Slice Localization for Three-Dimensional Breast Ultrasound Volume Using Deep Learning

Tai-feng YU\textsuperscript{1}, Paul LIU\textsuperscript{2}, Yu-lan PENG\textsuperscript{3,}\textsuperscript{*}, Jing-yan LIU\textsuperscript{3}, Hao YIN\textsuperscript{1} and Dong C LIU\textsuperscript{1}

\textsuperscript{1}Sichuan University, College of Computer Science, Yihuan Road, Chengdu, China
\textsuperscript{2}Stork Healthcare Co., Ltd, Chuangye Road, Hi-Tech Zone, Chengdu, China
\textsuperscript{3}West China Hospital, Department of Ultrasound, Wuhou District, Chengdu, China

\textsuperscript{*}Corresponding author

\textbf{Keywords:} Localization, Automated breast ultrasound system, Robotic arms, Fully connected neural network.

\section*{Abstract.} In routine ultrasonography, slice images of the internal organs of the human body are usually generated through an 1D array probe. The position and orientation of the probe is adjusted manually to obtain slice planes with pathological features. This is quite dependent on the experience and technique of a sonographer. This paper aims to locate 2D slice planes in a 3D breast ultrasound volume, which has significant application value in clinical ultrasound examinations. We propose a deep learning approach mapping all possible 2D image slices to their 3D coordinates parameters using a fully connected neural network implemented on MATLAB. We emphasize that this training must be done separately for each patient since the mammary tissue structure varies greatly from one person to another. The trained network can be interpreted as an image-slice location database for each patient. Our study is validated on GE ABUS (Automated Breast Ultrasound System) volume data. Each 2D image slice has four spatial parameters. The method achieves a prediction error of 0.14mm/0.25mm and 0.5 degree/0.3 degree for translation (x/y) and rotation (yaw/roll) parameters respectively, averaged over all practically scannable slices. It takes less than 0.1ms to predict the location of one 64x64 slice image. Thus, slice locations may be displayed with high accuracy in real-time when scanning with a conventional 1D probe, potentially allowing physicians to manipulate the probe to any scan planes of interest.

\section*{Introduction} Ultrasound (US) imaging is a medical diagnostic technology widely used in many areas of healthcare for years. Compared to other forms of imaging techniques, such as CT and MRI, US imaging is portable, noninvasive, radiation-free, and low cost \cite{1}. However, 2D conventional US scans the body in thin slices, and the images obtained in an US exam are highly dependent on the skill and expertise of the operators \cite{2}. It is difficult to adjust the probe to present the image at specific location in a potentially large anatomical volume \cite{3}. Therefore, it may be very beneficial to compute the slice locations from volume data when scanning. This will guide physicians to find slices of interest and reduce the exam time or repetitive operations.

Various techniques have been studied in 3D US imaging \cite{3}, including using a robotic arm to move the US probe. Multiple US diagnostic products equipped with robotic arms have been developed over the past decades. The KUKA LBR Med is a versatile robot for assisting tasks in the healthcare sector \cite{4}. By attaching an US probe to the robot, the probe can be moved with precise motion over the body to get the images. The MELODY System is a remote robotic ultrasound solution \cite{5}. The expert can maneuver the ultrasound probe on the patient’s body via the MELODY robotic arm in real-time. MELODY also has a universal interface for mounting most US probes available on the market. The GE ABUS is an innovative breast cancer screening technology \cite{6}. A highly standardized volume data can be obtained by the system scanning the entire breast with 15cm wide field-of-view.
high-frequency transducer. The transducer is driven by a mechanical arm moving with constant speed. Due to the soft nature of the breast tissue, it will deform substantially during scanning. This may cause tissue mis-registration in spatial compounding of multiple scan planes. The ABUS can overcome the problem above and it is currently the most accurate 3D US breast volume data without relying on the sonographers’ technique.

Our study aims to predict the locations of slices in breast US volume data. The slice locations signify the translation and rotation of the 1D US probe. Recently, deep learning algorithms have rapidly become a popular methodology in medical image domains [7]. Several works have applied deep learning techniques to locate slices in volumes. Soufiane et al. [8] use deep convolution network to spotting a particular slice (L3 slice) in a 3D spine CT scan using transfer learning. Ryou et al. [9] adopt transfer learning to localize the fetus and extract the best biometry planes for both the fetal head and abdomen in the axial plane. In their studies [8,9], sampled slices are all along the same axis. Li et al. [10] propose the Iterative Transformation Network to detect standard planes in 3D fetal US, and they take rotation parameters into consideration. But these methods [8,9,10] only extract specific or standard slices, while our target is locating arbitrary slices in US volume data. Hou et al. [11] propose a real-time approach that can predict rotations and translations of 2D image slices from 3D volumes, while their slice sampling methods and prediction accuracy for MRI or CT slice locations are not suitable for US volume data. The slice image at predicted location may miss crucial information such as small lesion area.

