RESEARCH ARTICLE

Brachial Plexus Nerve Trunk Recognition From Ultrasound Images: A Comparative Study of Deep Learning Models

DINGCHENG TIAN[1], **BINBIN ZHU**[2], **JIANLIN WANG**[3], **LINGSI KONG**[3], **BIN GAO**[2], **YU WANG**[4], **DECHAO XU**[1], **RUYI ZHANG**[1], AND **YUDONG YAO**[1], (Fellow, IEEE)

[1]Research Institute for Medical and Biological Engineering, Ningbo University, Ningbo 315211, China
[2]Department of Anesthesiology, The Affiliated Hospital of Medical School, Ningbo University, Ningbo 315020, China
[3]Department of Anesthesiology, Ningbo No.6 Hospital, Ningbo 315040, China
[4]College of Medicine and Biological Information Engineering, Northeastern University, Shenyang 110819, China

Corresponding author: Yudong Yao (yaoyudong@nbu.edu.cn)

ABSTRACT Brachial plexus block is a common regional anesthesia method widely used in upper limb surgery. Nowadays, ultrasound-guided brachial plexus block has been extensively used in clinical anesthesia. However, accurate brachial plexus block is highly dependent on the physician's experience, and a physician without extensive clinical experience may cause nerve injury when performing a nerve block. With the development of artificial intelligence technology, the deep learning method can automatically identify the brachial plexus in ultrasound images and assist doctors in completing the brachial plexus block accurately and quickly. In this paper, we aim to evaluate the performance of different deep learning models in identifying brachial plexus (i.e., segmentation of brachial plexus) from ultrasonic images to explore the best models and training strategies for this task. To this end, we use a new dataset containing 340 brachial plexus ultrasound images annotated by three experienced clinicians. Among the 12 deep learning models we evaluated, U-Net achieves the best segmentation accuracy, with an intersection over union (IoU) of 68.50%. However, the number of U-Net parameters is very large, and it can only process 15 images per second. Compared to U-Net, LinkNet can process 142 images per second and achieve the second-best segmentation accuracy with an IoU of 66.27%. It achieves the balance between segmentation accuracy and processing efficiency, which has a good potential for the brachial plexus's real-time segmentation task.

INDEX TERMS Deep learning, brachial plexus, nerve segmentation, nerve block.

I. INTRODUCTION

Brachial plexus block is a common regional nerve block method, which has been widely used in upper limb surgery. Compared with general anesthesia, brachial plexus block anesthesia has minimum negative impact on the patient's physiology, fewer complications, and is more effective. In clinical practice, medical staff has developed nerve block guidance techniques such as paraesthesia, nerve stimulation, and ultrasound guidance to enhance the success rate of

The associate editor coordinating the review of this manuscript and approving it for publication was Chulhong Kim.

regional anesthesia. Compared with other techniques, ultrasound guidance provides more intuitive and accurate nerve localization. It provides real-time images when blocking the target nerve, thus assisting the physician in achieving precise anesthesia [1]. However, ultrasound-guided brachial plexus blocks still face several challenges. Firstly, ultrasound images have many artifacts and noise, low contrast between tissues, and fuzzy boundaries [2], making it challenging to identify nerve block areas in clinical practice. Secondly, the use of ultrasound-guided localization techniques requires anesthesiologists who not only have a deep understanding of anatomical structures but also have extensive experience in

the use of ultrasound imaging [3]. Anesthesiologists with extensive clinical experience are in short supply, and doctors with insufficient clinical experience may cause nerve injury when performing nerve blocks. Thirdly, for certain patients, such as those with obesity, edema, and muscle atrophy, the quality of ultrasound imaging is poorer, and the area of nerve block can be difficult to identify [4]. Based on the discussions above, it can be seen that accurately locating nerves in ultrasound images has become an important issue in the development of nerve blocks. To overcome the above problems, an effective approach is to introduce a computer-aided diagnosis (CAD) system, which can be used to supplement doctors' personal experience and knowledge to improve the reliability and efficiency of the brachial plexus block.

In recent years, the progress of CAD and its great potential in medical diagnosis have received extensive attention [5]. Researchers used digital image processing combined with traditional machine learning techniques to identify ultrasonic nerve block areas. González *et al.* [6] first use the graph cut method to pre-segment the ultrasound image to obtain the region of interest. Features are then extracted from the ultrasound image using a nonlinear wavelet transform. Finally, the pixels are classified using a Gaussian process classifier to obtain the neural region. Vashishtha and Aju [7] combine the canny edge detection algorithm with the support vector machine (SVM) to achieve neural segmentation in the ultrasonic image. Jimenez *et al.* [8] use random undersampling (RUS) and SVM for neural segmentation of ultrasound images. Although traditional segmentation methods can accomplish segmentation of neural block regions on smaller datasets, they all need to design feature extraction methods manually. The implementation can be complex, and it is challenging to achieve accurate segmentation results.

In contrast to machine learning, deep learning does not need to design the feature extraction process manually, and it can automatically learn from data during the training process to obtain excellent feature representation. It has succeeded in computer vision, bioinformatics, and other fields. Currently, deep learning methods have been applied to the whole process of medical image processing and analysis. They have achieved remarkable results in classification [9], detection [10], segmentation [11], reconstruction [12], and registration [13] tasks. In the research of brachial plexus segmentation, the computer-aided diagnosis system based on deep learning has achieved great results and has become the mainstream direction of related research. Wei and Tong [14] propose a dual-path U-shaped network with an attention mechanism. In this network, two paths are used to replace the ordinary convolution layer, and the attention mechanism is used to improve the efficiency and accuracy of segmentation. Van Boxtel *et al.* [15] propose a hybrid model consisting of a classification and segmentation model to segment brachial plexus regions in ultrasound images. The experimental results

show that the segmentation performance of the hybrid model is significantly improved compared to the single segmentation model. Zhao and Sun [16] propose an end-to-end method based on U-Net for automatic segmentation of the brachial plexus from ultrasound images, which achieves good segmentation performance. In addition to brachial plexus segmentation, deep learning is also used to identify nerve blocks in other body parts. Huang *et al.* [17] use the U-Net network to identify femoral nerve block regions from ultrasound images. Smistad *et al.* [18] use a convolutional neural network to segment axillary nerves. Horng *et al.* [19] combine U-Net and recurrent neural network for localization and segmentation of median nerve.

Although many studies use deep learning to identify nerves in ultrasound images, the relative performance results are not clear when using different deep learning models, training strategies, or loss functions. Researchers have made many contributions to comparative research in the field of medicine based on machine learning or deep learning [20]–[24]. However, such investigations have been missing in the research topic of identifying brachial plexus nerves in ultrasound images using deep learning methods. We attempt to contribute to this comparative study. Therefore, this paper comprehensively evaluates the performance of different deep learning models in brachial plexus ultrasound image segmentation.

It is worth noting that all current studies in brachial plexus identification are to segment a nerve block region rather than individual nerve trunks. However, compared with segmenting only a nerve block region, segmenting individual nerve trunks can assist doctors in locating the position of each brachial plexus trunk more accurately to improve the effectiveness of the brachial plexus block [25]. Therefore, in this study, we segment individual brachial plexus trunks. We use a new dataset containing 340 ultrasound images of brachial plexus and labeled by three clinicians [25]. Then we implement 12 deep learning models and comprehensively evaluate their performance from the following aspects to find the model and training strategy suitable for this nerve segmentation task. (1) To evaluate the neural segmentation performance of different models (including lightweight real-time segmentation networks), we compare the segmentation accuracy and speed. (2) We explore the role of deep transfer learning in network training. (3) We evaluate the impact of different loss functions on the model segmentation accuracy. (4) We evaluate the model's generalization performance in identifying the brachial plexus nerve in images from a new ultrasound machine.

