for the steered image (SI), respectively. Thus, $(x_{\mathrm{tail}}, y_{\mathrm{tail}})^{\mathrm{T}}$ and $(x_{\mathrm{tip}}, y_{\mathrm{tip}})^{\mathrm{T}}$ are two unique points for $(x_{\mathrm{st\_cd}}, y_{\mathrm{st\_cd}})^{\mathrm{T}}$; in addition, $(x_{\mathrm{tail\_n}}, y_{\mathrm{tail\_n}})^{\mathrm{T}}$ and $(x_{\mathrm{tip\_n}}, y_{\mathrm{tip\_n}})^{\mathrm{T}}$ are two unique points for $(x_{\mathrm{nor\_cd}}, y_{\mathrm{nor\_cd}})^{\mathrm{T}}$. Then, (10) converts the coordinate locations $(x_{\mathrm{tail}}, y_{\mathrm{tail}})^{\mathrm{T}}$ and $(x_{\mathrm{tip}}, y_{\mathrm{tip}})^{\mathrm{T}}$ in the steered image to the normal-angled image locations $(x_{\mathrm{tail\_n}}, y_{\mathrm{tail\_n}})^{\mathrm{T}}$ and $(x_{\mathrm{tip\_n}}, y_{\mathrm{tip\_n}})^{\mathrm{T}}$.

$$\begin{pmatrix} x_{\mathrm{nor\_cd}} \\ y_{\mathrm{nor\_cd}} \end{pmatrix} = \begin{pmatrix} 1 & -\sin(SA) \\ 0 & \cos(SA) \end{pmatrix} \begin{pmatrix} x_{\mathrm{st\_cd}} \\ y_{\mathrm{st\_cd}} \end{pmatrix}, \tag{10}$$

where SA is the steer angle of the front-end (FE). If $|90^\circ - SA - \alpha_n| > 5^\circ$, then we set the steer angle SA to $90^\circ - \alpha_n$ to guarantee that the steered beam is perpendicular to the needle. The needle angle in the normal-angled image is

$$\alpha_n = \tan^{-1} \frac{(x_{\mathrm{tip\_n}} - x_{\mathrm{tail\_n}})}{(y_{\mathrm{tip\_n}} - y_{\mathrm{tail\_n}})}. \tag{11}$$

As described in Figs.A1(b)–A1(g) of the supplementary materials of [22], we employ (10) to convert such points with steered image (SI) coordinates $(x_{\mathrm{st\_cd}}, y_{\mathrm{st\_cd}})^{\mathrm{T}}$ in the red area into normal-angled image (NI) coordinates $(x_{\mathrm{nor\_cd}}, y_{\mathrm{nor\_cd}})^{\mathrm{T}}$. We define the set of these normal-angled image (NI) coordinates $(x_{\mathrm{nor\_cd}}, y_{\mathrm{nor\_cd}})^{\mathrm{T}}$ as $S^{\mathrm{mask}}_{M\times N}$, which can be imagined as an $M \times N$ normal-angled (NI) image. Here, M and N are the width and the height of the normal-angled (NI) image, respectively.

Finally, (12) embeds $S^{\mathrm{mask}}_{M\times N}$ into the normal angled-image.

$$S^{\mathrm{Fusion}}_{M\times N} = \begin{cases} \max\left(\gamma\left(S^{\mathrm{mask}}_{M\times N}, 255\right)\right), \\ \qquad \text{if} \quad S^{\mathrm{mask}}_{M\times N} > 0, \\ NI, \\ \qquad \text{if} \quad S^{\mathrm{mask}}_{M\times N} = 0, \end{cases} \tag{12}$$

where γ is a fusion coefficient that can be adjusted by the clinical user and the range of γ is from 1.0 to 2.0. It is noted that the needle fusion framework is used when the needle is detected in SI.

3 Results

The section will show the effect of NLEM, evaluate the impact of two major steps on NLEM, and compare NLEM with other state-of-the-art methods.

3.1 Effect of NLEM

Fig.3(a) and Fig.3(g) show that in the normal-angled images, the needle shaft and the needle tip are almost invisible in the bovine and porcine tissues, respectively. Fig.3(b) and Fig.3(h) are the steered images in the bovine and porcine tissues respectively, in which the needle is relatively prominent (shown by the pink box) compared with Fig.3(a) and Fig.3(g). Fig.3(c) and Fig.3(i) are the segmentation results $(S_{\mathrm{NS}}(SI;\theta_s), S_{\mathrm{LM}}(SI;\theta_s))$of the MTL network, respectively. The needle masks of the normal-angled images $(S^{\mathrm{mask}}_{M\times N})$ are illustrated in the red box of Fig.3(d) and Fig.3(j) for the bovine and porcine tissues, respectively, which are extracted from the SI images (Fig.3(b) and Fig.3(h)) by (10). (12) fuses the needles with the normal-angled images (Fig.3(a) and Fig.3(g)) to get Fig.3(f) and Fig.3(k). Comparing the enhanced results of Fig.3(f) and Fig.3(k) with the normal angle images of Fig.3(a) and Fig.3(g) respectively, the needle visualization is significantly enhanced.

