Ultrasound in Med. & Biol., Vol. 00, No. 00, pp. 1–14, 2020

0301-5629/$ - see front matter

https://doi.org/10.1016/j.ultrasmedbio.2020.03.017

● *Original Contribution*

DEEPNERVE: A NEW CONVOLUTIONAL NEURAL NETWORK FOR THE LOCALIZATION AND SEGMENTATION OF THE MEDIAN NERVE IN ULTRASOUND IMAGE SEQUENCES

MING-HUWI HORNG,* CHENG-WEI YANG,* YUNG-NIEN SUN,* and TAI-HUA YANG†

* Department of Computer Science and Information Engineering, National Pingtung University, Pingtung, Taiwan; and † Department of Biomedical Engineering, National Cheng Kung University Hospital, Tainan, Taiwan

Abstract—Carpal tunnel syndrome commonly occurs in individuals working in occupations that involve use of vibrating manual tools or tasks with highly repetitive and forceful manual exertion. In recent years, carpal tunnel syndrome has been evaluated by ultrasound imaging that monitors median nerve movement. Conventional image analysis methods, such as the active contour model, are typically used to expedite automatic segmentation of the median nerve, but these usually suffer from an arduous manual intervention. We propose a new convolutional neural network framework for localization and segmentation of the median nerve, called DeepNerve, that is based on the U-Net model. DeepNerve integrates the characteristics of MaskTrack and convolutional long short-term memory to effectively locate and segment the median nerve. On the basis of experimental results, the proposed model achieved high performance and generated average Dice measurement, precision, recall and *F*-score values of 0.8975, 0.8912, 0.9119 and 0.9015, respectively. The segmentation results of DeepNerve were significantly improved in comparison with those of conventional active contour models. Additionally, the results of Student's *t*-test revealed significant differences in four deformation measurements of the median nerve, including area, perimeter, aspect ratio and circularity. We conclude that the proposed DeepNerve not only generates satisfactory results for localization and segmentation of the median nerve, but also creates more promising measurements for applications in clinical carpal tunnel syndrome diagnosis. (E-mail: ynsun@mail.ncku.edu.tw)

Key Words: **Carpal tunnel syndrome, Median nerve, U-Net, MaskTrack, ConvLSTM, Ultrasonic.**

INTRODUCTION

Carpal tunnel syndrome (CTS) is the most frequently encountered type of peripheral compression neuropathy and is characterized by median nerve entrapment at the wrist, resulting in median nerve dysfunction. This phenomenon results in a thickened transverse carpal ligament, fibrotic changes of the subsynovial connective tissue and a narrowed space of the carpal tunnel. This causes compression or entrapment of the median nerve (Yoshii et al. 2009), which further leads to variable hand pain and paralysis. Figure 1 is a diagram of the carpal tunnel, which is bounded by the transverse carpal ligament on the volar side and eight carpal bones on the dorsal side. Nine flexor digital tendons and the median nerve pass through the carpal tunnel in the wrist. In recent years, ultrasound imaging has been widely used to diagnose CTS because of its convenience, lower cost, non-invasiveness and shorter examination times. The sensitivity and specificity of ultrasound imaging are as high as 94% and 98%, respectively (Duncan et al. 1999). For diagnosing CTS, several studies have found that the cross-sectional area and the flattening ratio of the median nerve in the ultrasound image are the most effective parameters in identifying median nerve swelling (Hammer et al. 2006). Nerve localization and segmentation are important methods in understanding the deviation and displacement of the median nerve during different hand movements.

Using ultrasonic imaging to track median nerve regions of interest is, however, challenging. Thus, numerous studies of median nerve segmentation and tracking in ultrasonic images have been published. For example, Alkhatib et al. (2018) proposed the adaptive

Address correspondence to: Yung-Nien Sun, Department of Computer Science and Information Engineering, National Pingtung University, No. 1 Ta-Hsueh Road, Tainan 701, Taiwan. E-mail: ynsun@mail.ncku.edu.tw

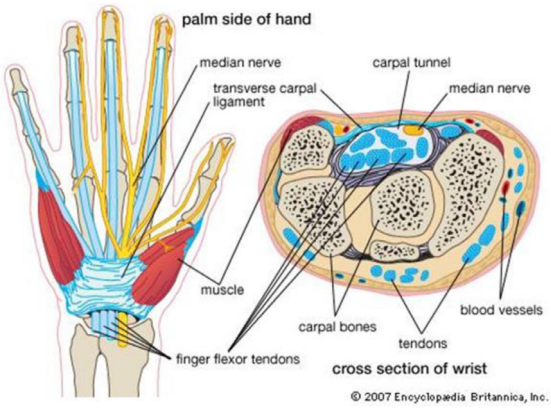

Fig. 1. Carpal tunnel structure (Reed 2005).

median binary pattern (AMBP) as the texture feature of a tracking algorithm and compared it with the particle filter, mean shift and Kanade−Lucas−Tomasi methods; they found that the accuracy of AMBP was 95%. Hadjerci et al. (2016a) proposed a segmentation system comprising pre-processing, feature extraction and support vector machine classifiers for selecting multiple targets of the median nerve, with a confidence decision that is ultimately used to determine the true object of the median nerve. This method generated an average Dice metric of 0.815, which is a confident determination of the true object of median nerve. Hafiane et al. (2014) used the phase-based probabilistic gradient vector flow (PGVF) algorithm to the track sciatic nerve region, which generated an average Dice metric of 0.90 and an average Hausdorff metric of 11.1. Hadjerci et al. (2016b) proposed a framework for median nerve detection and segmentation consisting of four stages: de-speckling filter, feature extraction, feature selection and classification and segmentation. These authors compared the PGVF method with localization + PGVF for segmentation of the median nerve. After testing 1900 ultrasonic images, the authors reported that the localization + PGVF method achieved an average Dice metric of 0.82 and a Hausdorff metric of 10.40; however, PGVF had an average Dice metric of only 0.71 and a Hausdorff metric of 12.51.

Recently, convolutional neural networks (CNNs) have been found to have enormous potential in the field of medical ultrasound analysis (Liu et al. 2019a, 2019b). The popular U-Net architecture was successfully used for precise biomedical image segmentation (Ronneberger et al. 2015; Liu et al. 2019a, 2019b). In U-Net, contextual information is propagated to upsampling layers by concatenating the output of lower layers to higher layers, which provides more feature channels. Kakade and Dumbali (2018) used the U-Net architecture to identify the brachial plexus nerve in ultrasound images. In their article, Kakade and Dumbali pre-processed all images using linear Gabor binary patterns that were then supplied to the U-Net for segmentation. The results revealed an average Dice metric of 0.669 and that use of U-Net to directly segment the median nerve is not effective. Hafiane et al. (2017) claimed that conventional CNNs were not sufficiently robust to locate nerve regions; thus, the authors used a CNN associated with spatial and temporal consistency to enhance localization. The proposed CNN consists of three layers—convolutional, pooling and fully connected layers—that detect the area of interest in the nerve and uses the PGVF method to delineate the regions of the median nerve. The results revealed average Dice and Hausdorff metrics of 0.85 and 10.72, respectively. Another deep learning algorithm is MaskTrack (Federico et al. 2017; Chen et al. 2018), which is a recursive neural network architecture that can effectively segment objects in a sequence image and adequately locate regions of the median nerve. Each step of MaskTrack coarsens previous predictions to preserve the approximate location and shape information and then combines the current image and previous predictions as an input image for segmenting targets. Hochreiter and Schmidhuber (1997) proposed long short-term memory (LSTM), a recursive neural network characterized by the ability to collect previously accumulated messages in the current neural network. This characteristic is suitable for analyzing image sequence data. The convolutional LSTM (ConvLSTM) model adds convolution operations to the input-to-state and state-to-state transitions of the original LSTM, which makes it a powerful method for addressing region segmentation (Romera-Paredes and Torr 2016).