The rest of this paper is organized as follows. Section II details the dataset, the 12 deep learning models used, the transfer learning process, and the loss function. Section III presents the experimental results and analysis, including the experimental setup and evaluation methods. In Section IV, we discuss the results. Finally, Section V concludes the paper with suggested future work.
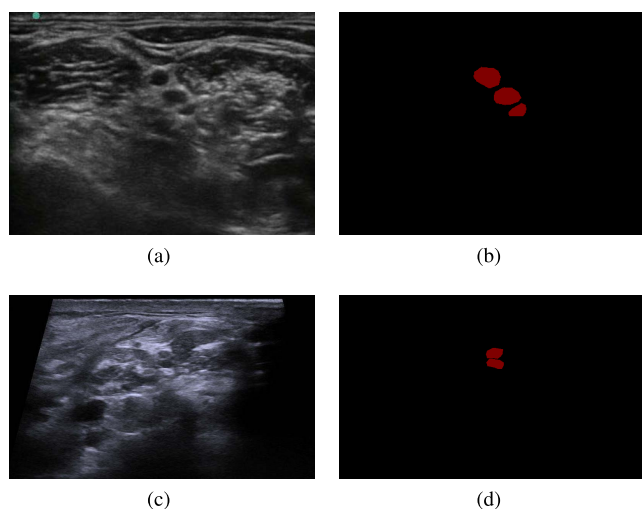
FIGURE 1. Example of brachial plexus ultrasound dataset. (a), (c) represent the original brachial plexus ultrasound images from YGY and BK3000 respectively. (b), (d) are doctor-annotated images of YGY and BK3000.

FIGURE 2. Example of femoral nerve dataset. (a) Ultrasound image of the femoral nerve. (b) The doctor annotates the image of the femoral nerve.

FIGURE 3. Example of an enhanced image using limiting contrast adaptive histogram equalization. (a) Original brachial plexus ultrasound image. (b) Enhanced ultrasound image of brachial plexus.

II. MATERIALS AND METHODS

A. DATASETS

1) DATA DESCRIPTION

To obtain brachial plexus ultrasound images, we cooperate with the Affiliated Hospital of Medicine School of Ningbo University and the Sixth Hospital of Ningbo to collect 340 brachial plexus ultrasound images [25]. To increase the heterogeneity of the data, we collect data from as many patients as possible rather than multiple images from the same patient. In addition, data is collected from two different ultrasound machines (YGY, BK3000). Our images are labeled by three professional anesthesiologists using Labelme [26]. They all have more than seven years of experience in ultrasound-guided nerve block anesthesia. An example of the brachial plexus ultrasound dataset is presented in Fig. 1. Furthermore, we directly annotate the neural trunk of the brachial plexus. To our knowledge, there is no publicly available brachial plexus ultrasound image dataset as annotated as ours. In addition, to validate the role of transfer learning in neural network model training, we use a publicly available dataset of femoral nerve ultrasound images. Huang *et al.* [17] created this publicly accessible dataset of femoral nerve ultrasound images in the GitHub repository. There are a total of 562 ultrasound images in this dataset. Since the femoral nerve is difficult to identify, they did not label it directly but annotated the connective tissue surrounded by the iliofascial membrane and the iliopsoas. This is a crucial area for identifying the femoral nerve. An example of the femoral nerve dataset is shown in Fig. 2.

2) DATA PREPROCESSING

Raw ultrasound data needs preprocessing before model training in order to obtain good neural segmentation performance results. Firstly, t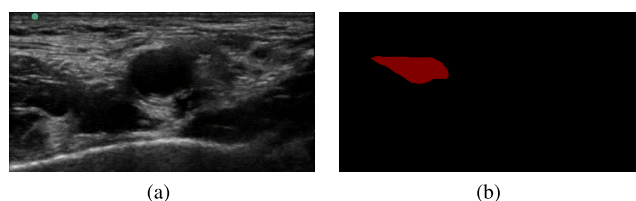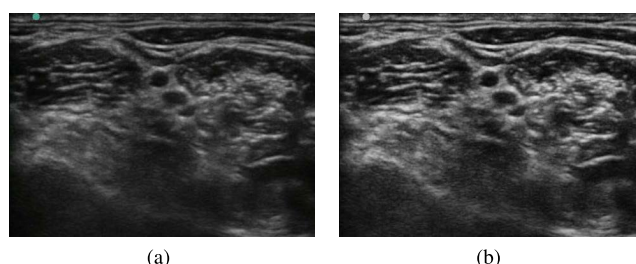he collected ultrasound images may contain ultrasound equipment and patient information, and we crop each image to ensure anonymity and remove redundant information. Secondly, the pixel values of the brachial plexus ultrasound images are mainly concentrated in low gray value areas, and the images are dark and low in contrast. Therefore, we perform enhancement processing on the brachial plexus ultrasound images using the contrast limited adaptive histogram equalization (CLAHE) method to improve the image quality, as illustrated in Fig. 3.

As we all know, training with large-scale labeled data is one of the important reasons for the success of deep learning in various fields. However, obtaining sufficient amount of labeled training data remains a significant challenge in medical image analysis. In addition, deep learning models are prone to overfitting problems when the amount of data is insufficient, making it difficult for deep learning models to obtain satisfactory performance results. Data augmentation aims to make small changes to existing data to increase the amount of training data. It is an important step in the training process of deep learning models, which can alleviate the problem of small data volume, thereby reducing the overfitting phenomenon in deep learning model training and improving the performance of neural networks.

In this study, we use the Albumentations library [27] to perform data augmentation operations, which efficiently implements various image transformation operations. More specifically, we implement two data augmentation methods, random cropping, and random flipping, using the Compose method in the Albumentations library. In order to reduce the chance of cropping out the image of the neural region due to random cropping, we first scale the image to 300×380 to make it similar to the size of the target image during random cropping. Then we randomly crop the image to 256×320.

TABLE 1. Overview of deep learning models.

Model	Ref.	Params size	Network depth	Remarks
FCN	[30]	76.68 MB	94	a. The full connection layer is changed to the convolution layer. b. Use jump connection to fuse information. c. The first end-to-end structure.
SegNet	[31]	112.39 MB	92	a. Use upsampling to restore image size. b. Use deconvolution to enrich image information.
PsPNet	[32]	250.62 MB	376	a. Propose a pyramid pooling module to aggregate contextual information in different regions. b. Introduce an auxiliary loss.
U-Net	[33]	142.54 MB	89	a. Strictly symmetrical U-shaped structure. b. Feature fusion is performed on the image channel.
U-Net++	[34]	286.63 MB	168	a. Dense jump connections and make full use of shallow and deep features. b. Deep supervision.
DeepLabv3+	[35]	208.31 MB	301	a. Apply Xception [28] structure to segmentation tasks. b. Use atrous spatial pyramid pooling (ASPP) module. c. Introduce deep separable convolution.
GCN	[36]	222.68 MB	399	a. Find a balance between classification and positioning. b. Alleviate the problem of large convolution kernel parameters.
LinkNet	[37]	44.00 MB	113	a. The encoder and the corresponding part of the encoder are directly connected. b. Fewer model parameters, fast processing.
ENet	[38]	1.33 MB	350	a. Encoder and decoder asymmetry. b. Use PReLU as activation function. c. Less calculation, fewer model parameters and fast speed.
BiseNet	[39]	45.37 MB	168	a. Bidirectional semantic segmentation network. b. Introduce feature fusion module and attention refinement module.
DFANet	[40]	9.06 MB	771	a. Propose an encoding module for semantic segmentation with a multi-connected structure. b. Change the Xception network structure and increase the attention layer to ensure the receptive field.
BiseNetV2	[41]	13.81 MB	218	a. A novel deep convolution-based lightweight network is designed on the semantic branch to enhance the receptive field and obtain rich intertextual information. b. An augmentation training strategy is introduced to improve the segmentation performance further.