Fig.4(a) demonstrates that the segmentation loss of the MTL framework converges after 10 iterations. Fig.4(b) demonstrates that the discrimination loss of the MTL framework oscillates in a narrow range around 0.5 after 10 iterations.

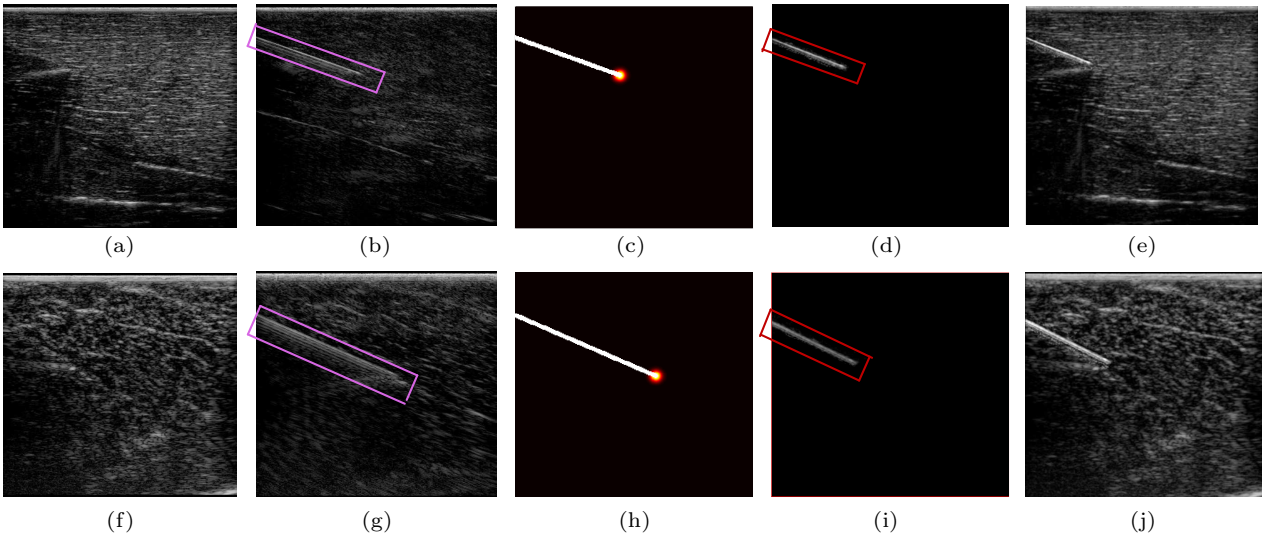

Fig.3. Results of our NLEM method. (a) Normal-angled images (NI) of bovine tissues. (b) Steered images (SI) corresponding to (a). (c) Outputs (S_{NS}, S_{LM}) of our segmentation network corresponding to (b). (d) Needle marks ($S^{\mathrm{mask}}_{M\times N}$) extracted through the needle fusion framework. (e) Fusion results of (a) and (d) ($\gamma = 1.5$). (f) Normal-angled image (NI) of porcine tissues. (g) Steered images (SI) corresponding to (f). (h) Outputs (S_{NS}, S_{LM}) of our segmentation network corresponding to (g). (i) Needle mark ($S^{\mathrm{mask}}_{M\times N}$) extracted through the needle fusion framework. (j) Fusion results of (f) and (i) ($\gamma = 1.6$).

3.2 Evaluation of the Impact of Two Major Steps on NLEM

Fig.1 shows that the NLEM algorithm consists of three steps. Since the beam steering for step 1 and the MTL framework of step 2 are the key components of the algorithm, we verify their impact on the localization, classification accuracy and processing time by five commonly used metrics [14, 16, 19, 20] (shown in Table 1).

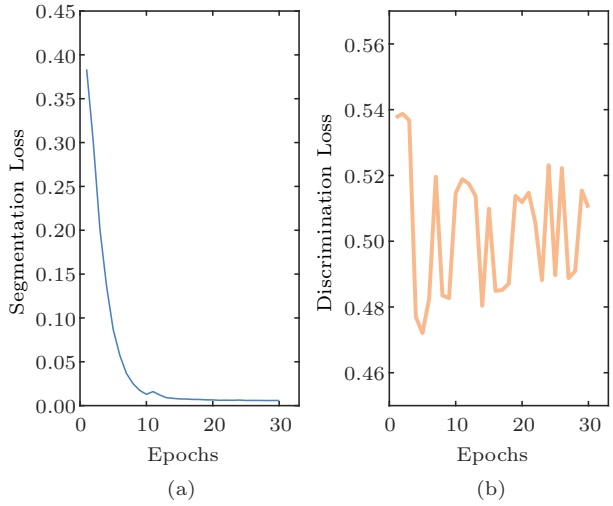

Fig.4. Mode performance during training. (a) Segmentation loss for semantic segmentation of the needle during training on the validation datasets. (b) Discrimination loss for the adversarial training on the validation datasets.