Our main aim in the work described here was to develop a CNN model for median nerve segmentation and localization based on U-Net architecture, called DeepNerve. DeepNerve uses MaskTrack technology Perazzi et al., 2017 to enhance location capability and ConvLSTM to record previous information and reinforce segmentation. The proposed DeepNerve has several advantages over conventional methods. Based on the original U-Net, four different deep learning models were compared: lightweight U-Net, U-Net + MaskTrack, ConvLSTM + U-Net + MaskTrack and DeepNerve. DeepNerve deployed ConvLSTM in the bottom layer of the decoder part of the lightweight U-Net, and was found to generate the best median nerve segmentation of the four models. Second, experimental results were compared with other conventional works based on the active contour model. DeepNerve produced superior results and outperformed the conventional active contour methods, which indicates that deep learning can achieve more precise segmentation. The best results were obtained with DeepNerve compared with existing deep learning models. Therefore, we are confident that DeepNerve is highly capable of performing nerve segmentation.

METHODS

Experimental materials

All ultrasound image sequences were acquired using an ACUSON S2000 ultrasound system (Siemen Medical Solutions, Mountain View, CA, USA) with an 18 L6 HD transducer that assigned an 18-MHz acquisition frequency. Image resolution was 1024 × 768 pixels (each pixel measured 0.075 × 0.075 mm^2). During ultrasonic imaging, the right hand of participants was kept flat and facing upward, as illustrated in Figure 2, and different guided tendon exercises were performed in the order of straight hand, platform position, straight fist, full fist and hook fist. The acquired imaging frequency was 24 fps.

Before experiments, all participants were informed of the study's aims and procedures, and participant identifiers were removed to ensure privacy protection. Participants signed consent forms approved by the institutional review board (IRB) of the National Cheng Kung University Hospital (IRB No. A-ER-107-415). Twenty-four image sequences were captured from six male participants: four normal cases and two CTS patients ranging in age from 18−28 y. Four image sequences per participant were taken; each image sequence consisted of approximately 420 frames. Each frame of the original sequence was first cropped to a region of interest (ROI) 512 × 256 pixels in size. The cropped ROI included the median nerve region. To simplify computation complexity, the size of all ROIs was further minified into 256 × 128 pixels with equal magnification.

All experiments were conducted according to the fourfold cross-validation method. Generally, all image sequences were randomly partitioned into four sets of equal-sized subsamples. Of the four sets, one set was retained to use as data for testing the model, and the remaining three sets were used as training data. In the training stage, each trained sequence was sequentially resampled into 10 equal interval subsequences, consisting of 42 frames. A clinical expert annotated the median nerve in each subsequent frame, which served as the ground truth. Figure 3 is a subsequence annotated median nerve. The image subsequences and the corresponding ground truth were used to train CNN models. To obtain enough training subsequence, each subsequence was augmented by two rotations of 15○ and −15○. One-sixth of all training subsequences were used as validation samples to select parameters for the CNN models. In the testing process, cross-validation was repeated four times in which each of the four sets was used once as test sample to evaluate classification performance. The four sets of test results were then averaged to produce a final evaluation.

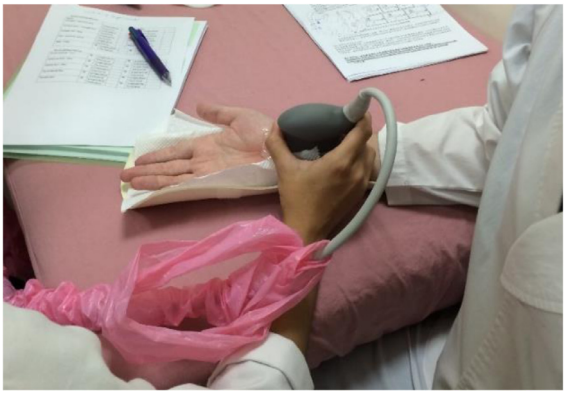

Fig 2. Ultrasound imaging design.

The U-Net, ConvLSTM and MaskTrack methods were found to be effective for identifying targets in a sequence of images. However, integrating their advantages to develop a more accurate algorithm for identification of the median nerve in an ultrasound image sequence remains an important task. In this section, four different deep learning models are described, including lightweight U-Net, U-Net + MaskTrack, ConvLSTM + U-Net + MaskTrack and DeepNerve. Development of these four models is based on the original four-layer U-Net. Figure 4 illustrates the basic structure of the four models. In the training stage, the intensity of each input image and corresponding ground truth were adjusted from [0, 255] to [0, 1]. To obtain stable convergence, an activation sigmoid function was added to the last layer of each model.

U-Net and lightweight U-Net

U-Net is an encoder−decoder-based five-layer architecture that has been effectively applied to biomedical image segmentation, as illustrated in Figure 5. In the encoder stage, the original U-Net uses the stacked convolution and max-pooling layers to extract features from the original image. In the decoder stage, U-Net uses the stacked convolution and deconvolution layers to restore the extracted feature map to the original image size. Additionally, U-Net crops the feature maps of the encoding phase and concatenates them to the deconvolution results of the decoding phase in the same layer. The copy and concatenation operation preserves the segmented messages in the original image and speed ups the convergence time during the learning procedures of the model. However, in our experiments with the original U-Net, we found that it is very difficult to reduce training loss and validation loss. Therefore, U-Net may require additional mechanisms to cope with overfitting by using the limited number of training image sequences.

Batch normalization is a frequently used to avoid premature convergence and to effectively prevent the

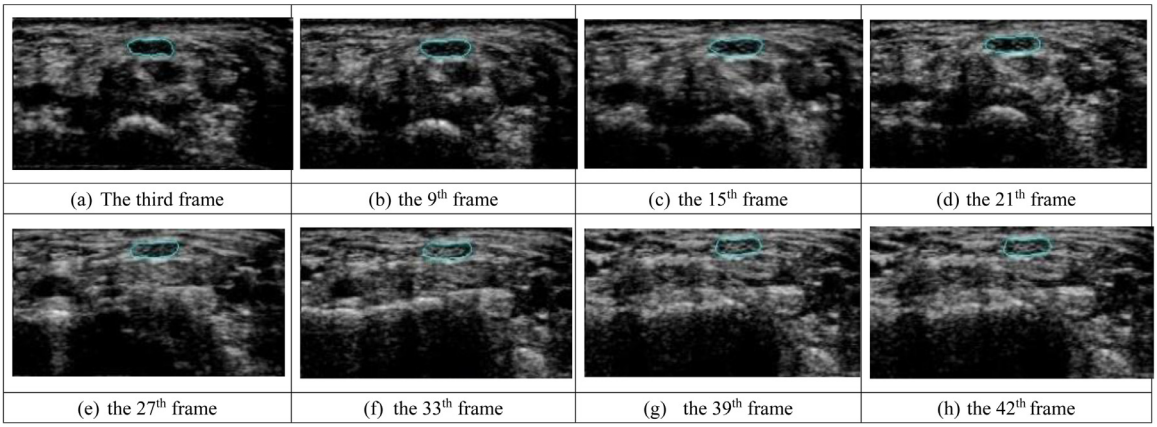

Fig 3. Image sequence. The order of these frames is from top to bottom and left to right. The region of the median nerve is annotated by the blue water contour. The pixel dimension of each frame is 512 × 256 with a frame rate of 24 fps. Each pixel is 0.075×0.075 mm^2.