Finally, we flip the image horizontally around the X-axis and horizontally around the Y-axis with a probability of 0.5. In our experiment, images are randomly processed in data augmentation and returned to the network during training.

B. DEEP LEARNING MODELS

The advancement of the deep learning technology has greatly promoted the development of the field of image segmentation. This study explores and evaluates the effect of different deep learning models in brachial plexus ultrasound image segmentation. Ultrasound is widely used to guide nerve blocks during surgery due to its flexibility, convenience, and rapid real-time imaging. In fact, rapid real-time imaging is an important reason for using ultrasound in clinical nerve blocks [29]. Therefore, the deep learning model needs to be accurate when identifying brachial plexus and, at the same time, it needs to have good real-time performance. This is an essential prerequisite for the clinical applications of related technologies. In this paper, we evaluate 12 deep learning models (FCN [30], SegNet [31], PsPNet [32], U-Net [33], U-Net++ [34], DeepLabv3+ [35], GCN [36], LinkNet [37],

ENet [38], BiseNet [39], DFANet [40], BiseNetV2 [41]), in which several are considered are real-time segmentation models (e.g., LinkNet, ENet, BiseNet, DFANet, BiseNetV2). A real-time segmentation model has faster speed and low requirements for hardware, which is convenient for deployment on mobile devices or embedded devices. Table 1 summarizes some parameters of the deep learning models used in our research, such as the number of parameters and the depth of the network (including activation layer, batch normalization and so on).

C. TRANSFER LEARNING

In recent years, the success of deep learning in many fields can be attributed to three factors, powerful computers, excellent algorithm models, and larger datasets [42]. However, obtaining enough training data for deep learning tasks in the medical field is still a significant challenge. Compared with traditional machine learning methods, deep learning is very much dependent on training data [43]. In medical image analysis, it is difficult for deep learning models to effectively extract image features by using only small samples for
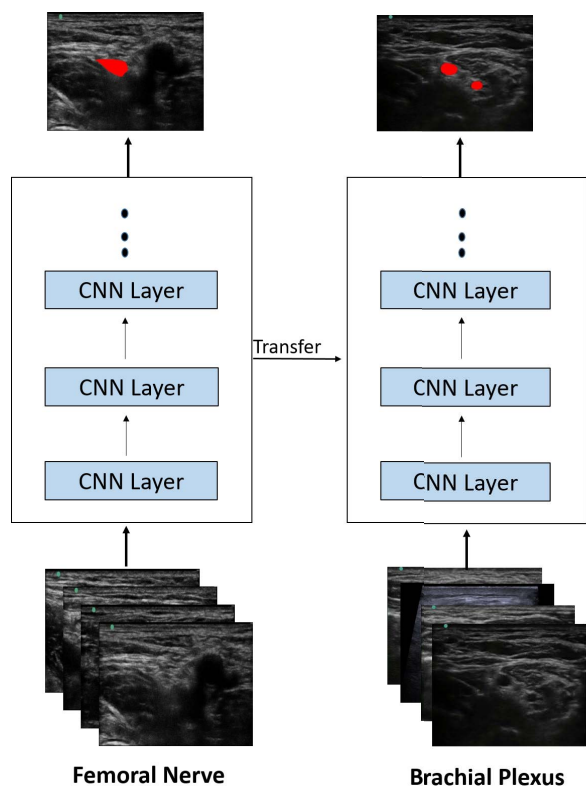
FIGURE 4. Workflow of brachial plexus identification using transfer learning.

training. In addition, the small dataset may lead to overfitting of the deep learning model, making it difficult for the model to achieve satisfactory performance results and accomplish the expected goal.

To solve the above problem, a common approach is transfer learning. Studies have shown that transfer learning has achieved many results in medical image analysis and has been widely used for various research [44]. Transfer learning aims to extract knowledge from one or more tasks and use it in an intended task. Therefore, we can adopt transfer learning to deal with the problem of insufficient training data in medical image analysis and alleviate the dependence of deep learning models on training data [45]. Currently, there are some classic models in PyTorch. These models are trained on large-scale benchmark datasets such as ImageNet, which can provide pre-training parameters for the implementation of various deep learning tasks. Our study uses these classic architectures when implementing segmentation models, such as VGG [46], ResNet [47], and other network structures as feature extractors. The parameters of these classic models are trained on the ImageNet dataset. In addition, we pre-train our networks using the femoral neural dataset [17] and this dataset is similar to our data and can provide better initialization parameters for our segmentation task. Fig. 4 shows the transfer learning process in our research.

D. LOSS FUNCTION

The loss function aims to measure the difference between labels and predicted results. It is an essential part of image segmentation methods based on deep learning. Researchers have designed many loss functions for different image segmentation tasks. The loss function that fits the data characteristics will positively impact the segmentation results. However, choosing a loss function suitable for a task is challenging. Therefore, in this study, we evaluate the effect of several commonly used image segmentation loss functions in brachial plexus segmentation of ultrasonic images. The purpose is to provide a reference or a basis for researchers to select proper loss functions. The loss functions we evaluated are cross entropy (CE) loss, dice loss [48], focal loss [49], combo loss [50], cross entropy loss with focal loss, and cross entropy loss with lovász-softmax loss [51].

Cross entropy is derived from the Kullback-Leibler (KL) divergence and is an indicator to measure the difference between two distributions [52]. Cross entropy loss is one of the most commonly used loss functions. Its effect is stable and can be used in most semantic segmentation scenarios. Dice loss can directly optimize the dice similarity coefficient (DSC), which is an indicator used to measure the similarity of sets and is usually used to evaluate the segmentation performance of the model. Dice loss is suitable for unbalanced samples. Focal loss originates from the direction of target detection and is an improvement of standard cross entropy loss. It aims to solve the unbalanced number of complex and easy samples and unbalanced foreground and background categories. Combo loss is a combination of cross entropy loss and dice loss. It attempts to use dice loss to solve the class imbalance problem and uses cross entropy loss to smooth the curve. Cross entry loss with focal loss is a combination of cross entropy loss and focal loss. Cross entropy loss with lovász-softmax loss is a combination of cross entropy loss and lovász-softmax loss. Lovász-softmax loss is used to optimize Jaccard directly. Since using the lovász-softmax loss alone does not achieve the desired results, we combine the cross entropy loss with the lovász-softmax loss.

III. EXPERIMENTS AND ANALYSIS

A. EXPERIMENTAL SETUP

We use PyTorch 1.10 deep learning framework to implement all deep learning models in our study. We complete all training and testing using two types of computers with Intel i3-10100F central processing units (CPU), NVIDIA GTX1050ti graphics processing units (GPU), and Intel i9-10900X CPU, NVIDIA RTX3090 GPU. For all segmentation models, the input image size is $(256 \times 320 \times 3)$. We set the initial learning rate to 0.0001, the batch size to 8, the number of epochs to 200, and use the Adam algorithm as the optimizer for model training. In addition, we use Pytorch's ReduceLROnPlateau method to adjust the learning rate during the training process. The learning rate decreases to the original by 50% when the results do not improve after ten

TABLE 2. Comparison of IoU results with different loss functions.