First, we verify the impact of beam steering on the NLEM algorithm. We remove the steered image (SI) in step 1, but keep the other configurations as before, which means that we only use the normal-angled (NI) images to train and test the MTL framework. Fig.5(a) defines this method as "NA_NLEM".

Second, we verify the impact of the MTL framework on the NLEM algorithm. We adopt network ablation studies to verify the effectiveness of the MTL framework, as previous studies [15, 28]. Because U-Net [21] is considered to be a standard state-of-the-art model for medical image segmentation, we train two single task U-Nets. One is used to segment the needle shaft (NS) region and the other is used to detect the needle tip landmark (LM). Here, we set the configuration of each single task U-Net to be identical as the segmentation model (S) of NLEM. As in the previous work [29], we implement a single-task convolution neural network using the same configurations as the proposed classification network (C) to verify the effectiveness of the classification network (C). Fig.5(a) defines this method as "STL_NLEM".

Third, the NLEM method employs adversarial training to reduce artifact noise in the segmentation. Therefore, we remove adversarial training from the NLEM method to investigate the effect of adversarial training on the performance of the NLEM method. Fig.5(a) defines this method as "WAT_NLEM".

Fig.5(a) and Table A2 in the supplementary materi-

Table 1. Metrics for Localization Performance of Needle

Name	Equation	Purpose & Description
TE [15, 20]	$\mathrm{TE} = \left\lvert d_{\mathrm{true}} - d_{\mathrm{pred}} \right\rvert$	Targeting error (TE) [15, 20] is used to measure the localization accuracy of needle orientation. Details of the TE definition refer to Fig.A3 in the supplementary materials of [22].
ED [16, 19]	$\mathrm{ED} = \left\lVert T\left(x_{\mathrm{true}}, y_{\mathrm{true}}\right) - T\left(x_{\mathrm{pred}}, y_{\mathrm{pred}}\right) \right\rVert_2$	Tip localization error (ED) [16, 19] is used to measure the localization error of the needle tip. $T\left(x_{\mathrm{true}}, y_{\mathrm{ture}}\right)$ and $T\left(x_{\mathrm{pred}}, y_{\mathrm{pred}}\right)$ are the ground truth and the prediction coordinate of needle tip, respectively.
SLE [15, 16]	$\mathrm{SLE} = \lvert\beta_1 - \beta_2\rvert$	Shaft localization error (SLE) [15, 16] is used to measure the angle localization error of the needle shaft. Details of the SLE definition refer to Fig.A3 in the supplementary materials of [22].
OPT		The overall processing time (OPT) is used to evaluate the efficiency of the algorithm, which means the sum of total processing time.
1EDLS	$\mathrm{1EDLS} = \frac{N_{(\mathrm{ED}<1)}}{N_{\mathrm{total}}} \times 100\%$	1EDLS is used to measure the success rate of tip localization, where N_{total} is the total number of test samples including the needle, and $N_{(\mathrm{ED}<1)}$ is the number of test samples in which the Euclidean distance between ground truth tip location and prediction tip location is less than 1 mm.

als of [22] compare the localization performance among NLEM, STL_NLEM, WAT_NLEM and NA_NLEM, demonstrating that 1EDLS of the NLEM method is significantly greater than those of the STL_NLEM, WAT_NLEM and NA_NLEM methods; the needle tip localization error (ED) of the NLEM method is significantly less than those of the NA_NLEM, WAT_NLEM and STL_NLEM methods; the needle shaft localization error (SLE) and targeting error (TE) of the NLEM method are significantly less than those of the STL_NLEM, WAT_NLEM and NA_NLEM methods; the overall processing time (OPT) of the NLEM method is significantly less than that of STL_NLEM. However, there is no statistical difference of OPT among NLEM, WAT_NLEM and NA_NLEM. The number of model parameters for NLEM (153.8k) is significantly less than that for the STL_NLEM method (324.6k), whereas NA_NLEM and WAT_NLEM exhibit the same number of model parameters as NLEM.

Next, Fig.5(b) and Table A3 in the supplementary materials of [22] employ four commonly used metrics [30] to validate the performance of our classification model (C) for the NLEM method, which show that the accuracy, precision, sensitivity and specificity are 98.4%, 97.7%, 99.7%, and 96.8%, respectively.

Furthermore, Fig.5(c) shows that approximately 77.5% of the classification failure cases occur when the insertion length of the needle is less than 10 mm.

Lastly, Fig.5(d) shows that approximately 58.9% of the cases of needle tip localization failure (ED >1 mm) occur when the needle tip is occluded by grating lobe artifacts, which is shown in Fig.A4 in the supplementary materials of [22].