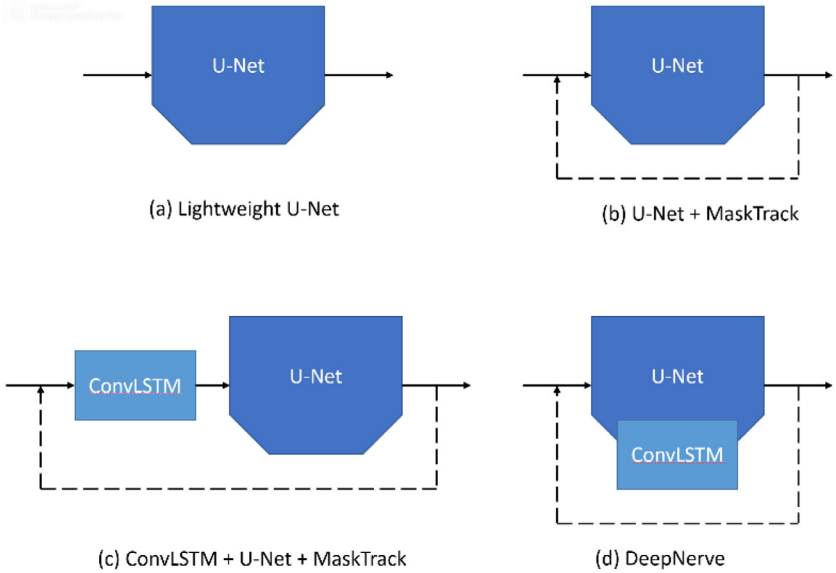

Fig. 4. Four different models proposed in this paper.

gradient from disappearing (Ioffe and Szegedy 2015). Therefore, batch normalization is used as a follow-up step to the first convolutional operation in each layer of the proposed lightweight U-Net. As illustrated in Figure 6, the designed lightweight U-Net contains a total of 7,699,074 trainable parameters with a floating number format.

U-Net + MaskTrack

As illustrated in Figure 7, a structure similar to MaskTrack was incorporated into the lightweight U-Net to enhance the accuracy of median nerve tracking. This second model, called U-Net + MaskTrack, is also a recursive neural network architecture that can effectively segment targets in an image sequence. Each step of this model coarsens the previous predictions to preserve the approximate shapes and corresponding locations. The current image and previous prediction are concatenated as an input of this model to segment targets and output them as the next prediction. The number of trainable parameters in this model is 7,699,585 with a floating number format, which is only a slight increase relative to lightweight U-Net.

ConvLSTM + U-Net + MaskTrack

The primary mechanism of ConvLSTM depends on its memory cell, c_t, which functions as an accumulator of state information. The memory cell is accessed, written and cleared by several self-parameterized controlling gates. When new input arrives, its information is

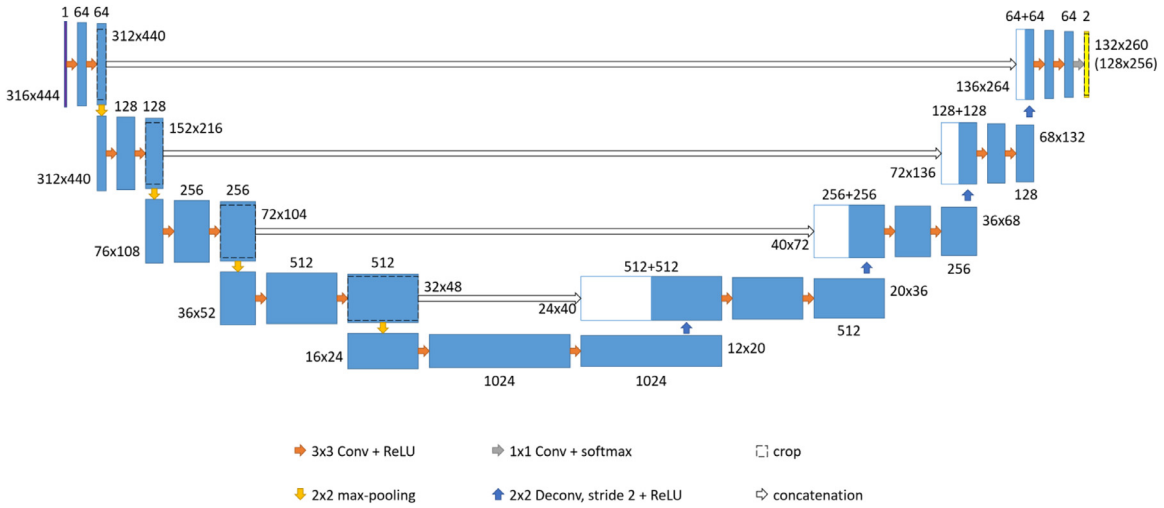

Fig. 5. Original U-Net used in this study. The *purple block* is the input image, and the *yellow block* represents the output. The size of each frame was 128 × 256 and was extended to 316 × 444 by using image padding.

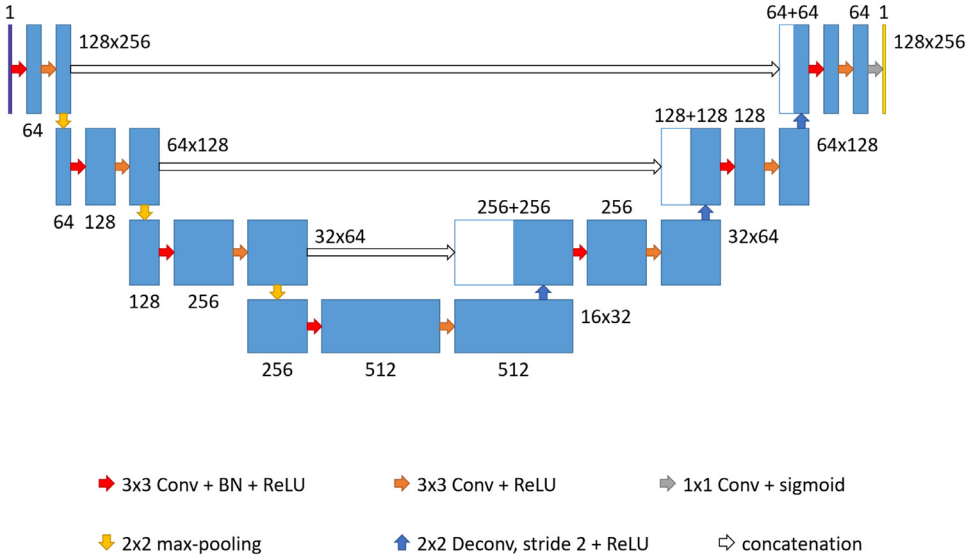

Fig. 6. The lightweight U-Net. The *purple block* is the input image and the *yellow block* represents the output.

accumulated in the cell if the input gate, i_t, is on. Simultaneously, the previous cell status, c_{t-1}, is forgotten in this process if the forget gate, f_t, is on. The latest cell output c_t is propagated to the final state, h_t, which is controlled by the output gate, o_t. The ConvLSTM network (Romera-Paredes and Torr 2016) uses 3 × 3 convolutional operations in both the input-to-state and state-to-state transitions. The operations of ConvLSTM are described in the equations

$$i_t = \sigma(\mathrm{Conv}(x_t; W_{xi}) + \mathrm{Conv}(h_{t-1}; W_{hi}) + b_i) \quad (1)$$

$$f_t = \sigma\big(\mathrm{Conv}\big(x_t; W_{xf}\big) + \mathrm{Conv}\big(h_{t-1}; W_{hf}\big) + b_f\big) \quad (2)$$

$$o_t = \sigma(\mathrm{Conv}(x_t; W_{xo}) + \mathrm{Conv}(h_{t-1}; W_{ho}) + b_o) \quad (3)$$

$$g_t = \phi\big(\mathrm{Conv}\big(x_t; W_{xg}\big) + \mathrm{Conv}\big(h_{t-1}; W_{hg}\big) + b_g\big) \quad (4)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t \quad (5)$$

$$h_t = o_t \odot \phi(c_t) \quad (6)$$

where W denotes the convolution kernel, x_t is the input, h_t represents the short-term state of the output, c_t is the long-term state at time t, b indicates the bias terms, σ denotes the sigmoid, ϕ signifies the hyperbolic tangent function, * denotes the convolution operation, and $\odot$ denotes the Hadamard product.