Model	CE loss	Dice loss	Focal loss	Comobo loss	CE + Focal loss	CE + Lovász-softmax loss
FCN	0.5962	0.5690	0.5746	0.6020	**0.6075**	0.5904
SegNet	0.6230	0.6246	0.5676	0.5929	0.6198	**0.6285**
PsPNet	0.5147	0.5308	**0.5463**	0.5364	0.5066	0.5436
U-Net	0.6561	0.6609	0.6282	0.6540	**0.6850**	0.6738
U-Net++	0.6381	0.6147	0.6171	0.6412	**0.6434**	0.6351
DeepLabv3+	0.5734	0.5633	0.5366	0.5806	0.5935	**0.6021**
GCN	0.5875	0.5644	0.5162	**0.5914**	0.5684	0.5853
LinkNet	0.6533	0.6488	0.6233	0.6571	0.6465	**0.6627**
ENet	0.5918	0.5713	0.6014	0.5833	0.6084	**0.6295**
BiseNet	**0.5937**	0.5503	0.5412	0.5680	0.5514	0.5586
DFANet	0.6116	0.5953	0.5612	0.6092	**0.6155**	0.5913
BiseNetV2	0.6130	0.5990	0.5900	0.6153	**0.6468**	0.5881
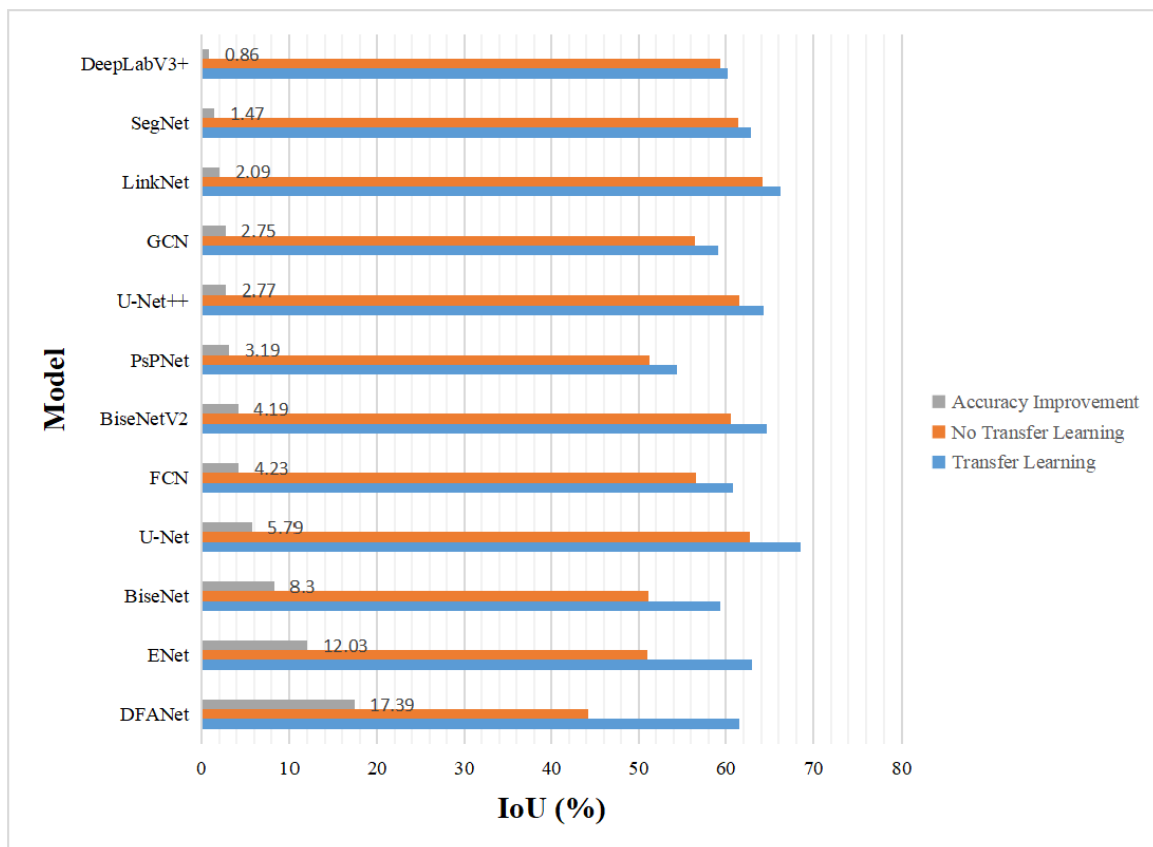
FIGURE 5. Performance comparison between transfer learning and non-transfer learning.

epochs. After completing training, we save the model weights that perform best on the validation set for testing. The above settings remain the same when running all models.

Our data contains 340 ultrasound images. In order to construct a reasonable dataset, we randomly select 34 images as the test dataset. To allow more data to be used for training, we divide the remaining ultrasound images into a training dataset and a validation dataset by 9:1. We do not divide the test dataset for the femoral neural dataset since our purpose is only to provide pre-training parameters for the neural network model to examine the effect of transfer learning.

B. EXPERIMENTAL RESULTS AND ANALYSIS

1) PERFORMANCE RESULTS OF DIFFERENT LOSS FUNCTIONS

The loss function is an important part of an image segmentation method based on deep learning. To determine the loss function suitable for the brachial plexus segmentation task, we evaluate several common image segmentation loss functions. The specific results are shown in Table 2. We use intersection over union (IoU) as the evaluation metric for segmentation accuracy. IoU measures the similarity between two sets. It is one of the most commonly used metrics to measure the accuracy of image segmentation, which is
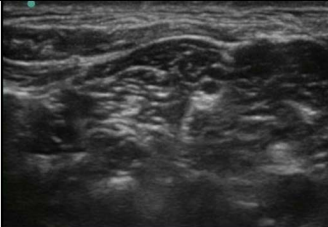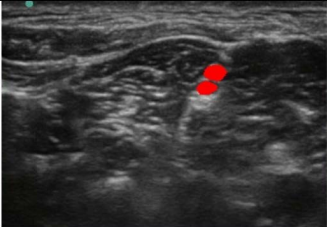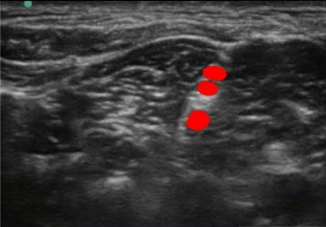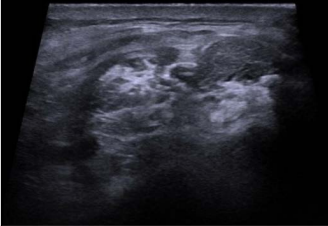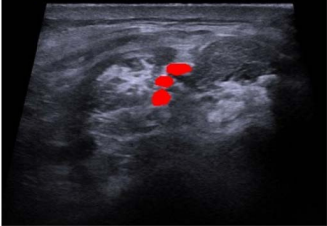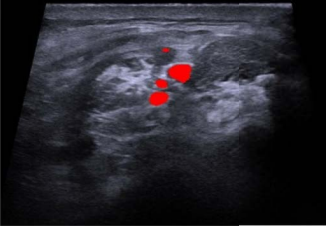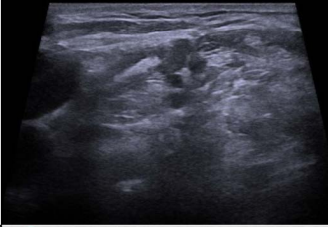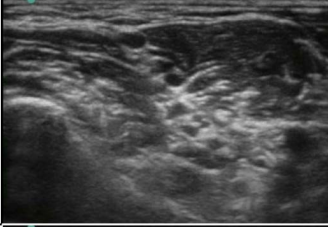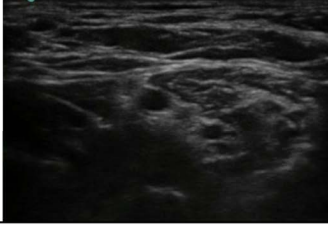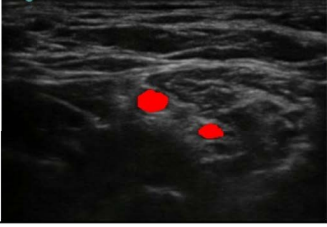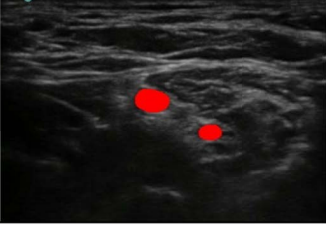
IoU	Original	Label	Predicted
0.4969			
0.5404			
0.6578			
0.7611			
0.8211			