3.3 Comparing NLEM with Other State-of-the-Art Methods

Fig.6 shows a comparison among the NLEM method and three commonly used needle localization algorithms, which are CASPER [16], Mwikirze *et al.*'s method [19] and Hatt *et al.*'s method [20]. It is important to note that not every evaluation metric is suitable for all methods. For example, Mwikirize *et al.*'s method [19] only realizes the localization of the needle tip and is unsuitable for the metrics of TE and SLE. In Fig.6, we display the overall processing time (OPT) in logarithmic scale.

Fig.6 demonstrates that the targeting error (TE) of NLEM is less than that of Hatt *et al.*'s method [20]; the tip localization error (ED) of NLEM is less than that of Mwikirze *et al.*'s method [19] and CASPER [16]; the shaft localization error (SLE) of NLEM is less than that of CASPER [16], and the overall processing time (OPT) of NLEM is far less than that of Hatt *et al.*'s method [20] and CASPER [16], and close to that of Mwikirze *et al.*'s method [19].

4 Discussions

To precisely locate and enhance the needle in real time and reduce the difficulty of clinical puncture intervention, we propose a robust needle localization and enhancement algorithm (NLEM) that can increase the needle localization accuracy, reduce the processing time, and enhance the needle visualization.

Shown in Fig.3, NLEM greatly increases the visualization capacity for the needle, since NLEM can

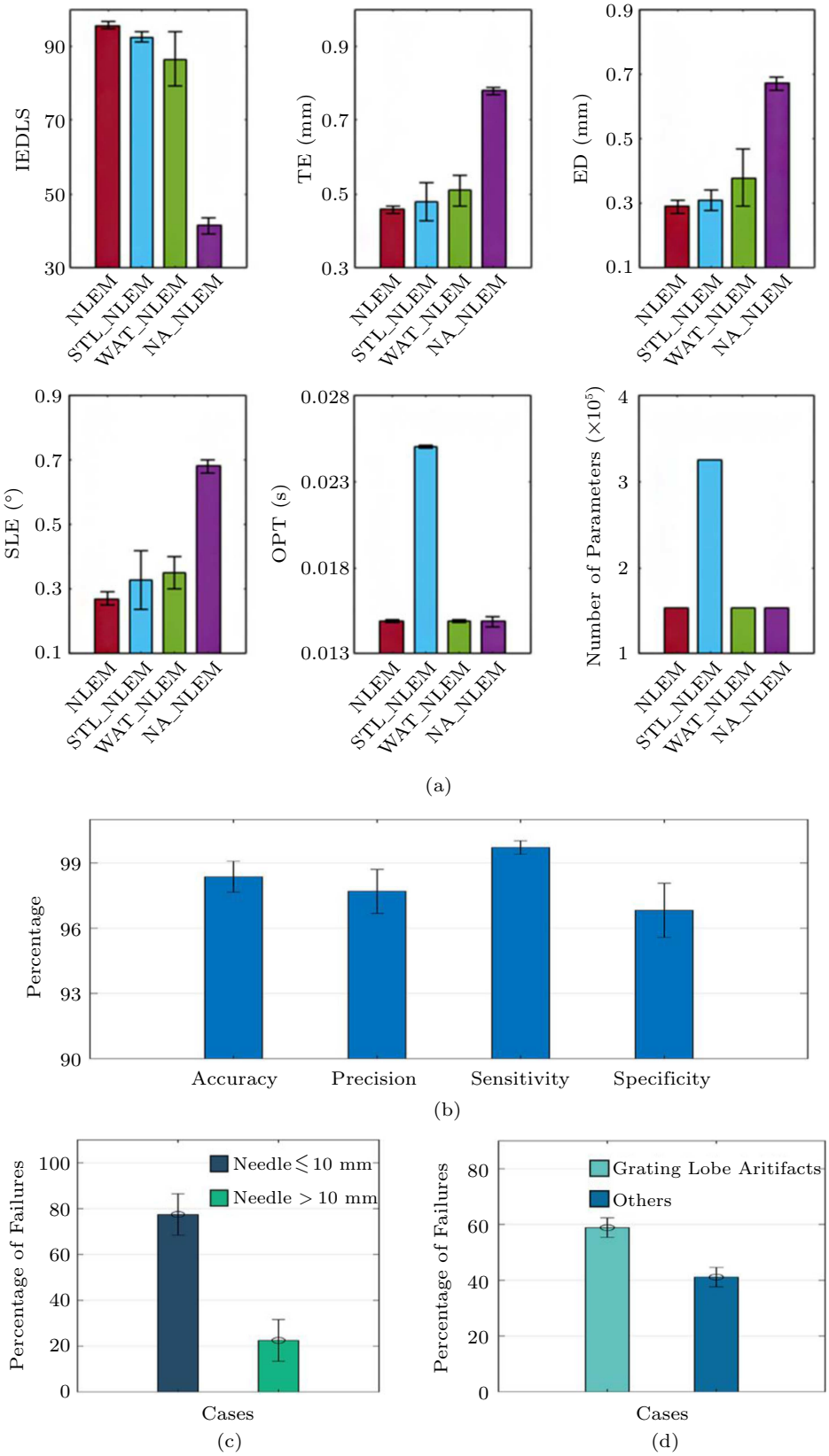

Fig.5. Performance comparison of related methods. (a) Comparing localization accuracy and overall processing time of the proposed NLEM, NA_NLEM, WAT_NLEM and STL_NLEM. (b) Performance of classification network of the proposed NLEM. (c) Failure cases analysis for needle classification. (d) Failure cases analysis for needle localization.