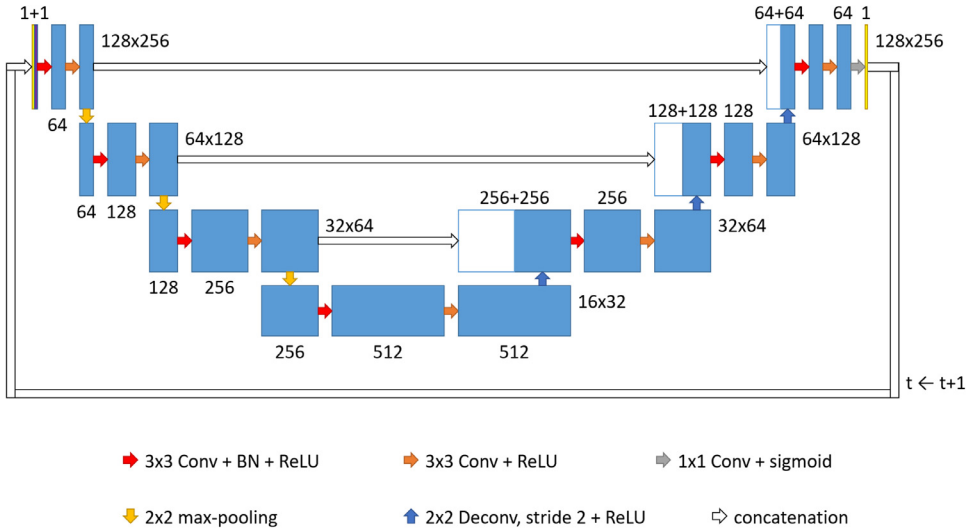

Fig. 7. U-Net + MaskTrack. This model concatenates the previous prediction output (*yellow box*) and the current image (*purple box*) as the model input.

This model uses a ConvLSTM structure to extract important features, then feeds them into U-Net to complete segmentation of the median nerve, as illustrated in Figure 8. The model concatenates the current image, the previous prediction and the context of short-term memory, h_t, as the input instance. The number of trainable parameters in this model is 7,718,081 with a floating number format.

DeepNerve

The ConvLSTM + U-Net + MaskTrack model uses the large size of feature maps to compute and update model parameters and, therefore, requires lengthy computation time. To mitigate this, we attempted to add ConvLSTM to the bottom layer of U-Net + MaskTrack to develop a new model called DeepNerve. Theoretically, the feature maps in the bottom layer contain a large amount of important information and have a relatively small size. Although DeepNerve was found to be superior to the other models, it uses the highest number of trainable parameters (26,576,001) with a floating point format. Figure 9 illustrates the architecture of this model.

Training strategy

Training errors in a deep learning algorithm usually accumulate and ultimately result in a debacle in the

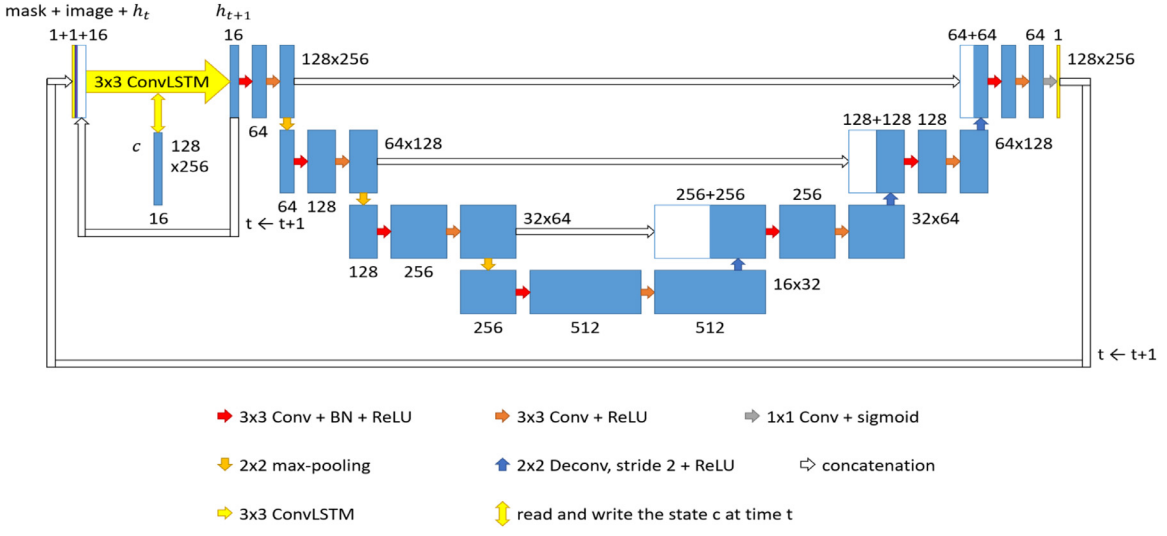

Fig. 8. ConvLSTM + U-Net + MaskTrack. This model concatenates the output of the previous prediction (*yellow box*), the current image (*purple box*) and the context of short-term memory (*white box*) as input.

Table 1. Parameters assigned to each deep learning model

Method	Batch size (image)	Training epochs	No. of trainable variables	Initial learning rate	Loss function	Optimizer	Average training time (s)
Five-layer U-Net	16	90	31,030,658	0.0002	Cross entropy	Adam (Kingma and Ba 2015)	46,175.36
Four-layer U-Net	16	90	7,696,258	0.002	Cross entropy	Adam	25,439.94
Lightweight U-Net	16	60	7,699,074	0.0005	Cross entropy	Adam	10,925.24
U-Net + Mask	16	60	7,699,585	0.0005	Mean square error	RMSprop	13,489.82
ConvLSTM + U-Net + Mask	16	60	7,718,081	0.0001	Mean square error	RMSprop	12,880.93
DeepNerve	16	60	26,576,001	0.0001	Mean square error	RMSprop	17,627.89

training mechanism, especially when using image sequences as training examples. Therefore, an additional strategy for updating parameters is needed to avoid divergence. When training a new image sequence, the first image frame and its corresponding ground truth are input to generate a prediction result, which is then concatenated with the next image frame as the next input. The steps are repeated until each frame is complete. When the number of trained frames is equal to the batch size, the model parameters are updated. In the test stage, the first frame and its rough segmentation of each test sequence are concatenated as an input sample to generate the next prediction for the subsequent frame concatenation.

All experiments were performed on a PC with an Intel Core i7 3.20-GHz CPU, 40-GB memory and NVIDIA Geforce GTX 1080 Ti graphics card. The proposed models were implemented based on the PyTorch framework in Python. The training parameters and training strategy of the models are outlined in Table 1, in which each assignment is optimal for each model. Different learning rates were employed for each model to achieve adequate convergence. With the DeepNerve model as an example, a learning rate of 0.0001 and 60 epochs were assigned. The RMSprop algorithm (Ruder 2017) was used as the optimizer, and the mean square error, as the loss function. A batch size of 16 was selected, and the average training time was 17,627.89 s.

The proposed system was assessed by several performance metrics including accuracy, recall, precision, F score, Dice similarity coefficient (DSC) and Hausdorff metric. For quantitative analysis of experimental results, these metrics were calculated as follows:

$$\text{Accuracy (AC)} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{7}$$

$$\text{Recall } (R) = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{8}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{9}$$

$$F \text{ score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{10}$$

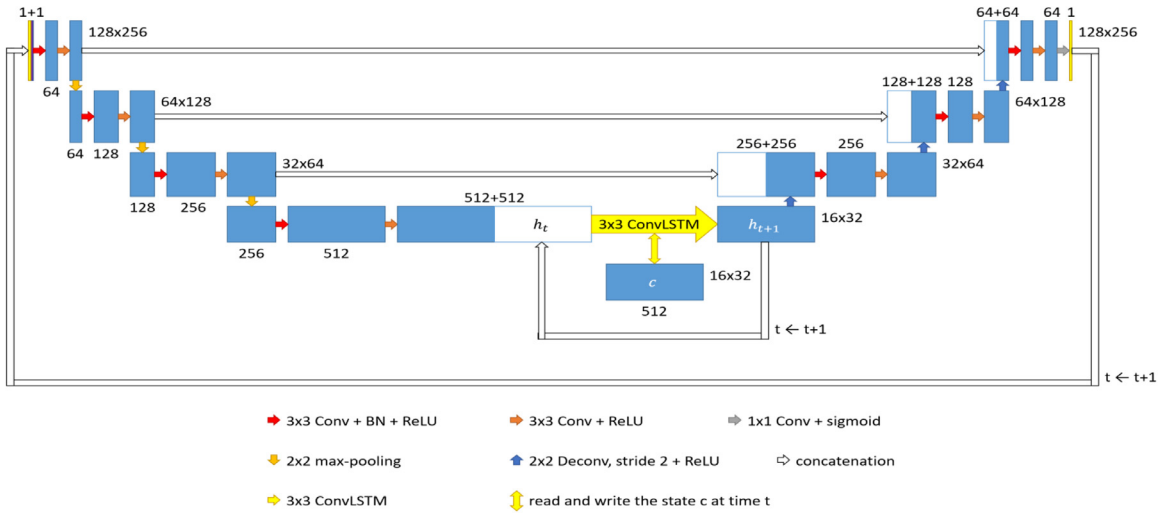

Fig. 9. DeepNerve. DeepNerve concatenates the output of the previous prediction (*yellow box*) and the current image (*purple box*) as input.