FIGURE 6. Representative results of U-Net with different IoU levels in test data. The first column gives the IoU value of the predicted result of the image. The second column represents the original images. The third column represents images with labels annotated by doctors, and the fourth column represents images with masks predicted by the deep learning model. Both masks annotated by doctors and predicted by the deep learning model are highlighted with red on the original image. The CPU used is an Intel i3-10100F, and the GPU is an NVIDIA GTX1050ti.

defined as

$$IoU = \frac{Intersection}{Union} \qquad (1)$$

where the Intersection is the overlapping part of predicted image and label, and the Union is the combination of the predicted image and label. The results in Table 2 show that no single loss function can achieve the best performance for all 12 deep learning models. But the compound loss functions (combo loss, cross entropy loss with focal loss, cross entropy loss with lovász-softmax loss) are the most robust loss, and they achieve the best performance for ten deep learning

models. In addition, although the cross entropy loss function achieves only one best performance result, it achieves relatively good results across all models, with the most robust performance results among the three single losses (CE loss, dice loss, focal loss).

2) PERFORMANCE RESULTS FOR TRANSFER LEARNING

Transfer learning is a common method for training neural networks. In our study, we pre-train the neural network using the femoral nerve dataset to obtain initialization parameters for brachial plexus segmentation. Fig. 5 shows the results

TABLE 3. Segmentation performance of different models.

| Model	MACs (G)	Params Size (MB)	Intel i9-10900X, NVIDIA RTX3090		Intel i3-10100F, NVIDIA GTX1050ti		IoU (%)	
			Time (ms)	fps	Time (ms)	fps	No-CLAHE	CLAHE
FCN	25.20	76.68	**5**	**200**	28	35	58.07	60.75
SegNet	51.65	112.39	9	111	52	19	60.62	62.85
PsPNet	24.59	250.62	27	37	46	21	53.71	54.63
U-Net	54.85	142.54	11	90	63	15	**64.99**	**68.50**
U-Net++	261.58	286.63	35	28	280	3	60.99	64.34
DeepLabv3+	25.89	208.31	17	58	50	20	58.44	60.21
GCN	19.05	222.68	28	35	54	18	56.88	59.14
LinkNet	3.23	44.00	6	166	**7**	**142**	64.01	66.27
ENet	**0.65**	**1.33**	14	71	15	66	57.51	62.95
BiseNet	3.79	45.37	6	166	9	111	55.03	59.37
DFANet	0.68	9.06	30	33	31	32	58.41	61.55
BiseNetV2	4.03	13.81	9	111	10	100	56.89	64.68
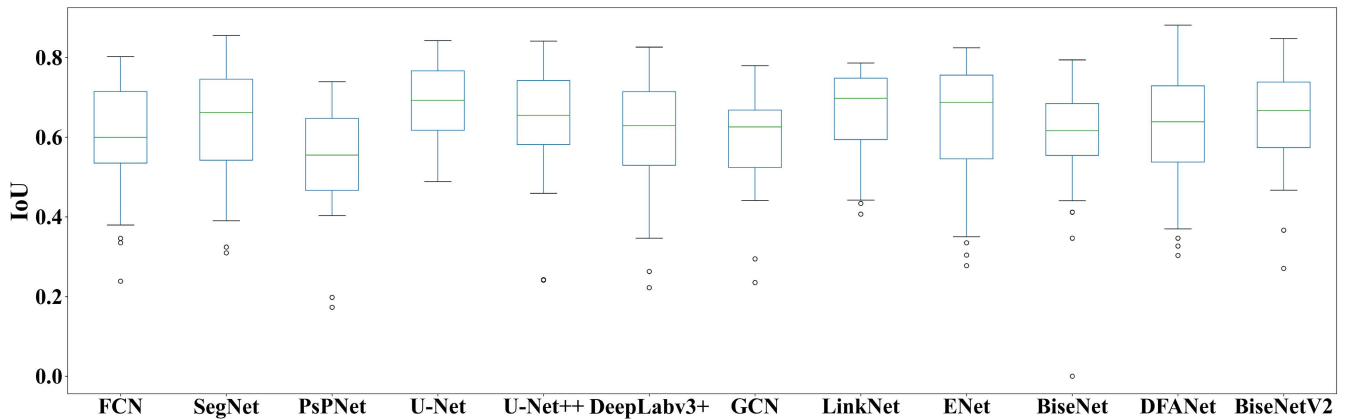
FIGURE 7. Comparison of IoU results using different deep learning models. The CPU used is an Intel i3-10100F, and the GPU is an NVIDIA GTX1050ti.

TABLE 4. Performance results on different datasets.

Model	Train:BK3000, Test:BK3000	Train:BK3000, Test:YGY	Train:YGY, Test:YGY	Train:YGY, Test:BK3000	Mixed
FCN	0.5681	0.3671	0.5956	0.2251	0.5704
SegNet	0.6005	0.4278	0.6087	0.2748	0.5725
PsPNet	0.4730	0.2866	0.4917	0.2447	0.5078
U-Net	**0.6302**	0.4935	0.6596	**0.4528**	0.6216
U-Net++	0.5876	0.4921	0.6361	0.3822	**0.6271**
DeepLabv3+	0.5578	0.2313	0.5709	0.2556	0.5300
GCN	0.5218	0.1982	0.5936	0.2421	0.5758
LinkNet	0.6263	**0.5024**	**0.6603**	0.4476	0.6096
ENet	0.5631	0.4881	0.5958	0.3649	0.5709
BiseNet	0.4659	0.4229	0.5807	0.2934	0.5134
DFANet	0.5407	0.4120	0.5992	0.3195	0.5758
BiseNetV2	0.5538	0.2353	0.5854	0.2421	0.6019

of using the transfer learning method and the results of not using the transfer learning method. It can be seen from the figure that the segmentation accuracy is improved when using the transfer learning method to train the network. Especially in the real-time or lightweight network, the performance improvement is more prominent. Therefore, we suggest using transfer learning methods in brachial plexus recognition when the amount of data is insufficient, especially, when using lightweight networks.

3) PERFORMANCE RESULTS OF REAL-TIME SEGMENTATION NETWORKS

To analyze the performance results of each deep learning model, we compare all deep learning models from three different evaluation metrics of segmentation accuracy, inference speed, and hardware requirements. Table 3 shows the results of all comparisons.