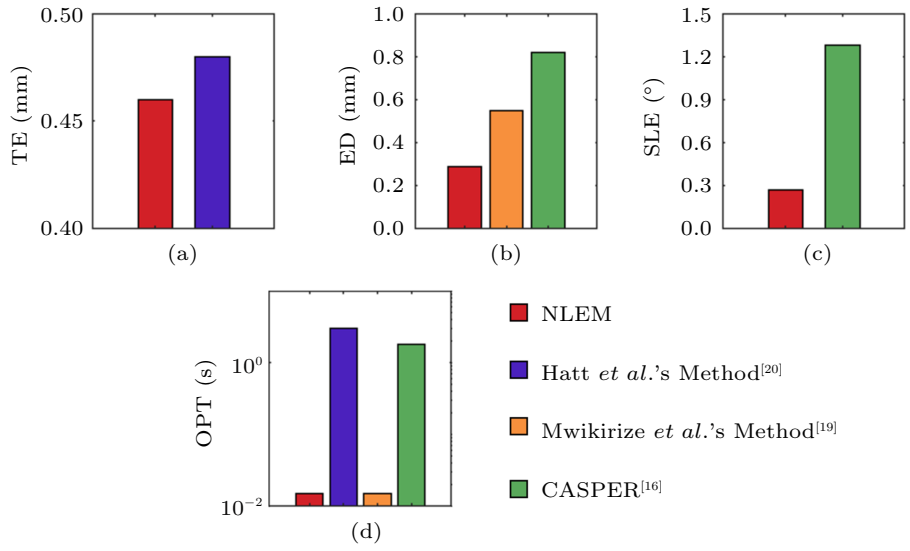

Fig.6. Performance comparison of the proposed NLEM and state-of-the-art methods. (a) TE. (b) ED. (c) SLE. (d) OPT.

extract the needle from real scanned SI images instead of low-quality, normal-angled images (NI). Moreover, Fig.4 turns out that the segmentation network (S) of the MTL framework is well trained.

We set the NA_NLEM method as the control. Fig.5(a) and Table A2 in the supplementary materials of [22] show the comparison among these methods, which are consistent in hardware, metrics and experimental datasets. As indicated in Fig.5(a) and Table A2 in the supplementary materials of [22], the proposed NLEM method outperforms the NA_NLEM, WAT_NLEM and STL_NLEM methods in the location performance, which is mainly attributed to the function of both the beam steering and the MTL framework (Fig.1).

In step 1 of NLEM, the steer angle (SA) of the ultrasound front-end is almost always perpendicular to the needle and thus the reflected signal of the needle is received by the ultrasound probe to the greatest extent (Fig.3(b) and Fig.3(h)). The one tailed paired t-test [31–36] determines that the beam steering mechanism has significant impact on the performance of the NLEM method, which greatly increases the localization accuracy (ED, SLE, TE, 1EDLS) compared with that of the NA_NLEM method (Fig.5(a)).

In step 2 of NLEM, the proposed NLEM employs a lightweight MTL network (153.8k parameters) to accurately segment the needle shaft region, detect the needle tip landmark location and determine whether an image frame has the needle, which not only significantly reduces the overall processing time but also makes NLEM meet the real-time requirements for clinical interventional applications. Moreover, Fig.5(a) and Table A2 in the supplementary materials of [22] show that the localization performance for NLEM is better than that of STL_NLEM and WAT_NLEM, because NLEM employs adversarial training [26] to effectively reduce artifact noise in the segmentation and employs multi-task learning (MTL) to decease processing time. Hence, we consider that the subsequent fusion framework (step 3 of Fig.1) can extract the needle more precisely.

To the best of our knowledge, current needle localization methods do not have a needle classification network. However, we consider that using a needle classification network could enhance the performance of the system. For this reason, we share the down-sampling layers of the segmentation model (S) and implement a classification model (C) by increasing only a few parameters (Fig.2). Fig.5(b) and Table A3 in the supplementary materials of [22] demonstrate that our classification model (C) can not only determine whether the needle is in the SI images, but also determine whether it significantly increases the computing efficiency (OPT of Fig.5(a)).

Fig.5(c) shows that approximately 77.5% of the classification failure cases occur when the insertion length for the needle is less than 10 mm, which implies that the classification model (C) has difficulty in processing samples that have a short length for the needle to insert into. We explain this phenomenon as follows: since the short length of the sample where the needle is inserted has fewer needle features than those of the long length sample where the needle is inserted, these samples are more vulnerable to background noise.