Table 2. Average indices of the quantitative evaluation of U-Net and lightweight U-Net

Method	Accuracy	Recall	Precision	*F* Score	DSC measurement	Hausdorff merit
Five-layer U-Net	0.9915 ± 0.0032	0.5442 ± 0.2153	0.7780 ± 0.1875	0.6404 ± 0.1921	0.6115 ± 0.2165	121.015 ± 0.003
Four-layer U-Net	0.9920 ± 0.0029	0.5843 ± 0.2048	0.7977 ± 0.1377	0.6745 ± 0.1632	0.6497 ± 0.19509	124.268 ± 41.249
Lightweight U-Net	0.9939 ± 0.0033	0.7016 ± 0.2618	0.7719 ± 0.2441	0.7350 ± 0.2014	0.7183 ± 0.2462	90.906 ± 34.608

DSC = Dice similarity coefficient.

Dice similarity coefficient (DSC)

$$= \sum_{i=1}^{n} \frac{2\left|GT_i \cap \mathrm{SR}_i\right|}{\left|GT_i\right| + \left|SR_i\right|} \tag{11}$$

where GT_i is the ground truth, SR_i is the segmented median nerve result of the ith frame, n is the frame number of the image sequence and $|\bullet|$ denotes the region size. TP, TN, FP and FN denote the pixel numbers of true positives, true negatives, false positives and false negatives, respectively.

The Hausdorff metric between the ground truth and the segmented median nerve region is defined as

$$\text{Hausdorff metric} = \frac{1}{n}\sum_{i=1}^{n} HD(\ \mathrm{GT}_i, \mathrm{SR}_i) \tag{12}$$

where HD is the Hausdorff distance between the shape of the SR_i and GT_i, which is defined as

$$\mathrm{HD}(\ \mathrm{GT}_i, \mathrm{SR}_i) = \max\left\{ \sup_{x \in \mathrm{GT}_i} \inf_{y \in \mathrm{SR}_i} \| x-y \| , \sup_{x \in \mathrm{SR}_i} \inf_{y \in \mathrm{GT}_i} \| x-y \| \right\} \tag{13}$$

In vivo *feasibility assessment*

To quantify the shape of the median nerve and to record any shape changes such as flattening, Yoshii et al. (2009) measured the nerve perimeter, cross-sectional area, aspect ratio of minimum enclosing rectangle and circularity of the median nerve. The minimum enclosing rectangle was obtained by fitting the smallest possible enclosing rectangle to the region of the median nerve. The aspect ratio of the minimum enclosing rectangle was defined as the ratio of the minor axis length to the major axis length of the minimum enclosing rectangle. They also computed the circularity measure, which is defined as

$$\text{circularity} = \frac{(\text{nerve perimeter})^2}{\text{area} \times 4\pi} \tag{14}$$

The four deformable parameters can characterize the relative motion and deformation of the median nerve on ultrasound images of the carpal tunnel during active finger motion.

RESULTS

Comparison of U-Net and lightweight U-Net performance

In general, recall, precision and *F* score were used to evaluate localization, while segmentation was analyzed based on the Dice measurement and Hausdorff metric (Hadjerci et al. 2016b). Table 2 outlines the performance indices for the five-layer U-Net, four-layer U-Net and lightweight U-Net using the fourfold cross-validation approach.

In the lightweight U-Net, batch normalizations were deployed immediately after the first convolution operation of each layer in the contracting encoder and expansive decoder paths, as illustrated in Figure 6. The proposed lightweight U-Net achieved an average *F* score, Dice coefficient and Hausdorff metric of 0.7350, 0.7183 and 90.9063, respectively. These results are superior to those of the four-layer U-Net. Figures 10–12 illustrate the segmentation results of the image sequence in Figure 3 using the five-layer, four-layer and lightweight U-Net, respectively.

Experimental comparison of the four proposed methods

In this section, the localization capability of the four models described earlier is compared. Figures 12–15 illustrate the segmentation results of the image sequence from Figure 3, and Table 3 lists the performance metrics of the four models. The DeepNerve model achieved an average *F* score measurement of 0.9015, which is better than the 0.7902 generated by ConvLSTM + U-Net + MaskTrack, 0.8127 by U-Net + MaskTrack and 0.7351 by lightweight U-Net. DeepNerve was highly effective at localization of the median nerve. Table 4 outlines the DSC measurement and Hausdorff metrics of lightweight U-Net, U-Net + MaskTrack, ConvLSTM + U-Net + MaskTrack and DeepNerve. The DeepNerve model achieved an average DSC measurement of 0.8975 and Hausdorff metric of 7.729, which are better than those produced by the other three methods. Tables 3 and 4 indicate that all performance indices of the DeepNerve are better than those of the ConvLSTM + U-Net +MaskTrack model. These results suggest that the bottom layer may be a good choice for deployment of the ConvLSTM in the U-Net + MaskTrack model. To verify the classification performance of DeepNerve, Student's *t*-test

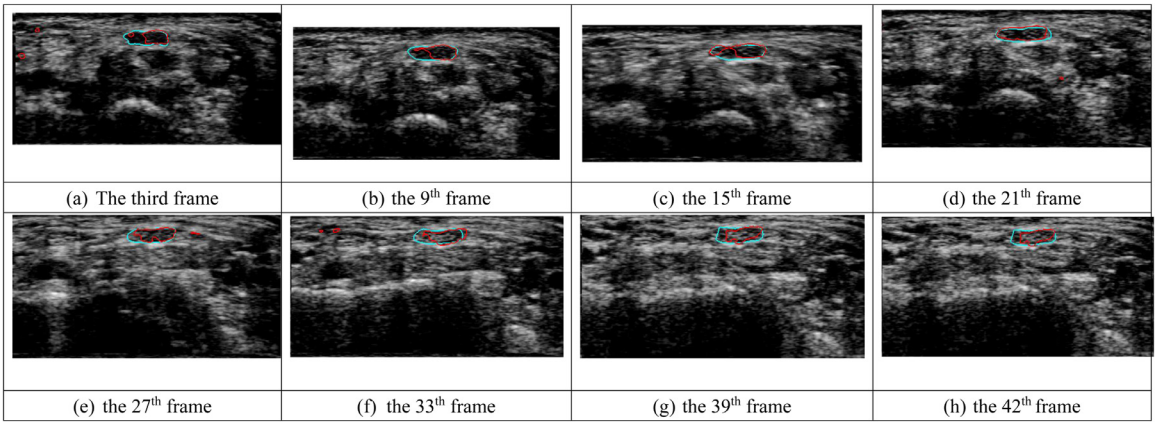

Fig. 10. Segmentation of Figure 3 using the five-layer U-Net. The order of these frames is from top to bottom and left to right. The region of the median nerve is annotated by the *blue water contour,* and the segmented region, by the *red contour*. The pixel dimension of each frame is 512 × 256 with a frame rate of 24 fps. Each pixel is 0.075×0.075 mm^2.