In this paper, a fully connected neural network is trained to map 2D slice images to their locations in the ABUS volume. Unlike other deep learning studies, the training process of our deep learning network is regarded as a recording process of the slice positions. Each patient has a unique image-based location network with separate weights for each patient. 2D slice images sampled from the volume data are the input to the network. Corresponding slice locations are the output, containing two translation parameters and two rotation parameters. After the network is trained, the four coordinate parameters of numerous slices in the entire volume can be predicted purely from the image intensity information. In this way, accurate slice locations can be estimated in real-time during free-hand scanning with a 1D conventional probe. This localization network can be used during a follow-up consultation. It can also serve as a training tool for inexperienced operators.

Materials and Methods

Resizing of 3D Volume Data

The study is based on ABUS data provided by the US department of West China Hospital of Sichuan University. For each patient’s individual ABUS data, we extract all the frames from the ABUS video to obtain 3D volume data. A right-hand coordinate system is built for ABUS, and let $z$ be axial, $x$ be lateral, $y$ be elevation dimensions as shown in Fig. 1(a). $Z$ is along the beam axis and $y$ is the direction in which the probe is moving.

![Figure 1.](image)

The video file has a resolution of 720×1280 pixels and 270 frames in which different frames represent different planes in the elevation direction. We consider an ROI of the original volume data...
which is 340 pixels in axial (Fig. 1(b)). We resize the volume by reducing the size of the lateral by one-half with Lanczos-2 kernel. The purpose of the resizing is to reduce the subsequent 2D slice plane extraction complexity by lowering the image size to the point where we cannot go lower due to resolution concerns. A pixel size 340×640×270 volume data is generated after resizing. The three-dimensional matrix can be understood as continuously storing 270 frames z-x plane grayscale images with a resolution of 340×640 pixels. The actual physical spacing information between voxels is extracted from the video. Units (cm/pixel) of z, x, y is about 0.0122, 0.0246, 0.0644 respectively. The size of the physical volume is about 4.148cm ×15.744cm×17.388cm.

Preparing Slice Data

As shown in Fig. 2, the motion of an object has six degrees of freedom in three-dimensional space. However, we only consider x, y, yaw and roll here, because a pitch rotation or a movement in the z direction will cause the 1D probe to leave the surface of the breast which is not allowed during the scanning.

![Figure 2](image)

Figure 2. The six degrees of freedom can be described as a sequence of x, y, z, yaw, pitch, and roll.

Based on above, multiple 2D slices are created in the four coordinate parameters (Yaw, Roll, X, Y) within a sub-region of the total volume data. X and Y are translation parameters, which present the left top of the slice plane (Fig. 3(a)); Yaw and Roll are rotation parameters (Fig. 3(b)(c)).

![Figure 3](image)

Figure 3. (a) Create one slice plane with four coordinate parameters. (b) Yaw (c) Roll.

Table 1 shows the ranges of four coordinate parameters, which ensures that the sampling of planes will not fall outside of the ABUS volume data. A total of 22,982,656 (181×31×64×64) slice planes are created. Each slice plane covers an area of 3.8cm by 3.8cm, having a grid that holds 64×64 pixels after downsampling. The choice of lateral 3.8cm is based on the size of a standard non-ABUS linear probe, which will represent conventional scanning.

<table>
<thead>
<tr>
<th>Coordinate parameters range</th>
<th>Yaw [degree]</th>
<th>Roll [degree]</th>
<th>X [cm]</th>
<th>Y [cm]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lower Bound</td>
<td>-90</td>
<td>-30</td>
<td>-3.1</td>
<td>-3.1</td>
</tr>
<tr>
<td>Upper Bound</td>
<td>90</td>
<td>30</td>
<td>3.2</td>
<td>3.2</td>
</tr>
<tr>
<td>Sampling Interval</td>
<td>1</td>
<td>2</td>
<td>0.1</td>
<td>0.1</td>
</tr>
</tbody>
</table>

The two rotation matrices for coordinate system in z, x, y (Fig.4) are given by Eq. 1 and Eq. 2, where α=Roll, β=Yaw.