Fig.5(d) determines that approximately 58.9% of the cases of needle tip localization failure (ED $>$ 1 mm) occur when the needle tip is occluded by grading lobe artifacts. We explain these findings as follows. Since the proposed NLEM method is based on beam steering (step 1 of Fig.1), it is prone to interference from ultrasound grating lobe artifacts. Such acoustic artifacts occur when the steered beams reflect from other strong tissue reflectors, such as muscle fascia, tendons, and bone interfaces, at large incident angles with respect to the steered beam [37, 38]. When an artifact happens to occlude the tip area, the tip localization accuracy of the NLEM method is reduced.

Fig.6 demonstrates that the proposed NLEM method performs better than other state-of-the-art methods [16, 19, 20] in localization accuracy, and its performance is very close to that of Mwikirize *et al.*'s method [19] in terms of its overall processing time. However, Mwikirize *et al.*'s method [19] can only locate the needle tip, and cannot locate the needle shaft, which greatly limits its visualization capacity in many interventional applications.

5 Conclusions

This study proposed a robust needle localization and enhancement algorithm (NLEM) that can increase the needle localization accuracy, reduce the processing time, and enhance the needle visualization. The experimental results showed that the tip localization errors for the NLEM method are approximately 0.29±0.02 mm and the overall processing time for NLEM is 0.014 9 s per frame, which meets the requests of real-time clinical intervention applications. Although this study already makes great progress in localization accuracy and overall processing time, our further work will focus on how to reduce the impact of grating lobe artifacts on the localization accuracy by ultrasound signal processing. Additionally, we plan to evaluate the NLEM method by using vivo ultrasound data in the distant future.

References

[1] Gulsen F, Kantarci F. Ultrasound-guided intervention around the hip joint. *American Journal of Roentgenology*, 2012, 198(1): W95. DOI: 10.2214/AJR.11.7627.

[2] Chin K J, Perlas A, Chan V W S, Brull R. Needle visualization in ultrasound-guided regional anesthesia: Challenges and solutions. *Regional Anesthesia and Pain Medicine*, 2008, 33(6): 532-544. DOI: 10.1016/j.rapm.2008.06.002.

[3] Fevre M C, Vincent C, Picard J, Vighetti A, Chapuis C, Detavernier M, Allenet B, Payen J F, Bosson J L, Albaladejo P. Reduced variability and execution time to reach a target with a needle GPS system: Comparison between physicians, residents and nurse anaesthetists. *Anaesthesia Critical Care & Pain Medicine*, 2018, 37(1): 55-60. DOI: 10.1016/j.accpm.2016.05.008.

[4] Stolka P J, Foroughi P, Rendina M, Weiss C R, Hager G D, Boctor E M. Needle guidance using handheld stereo vision and projection for ultrasound-based interventions. In *Proc. the 17th International Conference on Medical Image Computing and Computer-Assisted Intervention*, September 2014, pp.684-691. DOI: 10.1007/978-3-319-10470-6_85.

[5] Lu H, Li J, Lu Q, Bharat S, Erkamp R, Chen B, Drysdale J, Vignon F, Jain A. A new sensor technology for 2D ultrasound-guided needle tracking. In *Proc. the 17th International Conference on Medical Image Computing and Computer-Assisted Intervention*, September 2014, pp.389-396. DOI: 10.1007/978-3-319-10470-6_49.

[6] Xia W, West S J, Finlay M C, Mari J M, Ourselin S, David A L, Desjardins A E. Looking beyond the imaging plane: 3D needle tracking with a linear array ultrasound probe. *Sci. Rep.*, 2017, 7(1): Article No. 3674. DOI: 10.1038/s41598-017-03886-4.

[7] Ding M, Fenster A. Projection-based needle segmentation in 3D ultrasound images. In *Proc. the 6th International Conference on Medical Image Computing and Computer-Assisted Intervention*, November 2003, pp.319-327. DOI: 10.3109/10929080500079321.

[8] Qiu W, Yuchi M, Ding M. Phase grouping-based needle segmentation in 3-D trans-rectal ultrasound-guided prostate trans-perineal therapy. *Ultrasound in Medicine & Biology*, 2014, 40(4): 804-816. DOI: 10.1016/j.ultrasmedbio.2013.11.004.

[9] Beigi P, Rohling R, Salcudean T, Lessoway V A, Ng G C. Needle trajectory and tip localization in real-time 3-D ultrasound using a moving stylus. *Ultrasound in Medicine & Biology*, 2015, 41(7): 2057-2070. DOI: 10.1016/j.ultrasmedbio.2015.03.013.

[10] Zhao Y, Shen Y, Bernard A, Cachard C, Liebgott H. Evaluation and comparison of current biopsy needle localization and tracking methods using 3D ultrasound. *Ultrasonics*, 2017, 73: 206-220. DOI: 10.1016/j.ultras.2016.09.006.