U-Net achieves the best result among all deep learning models with an IoU of 68.50%. To visually show the deep learning results in identifying brachial plexus, we select representative results of different IoU levels from the test results of U-Net and highlight them on original ultrasound images, as shown in Fig. 6. Among all the models, the segmentation accuracy of PsPNet is the lowest, and the IoU is only 54.63%, which has a large gap with the best results. The segmentation performance of the real-time segmentation models (LinkNet, ENet, BiseNet, DFANet, BiseNetV2) exceeds 59%, and the IoU of LinkNet reaches 66.27%, which is the second-best
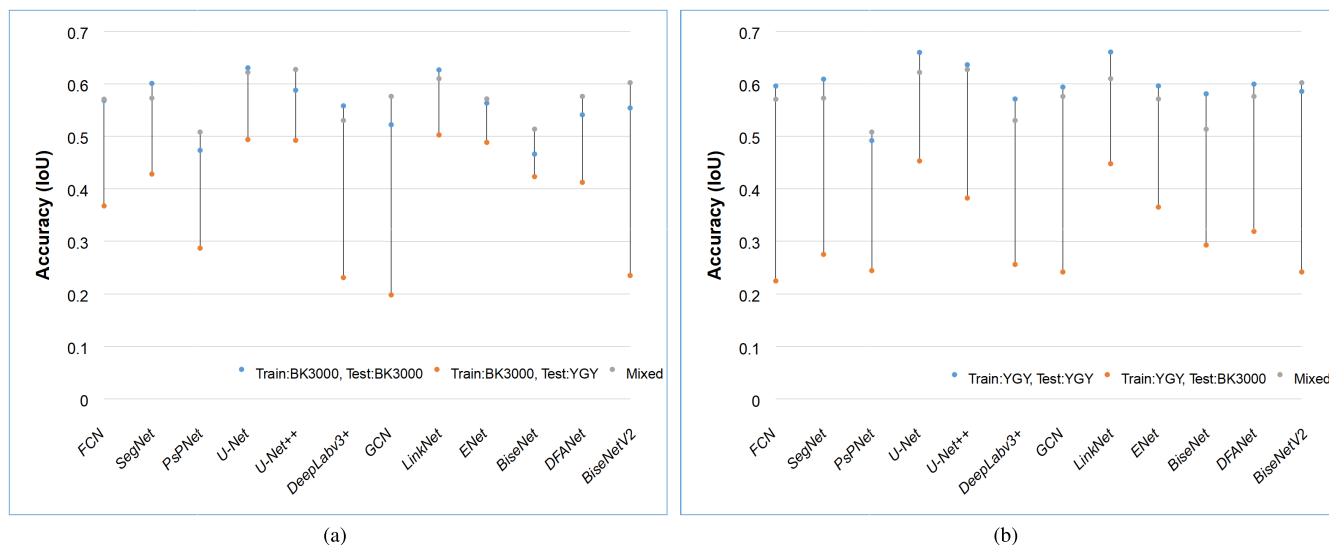
FIGURE 8. Difference in performance of each deep learning model on different data. (a) Performance results of the deep learning model are trained on the BK3000 dataset, tested on the BK3000 and YGY datasets, respectively, and trained and tested on the mixed dataset. (b) Performance results of the deep learning model are trained on the YGY dataset, tested on the YGY and BK3000 datasets, respectively, and trained and tested on the mixed dataset.

result among all models. Fig. 7 shows the distribution of IoU for different deep learning models on the test dataset.

The CLAHE algorithm optimizes the Adaptive Histogram Equalization (AHE) algorithm [53]. It avoids excessive image enhancement and overcomes the problem of excessive noise amplification by the AHE algorithm [54]. It is widely used in image enhancement processing. In our experiments, the ultrasound images are CLAHE processed using the built-in toolkit of the image processing tool OpenCV-Python, with clipLimit set to 1 and tileGridSize set to (8, 8). By analyzing Table 3, we observe that all models' segmentation accuracy improved after processing the images with CLAHE. Therefore, we believe that CLAHE has a significant role in improving the performance of the models in identifying the brachial plexus nerve in ultrasound images.

Model inference speed is a critical factor for applying the brachial plexus recognition method based on deep learning in the clinic. Therefore, we compare the inference speed of different deep learning models, including the inference time of a single image and the number of frames that can be processed per second. We set the image size to $256 \times 320 \times 3$ and perform inference speed tests on NVIDIA GTX1050ti graphics processing units and NVIDIA RTX3090graphics processing units. As shown in Table 3 U-Net++ has the slowest inference speed. It can process 28 images per second when using NVIDIA RTX3090 graphics processing units and only 3 images per second on NVIDIA GTX1050ti graphics processing units, much lower than the other models. Secondly, U-Net, SegNet, and GCN models can process fewer than 20 images per second on NVIDIA GTX1050ti graphics processing units, and the model inference speed is slow. However, LinkNet can process 142 images per second in NVIDIA GTX1050ti graphics processing units. The inference speed of ENet, BiseNet, and BiseNetV2 are also significantly higher

than other networks. Compared with models such as U-Net, real-time networks such as LinkNet are faster and more suitable for real-time segmentation tasks such as brachial plexus recognition.

To further compare the potential of these 12 deep learning models in mobile or embedded deployment, we analyze their requirements for hardware. The metrics used are the multiply and accumulations (MACs) operations and the storage required to save model parameters. The results show that some models, such as U-Net, are computationally intensive, have many parameters, and require large storage space. The large storage space and computational complexity make it challenging to be applied to various hardware platforms effectively. Real-time segmentation models such as LinkNet are computationally simple. At the same time, its parameter storage requires smaller space, which reduces the demand for hardware storage for model deployment.

4) PERFORMANCE ANALYSIS OF THE MODEL ON DIFFERENT DATASETS

When using deep learning models to identify brachial plexus in actual clinical practice, their performance may be affected by factors such as acquisition parameters, equipment, and methods. To evaluate the generalization performance of deep learning models in identifying brachial plexus nerves in new ultrasound images, we test their performance on different datasets based on two ultrasound machine data. Specifically, we first train and test the model using ultrasound data from one machine. We then test the model using data from another ultrasound machine. Finally, we train and test it using mixed data.

As shown in Table 4, the deep learning model works better in identifying the brachial plexus nerve in images from the same ultrasound machine as the training data. When the
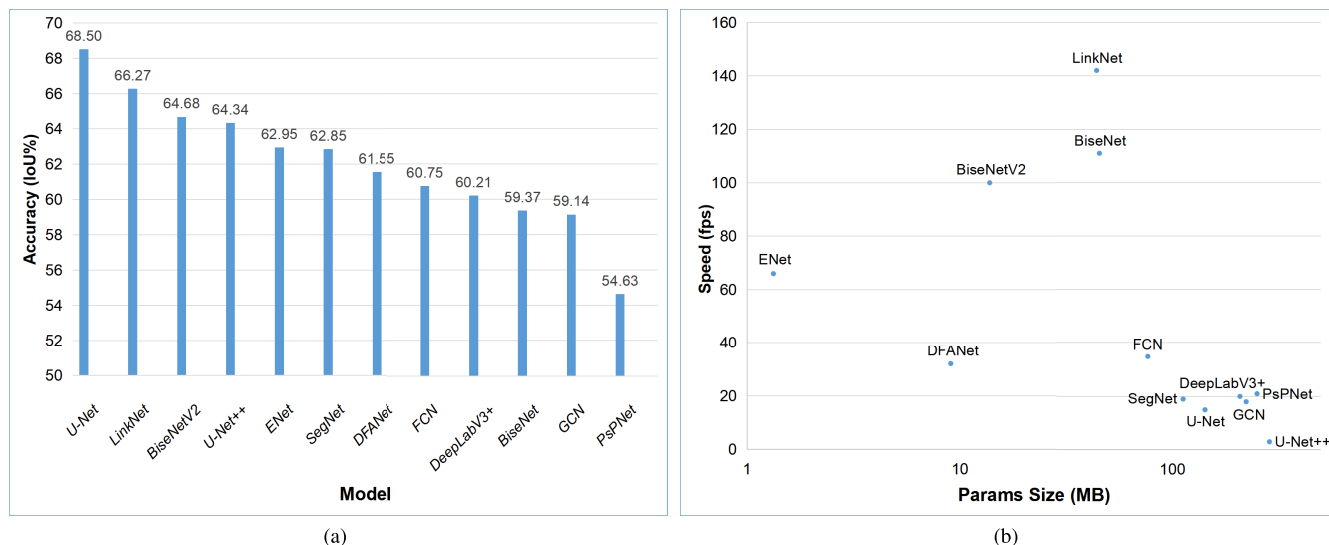
FIGURE 9. (a) Performance of different deep learning models in brachial plexus segmentation. (b) Inference speed and parameters size for different deep learning models.

model is trained using data from BK3000, its best IoU for segmenting the brachial plexus nerve on BK3000 is 63.02%. In comparison, its best IoU is only 50.24% when it is used to identify data from the YGY. When the model is trained using data from YGY, its best IoU for segmenting the brachial plexus nerve on YGY is 66.03%, while its best IoU is only 45.28% when it is used to identify data from BK3000. In addition, there is no significant difference in segmentation accuracy for each model when training with mixed data versus using one type of data alone. Fig. 8 shows the performance differences of each model on different datasets. In conclusion, it is difficult to obtain the same performance when applying a trained deep learning model to identify brachial plexus nerves in ultrasound images obtained by a new ultrasound machine.