\[
R_{\alpha,\beta}(\alpha) = \begin{pmatrix}
\cos \alpha & 0 & \sin \alpha \\
0 & 1 & 0 \\
-\sin \alpha & 0 & \cos \alpha
\end{pmatrix}
\]  

(1)
A single rotation matrix can be formed by multiplying the yaw and roll rotation matrices [12]. To take rotation into account, each sample point is further multiplied by the rotation matrix 

\[ R(\beta, \alpha) = R_{\text{yaw}}(\beta)R_{\text{roll}}(\alpha) \]

to obtain their final physical locations, then divided by units in z, x, y to get pixel locations. The grayscale value of each point is calculated by trilinear interpolation, which is determined by eight corners of a cube that surrounds it.

\[
R_{\text{yaw}}(\beta) = \begin{pmatrix}
1 & 0 & 0 \\
0 & \cos \beta & -\sin \beta \\
0 & \sin \beta & \cos \beta
\end{pmatrix}
\]

Before training the network, the coordinate parameters and grayscale value of images should be normalized first. Normalization will stabilize and speed up network training. The grayscale value of 64x64 pixels on each slice plane and four corresponding coordinate parameters (Yaw, Roll, X, Y) are saved. These data are the training data set used to train the network later.

**Network Model**

The neural network is a nonlinear statistical data modelling tool which has a robust fitting ability. It is suitable for predicting continuous data such as angles and distances. The proposed network model fits a one-to-one relationship between the slices and coordinate parameters. Fig. 4 shows the network architecture proposed in this paper. When all the neurons in adjacent layers are connected in pairs, a fully connected neural network is formed. The essence of a fully connected neural network is to fit the nonlinear relationship between known features and targets through parameters and activation functions.

![Network Architecture](image)

The first layer defines an image input layer. The input image to the network is 64x64x1, and the output size of the image input layer is 4096. The middle layers of the network comprise seven fully connected (FC) layers, each followed by a ReLU layer (activation layer). Output size for each fully connected layer is 2048. In the end, the network comes to a fully connected layer of size 4 and a regression output layer.

A fully connected layer multiplies the input by a weight matrix w and then adds a bias vector b, vector \( x = [x_1, x_2, ..., x_n]^T \) is the input, vector \( z = [z_1, z_2, ..., z_n]^T = w \cdot x + b = \sum_{i=1}^{n} w_i \cdot x_i + b \) is the result of this linear output. w and b will be adjusted to a more appropriate value each time the model learns.

An activation layer obtain the output vector \( h = [h_1, h_2, ..., h_n]^T \) through activation function Eq. 3. Rectified Linear Unit (ReLU) is a very common activation function; it performs threshold operations on each element in which input value less than zero is regard as zero, and otherwise remain unchanged.

\[
f(x) = \max(0, x) = \begin{cases} 
x, & x \geq 0 \\
0, & x < 0
\end{cases}
\]
The primary role of the activation function is to add nonlinear factors to solve the shortcomings of the linear model’s fitting ability, which is vital to the entire neural network. This transformation can retain the data features to the maximum extent. It is easy to calculate and has less effect on vanishing gradient in deep models. A regression layer follows the last fully connected layer, because establishing a mapping relationship from 2D images to 3D location parameters is a regression problem. The loss function of the regression layer is half-mean-squared-error.

The purpose of the architecture is to minimize the value of the loss function and update parameters. We choose stochastic gradient descent with momentum (SGDM) algorithm [13] as our optimizer. Based on SGD, SGDM optimizer introduces a first-order momentum, which is the exponential average moving of the gradients. The direction of the descent at a particular time is determined not only by the current gradient direction but also by the course of the descent accumulated before. SGDM has faster convergence and higher accuracy, which reduces oscillation during convergence. In this model, the value of momentum term is 0.9.

Experiment

The proposed method was implemented with MATLAB running on a machine equipped with Intel Core i7-8700K CPU at 3.70GHz and NVIDIA GeForce GTX1050Ti 4GB GPU. Deep Learning Toolbox in MATLAB provides a framework for designing and implementing deep neural networks [14], and it is utilized to build the learning model quickly. We create 22,982,656 slices from the ABUS video of one patient having breast nodules. For so many slices in the data set, 10,000 slice data files are more than enough to estimate the performance of the model. So we just randomly take 10,000 slices for testing, and all remaining data files are used for training. Each data file consists of two parts: the grayscale value of 64x64 pixels on each slice plane and a vector comprising of four coordinate parameters: Yaw, Roll, X, Y, which is known as it is manually decided.