[11] Ayvali E, Desai J P. Optical flow-based tracking of needles and needle-tip localization using circular hough transform in ultrasound images. *Annals of Biomedical Engineering*, 2015, 43(8): 1828-1840. DOI: 10.1007/s10439-014-1208-0.

[12] Uhercik M, Kybic J, Liebgott H, Cachard C. Model fitting using RANSAC for surgical tool localization in 3D ultrasound images. *IEEE Transactions on Biomedical Engineering*, 2010, 57(8): 1907-1916. DOI: 10.1109/TBME.2010.2046416.

[13] Zhao Y, Cachard C, Liebgott H. Automatic needle detection and tracking in 3D ultrasound using an ROI-based RANSAC and Kalman method. *Ultrasonic Imaging*, 2013, 35(4): 283-306. DOI: 10.1177/0161734613502004.

[14] Mwikirize C, Nosher J L, Hacihaliloglu I. Learning needle tip localization from digital subtraction in 2D ultrasound. *International Journal of Computer Assisted Radiology and Surgery*, 2019, 14(6): 1017-1026. DOI: 10.1007/s11548-019-01951-z.

[15] Mwikirize C, Nosher J L, Hacihaliloglu I. Convolution neural networks for real-time needle detection and localization in 2D ultrasound. *International Journal of Computer Assisted Radiology and Surgery*, 2018, 13(5): 647-657. DOI: 10.1007/s11548-018-1721-y.

[16] Beigi P, Rohling R, Salcudean S E, Ng G C. CASPER: Computer-aided segmentation of imperceptible motion—A learning-based tracking of an invisible needle in ultrasound. *International Journal of Computer Assisted Radiology and Surgery*, 2017, 12(11): 1857-1866. DOI: 10.1007/s11548-017-1631-4.

[17] Pourtaherian A, Ghazvinian Zanjani F, Zinger S, Mihajlovic N, Ng G C, Korsten H H M, De With P H N. Robust and semantic needle detection in 3D ultrasound using orthogonal-plane convolutional neural networks. *International Journal of Computer Assisted Radiology and Surgery*, 2018, 13(9): 1321-1333. DOI: 10.1007/s11548-018-1798-3.

[18] Arif M, Moelker A, Van Walsum T. Automatic needle detection and real-time Bi-planar needle visualization during 3D ultrasound scanning of the liver. *Medical Image Analysis*, 2019, 53: 104-110. DOI: 10.1016/j.media.2019.02.002.

[19] Mwikirize C, Nosher J L, Hacihaliloglu I. Single shot needle tip localization in 2D ultrasound. In *Proc. the 22nd International Conference on Medical Image Computing and Computer Assisted Intervention*, October 2019, pp.637-645. DOI: 10.1007/978-3-030-32254-0_71.

[20] Hatt C R, Ng G, Parthasarathy V. Enhanced needle localization in ultrasound using beam steering and learning-based segmentation. *Computerized Medical Imaging and Graphics*, 2015, 41: 46-54. DOI: 10.1016/j.compmedimag.2014.06.016.

[21] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. In *Proc. the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention*, October 2015, pp. 234-241. DOI: 10.1007/978-3-319-24574-4_28.

[22] Gao J, Liu P, Liu G D, Zhang L. Supplementary material. Technical Report, Sichuan University, 2020. https://github.com/gaojun0821/NLEM/blob/main/doc/-Supplementary_Material.pdf, Dec. 2020.

[23] Jiang B, Struthers A, Sun Z, Feng Z, Zhao X, Zhao K, Dai W, Zhou X, Berens M E, Zhang L. Employing graphics processing unit technology, alternating direction implicit method and domain decomposition to speed up the numerical diffusion solver for the biomedical engineering research. *International Journal for Numerical Methods in Biomedical Engineering*, 2011, 27(11): 1829-1849. DOI: 10.1002/cnm.1444.

[24] Jiang B N, Dai W Z, Khaliq A, Carey M, Zhou X B, Zhang L. Novel 3D GPU based numerical parallel diffusion algorithms in cylindrical coordinates for health care simulation. *Mathematics and Computers in Simulation*, 2015, 109: 1-19. DOI: 10.1016/j.matcom.2014.07.003.

[25] Zhang L, Jiang B, Wu Y, Strouthos C, Sun P Z, Su J, Zhou X. Developing a multiscale, multi-resolution agent-based brain tumor model by graphics processing units. *Theoretical Biology and Medical Modelling*, 2011, 8(1): Article No. 46. DOI: 10.1186/1742-4682-8-46.

[26] Isola P, Zhu J Y, Zhou T, Efros A. Image-to-image translation with conditional adversarial networks. In *Proc. the 2017 IEEE Conference on Computer Vision and Pattern Recognition*, July 2017, pp.5967-5976. DOI: 10.1109/CVPR.2017.632.