IV. DISCUSSIONS

At present, deep learning has been applied to the whole process of medical image processing and has made outstanding achievements in various medical image analysis tasks. Using deep learning technology to identify brachial plexus in ultrasound images to assist doctors in nerve blocks is of great significance for improving the safety and reliability of nerve blocks. In our study, we implement 12 deep learning models for automatic identification of brachial plexus and evaluate their performance.

Among the 12 deep learning models we used, U-Net achieves the best segmentation result with an IoU of 68.50%, as shown in Fig. 9. Similar structures like U-Net, such as U-Net++, FCN, and SegNet, have achieved good segmentation results. However, the inference speed of these models is low, and the number of parameters is large. It is challenging to deploy them efficiently in mobile and embedded devices. In contrast, real-time segmentation networks such as LinkNet have fast inference speed, fewer model parameters, and lower requirements for deployment platforms. At the same time,

these models can also achieve good segmentation results. Compared with models such as U-Net, LinkNet is faster and more suitable for real-time imaging methods such as ultrasound. In addition, the deep learning model obtains better segmentation results when using the same data (machine) for training and testing or using mixed data for training and testing. It is difficult to obtain good identification performance in new ultrasound-acquired images without training the source data.

Transfer learning is a common neural network training strategy. Through comparative experiments, we find that using transfer learning improved the segmentation accuracy of all models. Especially in lightweight networks, performance improvement is more prominent. Therefore, we suggest employing transfer learning in identifying the brachial plexus and using data from similar tasks to assist with the intended task. In comparing loss functions, we find that no loss can achieve the best performance in all 12 deep learning models, but the compound loss function is the most robust loss. In addition, the cross entropy loss is stable among the three single losses, and it can achieve a decent result in all models. Therefore, we suggest that the cross entropy loss be selected first when conducting related research. Then the compound loss function can be selected according to the characteristics of the data set to optimize the results.

V. CONCLUSION

In this study, we implemented 12 deep learning models for automatic segmentation of brachial plexus in ultrasound images and thoroughly evaluated their performance. The results show that complex models (e.g., FCN, SegNet, PsPNet, U-Net, U-Net++, DeepLabv3+, GCN) can often obtain better segmentation results, but the large storage space and computing resource consumption limit their applications on various hardware platforms. The real-time

segmentation networks (e.g., LinkNet, ENet, BiseNet, DFANet, BiseNetV2) can achieve good segmentation results while improving the model's speed, which is more suitable for using real-time imaging equipment such as ultrasound. In addition, we also discussed the strategy of model training, the selection of loss functions, and the model's generalization performance on new data, which can be considered for future research.

Our work has a number of limitations, which we hope to address in the future. First, our dataset is small, with only 340 images. Although data augmentation can partially alleviate model overfitting and improve model performance, we will be able to obtain better results if more data are available. Therefore, we will continue to collect relevant data to establish a larger dataset of brachial plexus ultrasound images. Second, the highest IoU result of the 12 deep learning models we evaluated was only 68.50%. We will investigate new neural network models to improve the performance of brachial plexus recognition.

REFERENCES

[1] Z. Wu and Y. Wang, "Development of guidance techniques for regional anesthesia: Past, present and future," *J. Pain Res.*, vol. 14, p. 1631, Jun. 2021.

[2] L. Wu, Y. Xin, S. Li, T. Wang, P.-A. Heng, and D. Ni, "Cascaded fully convolutional networks for automatic prenatal ultrasound image segmentation," in *Proc. IEEE 14th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2017, pp. 663–666.

[3] J. Bowness, O. Varsou, L. Turbitt, and D. Burkett-St Laurent, "Identifying anatomical structures on ultrasound: Assistive artificial intelligence in ultrasound-guided regional anesthesia," *Clin. Anatomy*, vol. 34, no. 5, pp. 802–809, Jul. 2021.

[4] M. Henderson and J. Dolan, "Challenges, solutions, and advances in ultrasound-guided regional anaesthesia," *BJA Educ.*, vol. 16, no. 11, pp. 374–380, Nov. 2016.

[5] J. Yanase and E. Triantaphyllou, "A systematic survey of computer-aided diagnosis in medicine: Past and present developments," *Expert Syst. Appl.*, vol. 138, Dec. 2019, Art. no. 112821.

[6] J. Gil Gonzalez, M. A. Alvarez, and A. A. Orozco, "Automatic segmentation of nerve structures in ultrasound images using graph cuts and Gaussian processes," in *Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2015, pp. 3089–3092.

[7] V. Vashishtha and D. Aju, "Nerve segmentation in ultrasound images," in *Proc. Innov. Power Adv. Comput. Technol. (i-PACT)*, Apr. 2017, pp. 1–5.

[8] C. A. Jimenez-Castaño, A. M. Álvarez-Meza, O. D. Aguirre-Ospina, D. A. Cárdenas-Peña, and Á. A. Orozco-Gutiérrez, "Random Fourier features-based deep learning improvement with class activation interpretability for nerve structure segmentation," *Sensors*, vol. 21, no. 22, p. 7741, Nov. 2021.

[9] K. Kawauchi, S. Furuya, K. Hirata, C. Katoh, O. Manabe, K. Kobayashi, S. Watanabe, and T. Shiga, "A convolutional neural network-based system to classify patients using FDG PET/CT examinations," *BMC Cancer*, vol. 20, no. 1, pp. 1–10, Dec. 2020.

[10] S. Chibuta and A. C. Acar, "Real-time malaria parasite screening in thick blood smears for low-resource setting," *J. Digit. Imag.*, vol. 33, no. 3, pp. 763–775, Jun. 2020.

[11] S. Gaj, M. Yang, K. Nakamura, and X. Li, "Automated cartilage and meniscus segmentation of knee MRI with conditional generative adversarial networks," *Magn. Reson. Med.*, vol. 84, no. 1, pp. 437–449, Jul. 2020.

[12] W. Zeng, J. Peng, S. Wang, and Q. Liu, "A comparative study of CNN-based super-resolution methods in MRI reconstruction and its beyond," *Signal Process., Image Commun.*, vol. 81, Feb. 2020, Art. no. 115701.

[13] S. M. Galib, H. K. Lee, C. L. Guy, M. J. Riblett, and G. D. Hugo, "A fast and scalable method for quality assurance of deformable image registration on lung CT scans using convolutional neural networks," *Med. Phys.*, vol. 47, no. 1, pp. 99–109, Jan. 2020.

[14] P. Wei and M. Tong, "Attention-DPU: Dual-path UNet with an attention mechanism for ultrasound image segmentation," *J. Phys., Conf. Ser.*, vol. 1693, no. 1, Dec. 2020, Art. no. 012155.