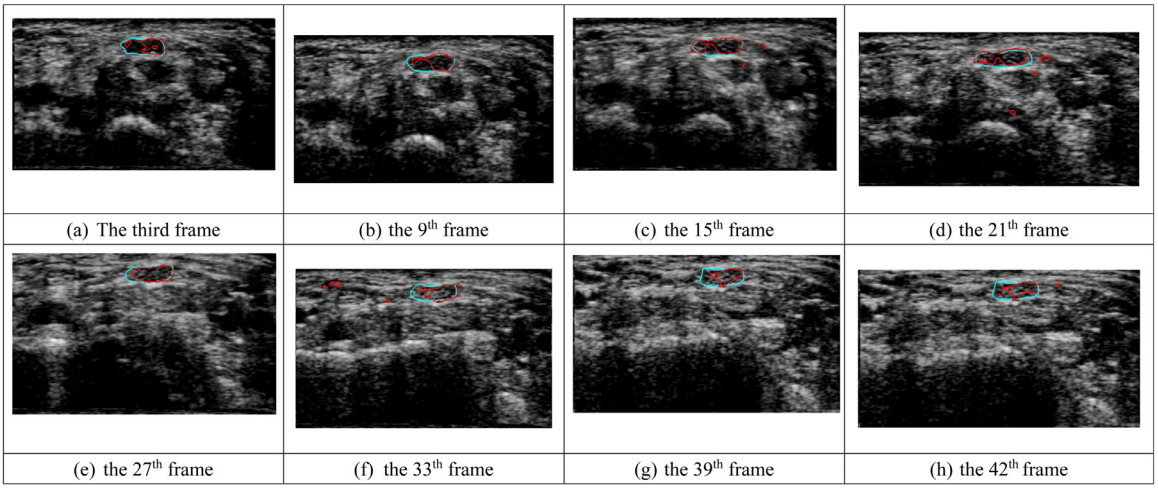

Fig. 11. Segmentation of Figure 3 using the four-layer U-Net. The order of these frames is from top to bottom and left to right. The region of the median nerve is annotated by the *blue water contour,* and the segmented region, by the *red contour*. The pixel dimension of each frame is 512 × 256 with frame rate 24 fps with frame rate 24 fps. Each pixel is 0.075×0.075 mm^2.

was performed, in which the null hypothesis is assumed to be the two measured samples with the same distribution. The results of the test are summarized in Table 5. In this study, all statistical tests were performed by using a software toolkit designed by Brownlee (2018). Additionally, experiments were performed by using a sixfold cross-validation, in which all participants' scans were used as training or test samples. The experimental results are summarized in Table 6. Table 7 outlines the performance of the CNN networks when they were trained only on scans obtained from healthy patients and tested on CTS scans.

Comparison of localization with conventional active contour methods

The segmentation results generated by well-known methods and those used to locate regions of the median nerve are now discussed. Table 8 summarizes the comparative results of our proposed DeepNerve method and several conventional methods, such as the Chan−Vese (CV) (Chan and Vese 2001), the distance regularized level set (DRLS) (Li et al. 2010), the PVGF (Hafiane et al. 2014), the localization + CV (Hadjerci et al. 2016b), the localization + DRLS (Hadjerci et al. 2016b) and localization + PVGF (Hadjerci et al. 2016b). Of these methods, the Localization + PVGF method was the best performer and generated a Dice measurement of 0.81 and Hausdorff metric of 12.06.

Comparison of segmentation with other deep learning methods

To evaluate the segmentation capability of the deep learning algorithm, we compared the U-Net (Kakade and

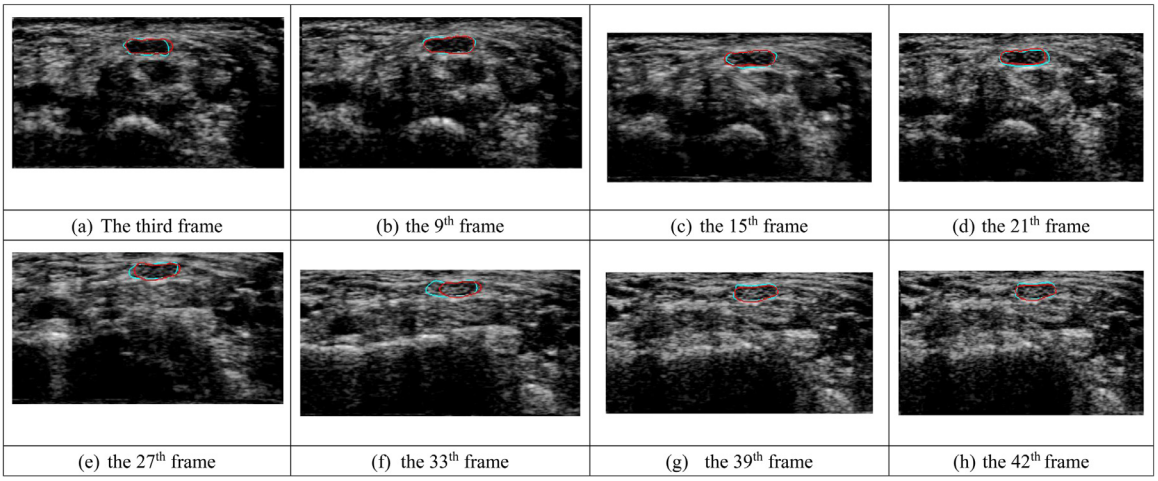

Fig. 12. Segmentation of Figure 3 using the lightweight U-Net. The order of these frames is from top to bottom and left to right. The region of the median nerve is annotated by the *blue water contour*, and the segmented region, by the *red contour*. The pixel dimension of each frame is 512 × 256 with a frame rate of 24 fps. Each pixel is 0.075×0.075 mm^2.

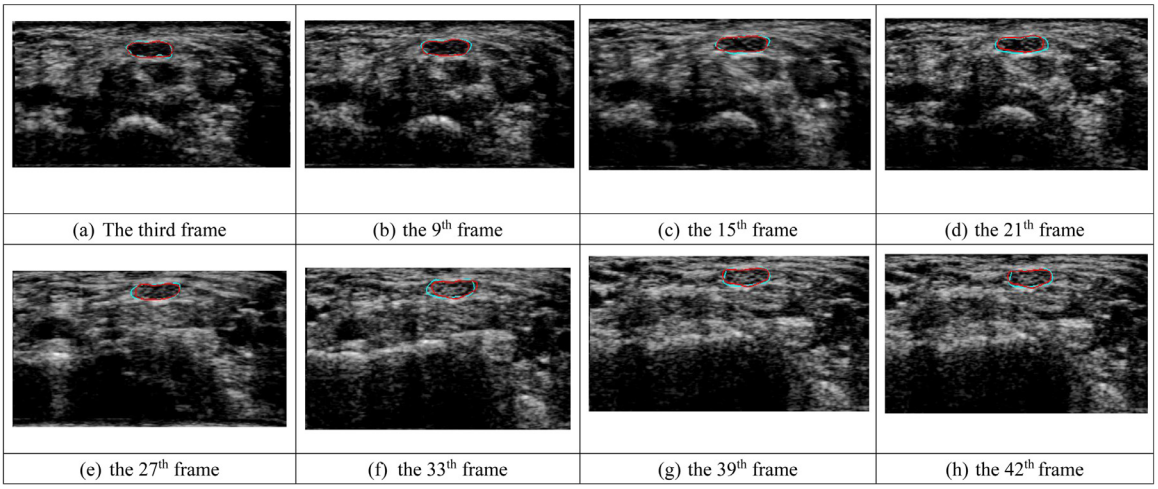

Fig. 13. Segmentation of Figure 3 by using U-Net + MaskTrack. The order of these frames is from top to bottom and left to right. The region of the median nerve is annotated by the *blue water contour*, and the segmented region, by the *red contour*. The pixel dimension of each frame is 512 × 256 with a frame rate of 24 fps. Each pixel is 0.075×0.075 mm^2.

Dumbali 2018) and CNN localization + PVGF (Hafiane et al. 2017) algorithms. The CNN localization + PVGF algorithm used a CNN to locate the region of the median nerve and then applied the PVGF algorithm to segment the median nerve. The results are summarized in Table 9.