[27] Wu W, Song L, Yang Y, Wang J, Liu H, Zhang L. Exploring the dynamics and interplay of human papillomavirus and cervical tumorigenesis by integrating biological data into a mathematical model. *BMC Bioinformatics*, 2020, 21(Suppl 7): Article No. 152. DOI: 10.1186/s12859-020-3454-5.

[28] Ren S, He K, Girshick R, Sun J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(6): 1137-1149. DOI: 10.1109/TPAMI.2016.2577031.

[29] Xu Z, Huo Y, Park J, Landman B, Milkowski A, Grbic S, Zhou S. Less is more: Simultaneous view classification and landmark detection for abdominal ultrasound images. In *Proc. the 21st International Conference on Medical Image Computing and Computer-Assisted Intervention*, September 2018, pp.711-719. DOI: 10.1007/978-3-030-00934-2_79.

[30] Powers D. Evaluation: From precision, recall and F-factor to ROC, informedness, markedness & correlation. arXiv:2010.16061, 2008. https://arxiv.org/abs/2010.16061, Jan. 2021.

[31] Xia Y, Yang C, Hu N, Yang Z, He X, Li T, Zhang L. Exploring the key genes and signaling transduction pathways related to the survival time of glioblastoma multiforme patients by a novel survival analysis model. *BMC Genomics*, 2017, 18(Suppl 1): Article No. 950. DOI: 10.1186/s12864-016-3256-3.

[32] Zhang L, Bai W, Yuan N, Du Z. Comprehensively benchmarking applications for detecting copy number variation. *PLoS Computational Biology*, 2019, 15(5): Article No. e1007069. DOI: 10.1371/journal.pcbi.1007069.

[33] Zhang L, Li J, Yin K, Jiang Z, Li T, Hu R, Yu Z, Feng H, Chen Y. Computed tomography angiography-based analysis of high-risk intracerebral haemorrhage patients by employing a mathematical model. *BMC Bioinformatics*, 2019, 20(Suppl 7): Article No. 193. DOI: 10.1186/s12859-019-2741-5.

[34] Zhang L, Liu Y, Wang M, Wu Z, Li N, Zhang J, Yang C. EZH2-, CHD4-, and IDH-linked epigenetic perturbation and its association with survival in glioma patients. *J. Mol. Cell Biol.*, 2017, 9(6): 477-488. DOI: 10.1093/jmcb/mjx056.

[35] Kremkau F W, Taylor K J. Artifacts in ultrasound imaging. *Journal of Ultrasound in Medicine*, 1986, 5(4): 227-237. DOI: 10.7863/jum.1986.5.4.227.

[36] Paul Y, Barthez D, Léveillé R, Peter V, Scrivani D. Side lobes and grating lobes artifacts in ultrasound imaging. *Veterinary Radiology & Ultrasound*, 1997, 38(5): 387-393. DOI: 10.1111/j.1740-8261.1997.tb02104.x.

[37] Matalon T A, Silver B. US guidance of interventional procedures. *Radiology*, 1990, 174(1): 43-47. DOI: 10.1148/radiology.174.1.2403684.

[38] Sofka C, Collins A J, Adler R. Use of ultrasonographic guidance in interventional musculoskeletal procedures: A review from a single institution. *Journal of Ultrasound*

in Medicine: Official Journal of the American Institute of Ultrasound in Medicine, 2001, 20(1): 21-26. DOI: 10.7863/jum.2001.20.1.21.

Jun Gao received his B.S. and M.S. degrees in computer science from Sichuan University, Chengdu, in 2005 and 2008 respectively. Currently, he is a Ph.D. candidate at the College of Computer Science, Sichuan University, Chengdu. His research interests involve medical image processing, computer vision, and machine learning.

Paul Liu received his B.S. degrees in electrical engineering and mathematics from the University of Washington, Seattle, in 2003, and his M.S. degree in electrical engineering from the University of California, Berkeley, in 2006. He earned his Ph.D. degree in computer science from Sichuan University, Chengdu, in 2014. Currently he is a co-founder and principal engineer at Stork Healthcare, a handheld ultrasound and ultrasound image analysis company. His research interests involve medical ultrasound imaging, signal processing, and machine learning.

Guang-Di Liu received his M.S. degree in computer science from Xihua University, Chengdu, in 2014, and his Ph.D. degree in mathematics from Southwest University, Chongqing, in 2020. His research interests involve bioinformatics, artificial intelligence, and data mining.

Le Zhang received his B.S. degree in computer science from Beijing Institute of Technology, Beijing, in 1999, and his M.S. and Ph.D. degrees in computer science, and computational analysis and modeling respectively from Louisiana Tech University in 2005, and the postdoctoral training from 2005 to 2008 in Harvard Medical School. Now he is a full professor at the College of Computer Science, Sichuan University, Chengdu. His research interests involve bioinformatics, computational biology, artificial intelligence, and high-performance computing.