[15] J. Van Boxtel, V. Vousten, J. Pluim, and N. M. Rad, "Hybrid deep neural network for brachial plexus nerve segmentation in ultrasound images," in *Proc. 29th Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2021, pp. 1246–1250.

[16] H. Zhao and N. Sun, "Improved U-Net model for nerve segmentation," in *Proc. Int. Conf. Image Graph.* Cham, Switzerland: Springer, 2017, pp. 496–504.

[17] C. Huang, Y. Zhou, W. Tan, Z. Qiu, H. Zhou, Y. Song, Y. Zhao, and S. Gao, "Applying deep learning in recognizing the femoral nerve block region on ultrasound images," *Ann. Transl. Med.*, vol. 7, no. 18, p. 453, Sep. 2019.

[18] E. Smistad, K. F. Johansen, D. H. Iversen, and I. Reinertsen, "Highlighting nerves and blood vessels for ultrasound-guided axillary nerve block procedures using neural networks," *J. Med. Imag.*, vol. 5, no. 4, Nov. 2018, Art. no. 044004.

[19] M.-H. Horng, C.-W. Yang, Y.-N. Sun, and T.-H. Yang, "DeepNerve: A new convolutional neural network for the localization and segmentation of the median nerve in ultrasound image sequences," *Ultrasound Med. Biol.*, vol. 46, no. 9, pp. 2439–2452, Sep. 2020.

[20] M. M. Rahaman, C. Li, Y. Yao, F. Kulwa, M. A. Rahman, Q. Wang, S. Qi, F. Kong, X. Zhu, and X. Zhao, "Identification of COVID-19 samples from chest X-ray images using deep learning: A comparison of transfer learning approaches," *J. X-Ray Sci. Technol.*, vol. 28, no. 5, pp. 821–839, Sep. 2020.

[21] P. Bizopoulos, N. Vretos, and P. Daras, "Comprehensive comparison of deep learning models for lung and COVID-19 lesion segmentation in CT scans," 2020, *arXiv:2009.06412*.

[22] Z. Cao, L. Duan, G. Yang, T. Yue, and Q. Chen, "An experimental study on breast lesion detection and classification from ultrasound images using deep learning architectures," *BMC Med. Imag.*, vol. 19, no. 1, pp. 1–9, Dec. 2019.

[23] F. Shahidi, S. M. Daud, H. Abas, N. A. Ahmad, and N. Maarop, "Breast cancer classification using deep learning approaches and histopathology image: A comparison study," *IEEE Access*, vol. 8, pp. 187531–187552, 2020.

[24] K. K. Bressem, L. C. Adams, C. Erxleben, B. Hamm, S. M. Niehues, and J. L. Vahldiek, "Comparing different deep learning architectures for classification of chest radiographs," *Sci. Rep.*, vol. 10, no. 1, pp. 1–16, Dec. 2020.

[25] Y. Wang, B. Zhu, L. Kong, J. Wang, B. Gao, J. Wang, D. Tian, and Y. Yao, "Brachial plexus nerve trunk segmentation using deep learning: A comparative study with doctors' manual segmentation," 2022, *arXiv:2205.08143*.

[26] B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman, "LabelMe: A database and web-based tool for image annotation," *Int. J. Comput. Vis.*, vol. 77, nos. 1–3, pp. 157–173, 2008.

[27] A. Buslaev, V. I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, and A. A. Kalinin, "Albumentations: Fast and flexible image augmentations," *Information*, vol. 11, no. 2, p. 125, Feb. 2020.

[28] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1251–1258.

[29] P. Stefan, J. Traub, C. Hennersperger, M. Esposito, and N. Navab, "Challenges in computer assisted interventions: Challenges in design, development, evaluation, and clinical deployment of computer assisted intervention solutions," in *Handbook of Medical Image Computing and Computer Assisted Intervention*. Amsterdam, The Netherlands: Elsevier, 2020, pp. 979–1012.

[30] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.

[31] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder–decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[32] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2881–2890.

[33] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2015, pp. 234–241.

[34] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-Net architecture for medical image segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Cham, Switzerland: Springer, 2018, pp. 3–11.

[35] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder–decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 801–818.

[36] C. Peng, X. Zhang, G. Yu, G. Luo, and J. Sun, "Large kernel matters–improve semantic segmentation by global convolutional network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4353–4361.

[37] A. Chaurasia and E. Culurciello, "LinkNet: Exploiting encoder representations for efficient semantic segmentation," in *Proc. IEEE Vis. Commun. Image Process. (VCIP)*, Dec. 2017, pp. 1–4.

[38] A. Paszke, A. Chaurasia, S. Kim, and E. Culurciello, "ENet: A deep neural network architecture for real-time semantic segmentation," 2016, *arXiv:1606.02147*.

[39] C. Yu, J. Wang, C. Peng, C. Gao, G. Yu, and N. Sang, "BiSeNet: Bilateral segmentation network for real-time semantic segmentation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 325–341.

[40] H. Li, P. Xiong, H. Fan, and J. Sun, "DFANet: Deep feature aggregation for real-time semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 9522–9531.

[41] C. Yu, C. Gao, J. Wang, G. Yu, C. Shen, and N. Sang, "BiseNet V2: Bilateral network with guided aggregation for real-time semantic segmentation," *Int. J. Comput. Vis.*, vol. 129, pp. 3051–3068, Sep. 2021.

[42] D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis," *Annu. Rev. Biomed. Eng.*, vol. 19, pp. 221–248, Jul. 2017.

[43] H. Riaz, J. Park, H. Choi, H. Kim, and J. Kim, "Deep and densely connected networks for classification of diabetic retinopathy," *Diagnostics*, vol. 10, no. 1, p. 24, Jan. 2020.

[44] Y. Wang, X. Ge, H. Ma, S. Qi, G. Zhang, and Y. Yao, "Deep learning in medical ultrasound image analysis: A review," *IEEE Access*, vol. 9, pp. 54310–54324, 2021.

[45] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, "A survey on deep transfer learning," in *Proc. Int. Conf. Artif. Neural Netw.* Cham, Switzerland: Springer, 2018, pp. 270–279.

[46] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[47] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[48] M. Drozdzal, E. Vorontsov, G. Chartrand, S. Kadoury, and C. Pal, "The importance of skip connections in biomedical image segmentation," in *Deep Learning and Data Labeling for Medical Applications*. Cham, Switzerland: Springer, 2016, pp. 179–187.

[49] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988.

[50] S. A. Taghanaki, Y. F. Zheng, S. K. Zhou, B. Georgescu, P. Sharma, D. Xu, D. Comaniciu, and G. Hamarneh, "Combo loss: Handling input and output imbalance in multi-organ segmentation," *Comput. Med. Imag. Graph.*, vol. 75, pp. 24–33, Oct. 2019.

[51] M. Berman, A. R. Triki, and M. B. Blaschko, "The lovasz-softmax loss: A tractable surrogate for the optimization of the intersection-over-union measure in neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4413–4421.

[52] J. Ma, J. Chen, M. Ng, R. Huang, Y. Li, C. Li, X. Yang, and A. L. Martel, "Loss Odyssey in medical image segmentation," *Med. Image Anal.*, vol. 71, Jul. 2021, Art. no. 102035.

[53] G. Yadav, S. Maheshwari, and A. Agarwal, "Contrast limited adaptive histogram equalization based enhancement for real time video system," in *Proc. Int. Conf. Adv. Comput., Commun. Informat. (ICACCI)*, Sep. 2014, pp. 2392–2397.

[54] Y. Chang, C. Jung, P. Ke, H. Song, and J. Hwang, "Automatic contrast-limited adaptive histogram equalization with dual gamma correction," *IEEE Access*, vol. 6, pp. 11782–11792, 2018.

• • •