The entire training process lasted about 105 hours for 32 epochs, and the training data are shuffled before each training epoch. Learning rate is 1e-2 in the initial training stage and 1e-3 in the fine-tuning training stage. The data are divided into groups, and each group of data is looped into the network to train by using a batch size of 128.

After the network is trained, we verify its performance on the test set. Input a slice image, and the network will predict a location, then compare it with the actual location. Utilize RMSE (root mean square error) to measure the differences between locations predicted by the network and the actual location.

Result

As shown in Table 2, RMSE of testing data in lateral dimension is about 0.01 cm, in elevation dimension is about 0.02 cm as elevation direction has a lower acoustic resolution, both achieving sub-pixel accuracy. For yaw and roll, the RMSE is less than 1 degree, which is sensitive enough for the rotation of the probe. It takes an average of 0.1 milliseconds to predict the location of one input image, which meets real-time requirements. The outliers are defined as values more than three scaled median absolute deviations away from the data’s median.

<table>
<thead>
<tr>
<th></th>
<th>Yaw [degree]</th>
<th>Roll [degree]</th>
<th>X [cm]</th>
<th>Y [cm]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of outliers</td>
<td>100</td>
<td>104</td>
<td>183</td>
<td>97</td>
</tr>
<tr>
<td>RMSE of the entire test set</td>
<td>0.5276</td>
<td>0.2768</td>
<td>0.0137</td>
<td>0.0249</td>
</tr>
<tr>
<td>RMSE without outliers</td>
<td>0.4989</td>
<td>0.2626</td>
<td>0.0121</td>
<td>0.0236</td>
</tr>
</tbody>
</table>

According to the test result of 10,000 samples in Table 2, though there are outliers in each dimension, no sample has outliers in all four dimensions, only one sample has outliers in all three dimensions. 31 samples have outliers in two dimensions, and 419 samples have outliers in one dimension.
Table 3 shows GT and predicted slices parameters of six pairs of samples. Fig. 5(a)~(c) are samples of non-outliers. (a)~(c) shows the slice images in non-outlier samples extracted from the initial volume data using these parameters, the images at predicted locations are almost identical to the GT images. (d) is a sample has outliers in three dimensions, (e) are samples have outliers in two dimensions, (f) are samples have outliers in one dimension, bold values indicate outliers. Fig. 5(d)~(f) shows there is not much difference between actual and predicted slice images, and they have similar textures such as tissue layers and hypoechoic structures. Slice images determined by outlier parameters are within acceptable tolerances.

Table 3. GT and predicted slices parameters of samples.

<table>
<thead>
<tr>
<th></th>
<th>GT</th>
<th>Predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>x</td>
<td>y</td>
</tr>
<tr>
<td>Non-outlier Samples</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>2.8</td>
<td>2.2</td>
</tr>
<tr>
<td>b</td>
<td>2.4</td>
<td>-0.3</td>
</tr>
<tr>
<td>c</td>
<td>-2.9</td>
<td>0.8</td>
</tr>
<tr>
<td>Outlier Samples</td>
<td></td>
<td></td>
</tr>
<tr>
<td>d</td>
<td>-0.2</td>
<td>-0.5</td>
</tr>
<tr>
<td>e</td>
<td>-0.2</td>
<td>-0.2</td>
</tr>
<tr>
<td>f</td>
<td>-2.4</td>
<td>-1.1</td>
</tr>
</tbody>
</table>

Discussion and Conclusions

We have presented a fully connected network to locate 2D slices in 3D ABUS volume data. By inputting slice images to the network, translation and rotation parameters can be estimated in real-time. A potential application is precise guidance to the scan slice planes of interest. The US physicians will know where and how to move the probe interactively according to the prompt of the output location of the ongoing scan slice. Since no two patients are identical, the network should be trained every time a new patient comes. After that, this trained network can be seen as the patient’s image-based “patient record” useful for follow-up scanning using traditional non-ABUS ultrasound devices with similar image resolutions. This avoids having to repeat another ABUS scan, which is time consuming and expensive.

In further study, the dedicated breast exam ABUS platform can be generalized to other platform, such as volume data reconstructed with wide-field scanning using robotics arm equipped with conventional 3D probes. Moreover, this proposed method may also be extended to automatically find the slice of interest in 3D medical data of other body parts.

Acknowledgements

This work is supported by the New Industrial Technology Research Institute, Technological Innovation Project of Chengdu Science and Technology Bureau (Breast ultrasound artificial intelligence technology and product development, No.2017-CY02-00027-GX). Our images are provided by the Department of Ultrasound, West China Hospital.
References