In vivo *feasibility assessment using DeepNerve*

To assess the severity of CTS for clinical diagnosis, Lin et al. (2014) and Yoshii et al. (2009, 2013) obtained four deformation measurements of the median nerve, including area, perimeter, aspect ratio and circularity. To verify the feasibility of using the DeepNerve for *in vivo* assessment, the ultrasonic scans of six additional patients were collected to compute the four deformation measurements. The six participants included three normal patients and three CTS patients ranging in age from 21−38 y. Table 10 lists the averages of the four measurements of the six patients, and Table 11 outlines the statistical significance test results of the normal and CTS diseased samples.

DISCUSSION

Based on the performance indices listed in Table 2, it is apparent that the four-layer U-Net performed better than the five-layer U-Net. This result suggests that the streamlining of the U-Net enhanced localization and segmentation accuracy, which motivated us to develop the lightweight U-Net based on the four-layer U-Net. It is evident that the localization and segmentation of the

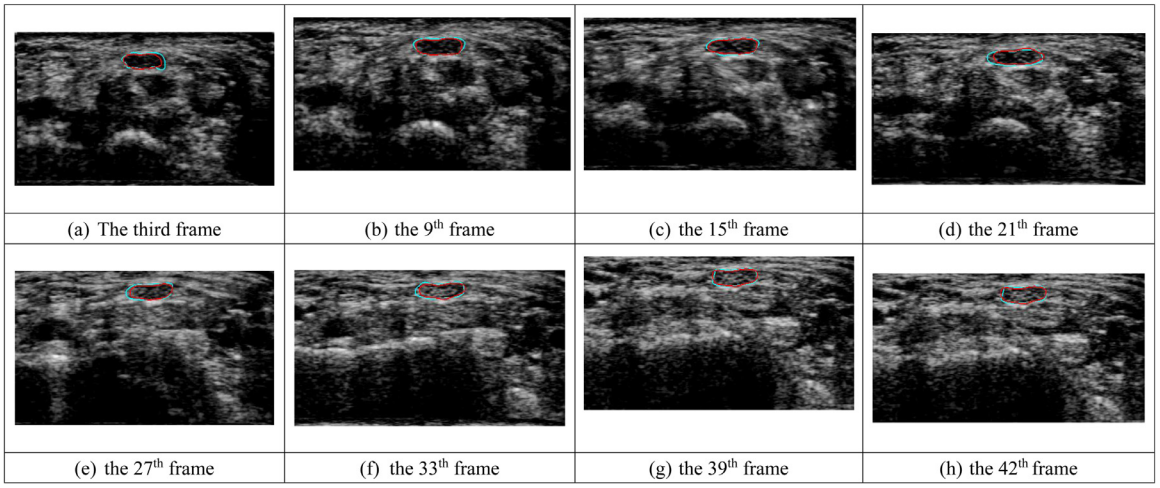

Fig. 14. Segmentation of Figure 3 using ConvLSTM + U-Net + MaskTrack. The order of these frames is from top to bottom and left to right. The region of the median nerve is annotated by the *blue water contour*, and the segmented region, by the *yellow contour*. The pixel dimension of each frame is 512 × 256 with a frame rate of 24 fps. Each pixel is 0.075×0.075 mm^2.

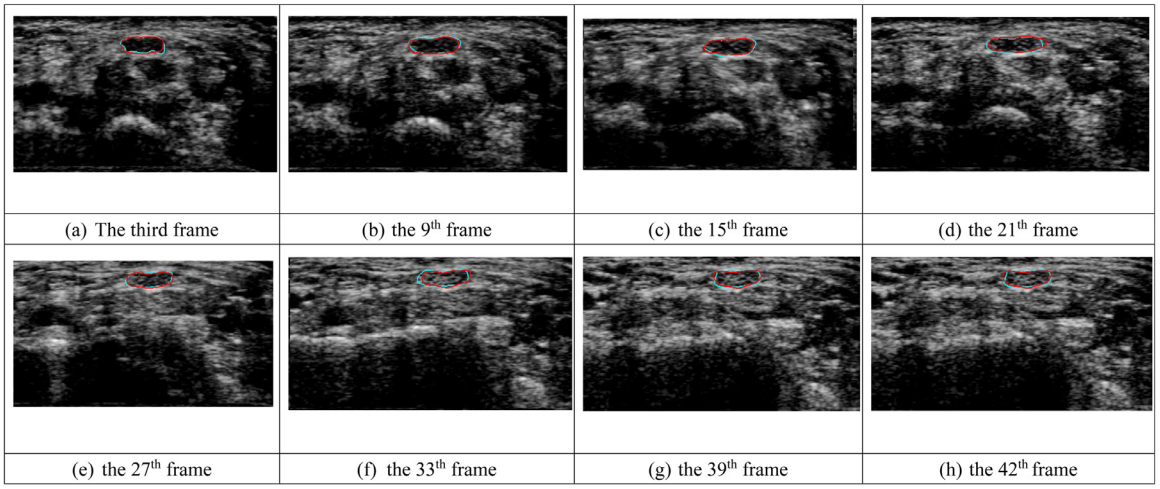

Fig. 15. Segmentation of Figure 3 using DeepNerve. The order of these frames is from top to bottom and left to right. The region of the median nerve is annotated by the *blue water contour,* and the segmented region, by the *yellow contour*. The pixel dimension of each frame is 512 × 256 with a frame rate of 24 fps. Each pixel is 0.075×0.075 mm^2.

Table 3. Performance indices of the four developed models

Method	Accuracy	Recall	Precision	*F* Score
Lightweight U-Net	0.9939 ± 0.0033	0.7016 ± 0.2618	0.7719 ± 0.2441	0.7351 ± 0.2387
U-Net + MaskTrack	0.9938 ± 0.0048	0.8974 ± 0.1345	0.76680 ± 0.1520	0.8127 ± 0.1391
ConvLSTM + U-Net + MaskTrack	0.9950 ± 0.0037	0.7879 ± 0.2053	0.7925 ± 0.2324	0.7902 ± 0.2214
DeepNerve	0.9970 ± 0.0007	0.9119 ± 0.0438	0.8912 ± 0.047	0.9015 ± 0.0418

lightweight U-Net outperformed the other two models. Therefore, the lightweight U-Net was used to develop new models for locating and segmenting the median nerve.

Interestingly, the quantitative indices of ConvLSTM + U-Net +MaskTrack were worse than those of U-Net + MaskTrack, as outlined in Tables 3 and 4. Although the inappropriate use of ConvLSTM may

Table 4. DSC and Hausdorff metric results using the fourfold cross-validation method

Method	DSC measurement	Hausdorff metric
Lightweight U-Net	0.7183 ± 0.2462	90.906 ± 34.608
U-Net + MaskTrack	0.8175 ± 0.1295	14.4676 ± 13.584
ConvLSTM + U-Net + MaskTrack	0.7925 ± 0.2324	57.765 ± 39.971
DeepNerve	0.8975 ± 0.0258	7.729 ± 3.203

DSC = Dice similarity coefficient.

Table 6. DSC measurement and Hausdroff metric using six-fold cross-validation method*

Method	DSC measurement	Hausdorff metric
Lightweight U-Net	0.7015 ± 0.1517	92.952 ± 77.289
U-Net + MaskTrack	0.8108 ± 0.1076	17.79 ± 16.969
ConvLSTM + U-Net + MaskTrack	0.8004 ± 0.0748	37.721 ± 12.735
DeepNerve	0.8877 ± 0.0862	9.357 ± 10.311

DSC = Dice similarity coefficient.
* In the sixfold cross-validation method, no scans of the same person were used as both trained and test samples

Table 7. DSC measurement and Hausdroff metric using the scans of healthy persons as trained samples and testing in scans of carpal tunnel syndrome patients

Method	DSC measurement	Hausdorff metric
Lightweight U-Net	0.6935 ± 0.1675	88.341 ± 50.713
U-Net + MaskTrack	0.8378 ± 0.0966	16.287 ± 16.400
ConvLSTM + U-net + MaskTrack	0.8146 ± 0.0495	37.721 ± 16.789
DeepNerve	0.8654 ± 0.0654	12.212 ± 8.563

Table 8. Performance indices of median nerve segmentation by conventional active contour methods

Method	Dice measurement	Hausdorff metric
CV (Chan and Vese 2001; Hadjerci et al. 2016b)	0.09 ± 0.13	350 ± 19.8
DRLS (Li et al. 2010; Hadjerci et al. 2016b)	0.13 ± 0.03	267 ± 11.2
PGVF (Hafiane et al. 2014)	0.75 ± 0.15	23.9 ± 25.9
Localization + CV (Hadjerci et al. 2016b)	0.69 ± 0.11	37.01 ± 30.9
Localization + DRLS (Hadjerci et al. 2016b)	0.71 ± 0.18	27.12 ± 17.73
Localization + PGVF (Hadjerci et al. 2016b)	0.81 ± 0.10	12.06 ± 2.81
DeepNerve	0.8975 ± 0.0258	7.729 ± 3.203

CV = Chan–Vese method; DRLS = distance regularized level set method; PGVF = probabilistic gradient vector flow method.

decrease segmentation capability, the DeepNerve model generated an average Dice measurement of 0.8975 and a Hausdorff metric of 7.729, which were superior compared with those for the other three models. This result indicated that deployment of ConvLSTM in U-Net + MaskTrack is an important factor in reinforcing segmentation. Finally, the U-Net + MaskTrack was more effective than the lightweight U-Net and had a higher Dice measurement and lower Hausdorff metric. This finding suggests that backtracking of the prediction significantly improved segmentation of the median nerve. In addition, the results of Student's *t*-test revealed the advantages of DeepNerve relative to the other three methods. In Table 5, Student's *t*-tests rejected all null hypotheses between DeepNerve and the other methods; in other words, there are significant differences between DeepNerve and the other models. We also used Student's *t*-test ($p < 0.05$) to determine the differences in DSC measurements associated with DeepNerve, outlined in Tables 4, 6 and 7. The hypotheses were accepted by Student's *t*-test, revealing that DeepNerve is not susceptible to image changes of the median nerve with CTS.

The active contour model is generally used as the basis for designing conventional image analysis methods for median nerve segmentation. However, this led to difficulties with selecting the initial contour, which subsequently caused segmentation errors. Table 8 indicates that the best method, Localization + PVGF, generated acceptable segmentation results of an average Dice measurement of 0.81 and Hausdorff metric of 12.06, which were worse than those for DeepNerve.

In Table 9, results generated by the deep learning techniques, including the original U-Net with principal component analysis transformation and the spatiotemporal consistency-based U-Net Localization + PGVF, are compared with the results generated by our proposed DeepNerve method. The highest Dice measurement and lowest Hausdorff metric consistently indicated that segmentation using DeepNerve performed the best of all methods. DeepNerve offers an end-to-end mechanism for annotating the median nerve of the test image

Table 5. Significance test of DSC measurement of DeepNerve versus one of the other three methods using Student's *t*-test at $p < 0.05$*

Lightweight U-Net versus DeepNerve	U-Net + MaskTracker versus DeepNerve	ConvLSTM + Tracker versus DeepNerve
Rejected H0 (statistic = 22.500) $p = 0.000$	Rejected H0 (statistic = 18.830) $p = 0.012$	Rejected H0 (statistic = 14.011) $p = 0.000$

* Null hypothesis: Two samples are the same distribution.

Table 9. Comparison of our framework with other deep learning models

Method	Dice measure	Hausdorff metric
Four-layer U-Net	0.6497 ± 0.19509	124.268 ± 41.249
Lightweight U-Net	0.7183 ± 0.2462	90.906 ± 34.608
Original U-Net with PCA transformation (Kakade and Dumbali 2018)	0.68828	在這裡鍵入方程式.
Spatiotemporal consistency-based U-Net localization + PGVF (Hafiane et al. 2017)	0.85 ± 0.15	10.72 ± 3.91
DeepNerve	0.897 ± 0.025	7.72 ± 3.20

PCA = principal component analysis; PGVF = probabilistic gradient vector flow method.

Table 10. Averages of the four powerful measurements

Sample	Average of area (pixels)	Average perimeters (pixels)	Average aspect ratio	Average circularity
Normal cases	1775.44 ± 153.54	195.97 ± 17.78	1.73 ± 0.26	0.42 ± 0.15
CTS patients	2002.43 ± 95.12	219.57 ± 14.96	1.92 ± 0.21	0.38 ± 0.08

CTS = carpal tunnel syndrome.

Table 11. Significance test of normal cases versus CTS patients of the four measurements using Student's *t*-test at $p < 0.05$

Deformable parameter	Area	Perimeters	Aspect ratio	Circularity
Results of Student's *t*-test	Rejected H0 (statistic = −24.246)	Rejected H0 (statistic = −20.424)	Rejected H0 (statistic = 4.290)	Rejected H0 (statistic = −11.001)

sequence without the need for any manual intervention, and effectively locates regions of the median nerve and directly segments them without the need for an initial localization procedure.

The results of significance testing in Tables 10 and 11 revealed that there were significant differences in the four measurements of deformation generated by DeepNerve corresponding to the normal and diseased patients. Therefore, DeepNerve is a potential tool for ultrasonic CTS diagnosis.

CONCLUSIONS

The main contribution of this article is the proposal of a new deep learning model called DeepNerve for median nerve segmentation and localization. DeepNerve integrates the characteristics of MaskTrack and ConvLSTM based on the lightweight version of U-Net. DeepNerve directly and effectively locates and segments the median nerve in an end-to-end manner. In six participants with different disease status, significant differences were found in the four median nerve deformation measurements generated by DeepNerve. It appears that DeepNerve is capable of assessing disease severity in patients with CTS. To further determine the feasibility of using DeepNerve, future studies will consider several ultrasonic image sequences of patients with CTS.

Acknowledgments—The authors thank the Ministry of Science and Technology, ROC (Project Nos. MOST-104-2221-E-006-097-MY3, MOST 109-2634-F-006-012 and MOST-108-2622-8-006 -014), for support of this work, which was clinically reviewed under Institutional Review Board No. A-ER-107-415.

REFERENCES

Alkhatib M, Hafiane A, Tahri O, Vieyres P, Delbos A. Adaptive median binary patterns for fully automatic nerves tracking in ultrasound images. Comput Methods Programs Biomed 2018;160:129–140.

Brownlee J. Statistical methods for machine learning: Discover how to transform data into knowledge with Python. : Machine Learning Mastery Pty. Ltd.; 2018.

Chan TF, Vese LA. Active contours without edges. IEEE Trans Image Proc 2001;10:266–277.

Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Trans Pattern Anal Mach Intell 2018;40:834–848.

Duncan I, Sullivan P, Lomas F. Sonography in the diagnosis of carpal tunnel syndrome. AJR Am J Roentgenol 1999;173:681–684.

Federico P, Anna K, Rodrigo B, Bernt S. Alexander SH. Learning video object segmentation from static images. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA. : IEEE; 2017. p. 2500–3491.

Hadjerci Q, Hafiane A, Conte D, Makris P, Vieyres P, Delbos A. Computer-aided detection system for nerve identification using ultrasound images: A comparative study. Informatics Med Unlocked 2016;3:29–43.

Hadjerci Q, Hafiane A, Moretee N, Novales C, Vieyres P, Delbos D. Assistive system based on nerve detection and needle navigation in ultrasound images for regional anesthesia. Expert Syst Appl 2016;61:64–77.

Hafiane A, Vieyres P, Delbos A. Phase-based probabilistic active contour for nerve detection in ultrasound image for regional anesthesia. Comput Biol Med 2014;52:88–95.

Hafiane A, Vietres P, Delbos A. Deep learning with spatiotemporal consistence for nerve segmentation in ultrasound images. arXiv:170605870 [cs.CV], 2017.

Hammer HB, Hovden IA, Haavardsholm EA, Kvien TK. Ultrasonography shows increased cross-sectional area of the median nerve in patients with arthritis and carpal tunnel syndrome. Rheumatology 2006;45:584–588.

Hochreiter S, Schmidhuber J. Long short-term memory. Neural Computation 1997;9:1735–1780.

Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv:1502.03167, 2015.

Kakade A, Dumbali J. Identification of nerve in ultrasound images using U-Net architecture. 2018 International Conference on Communication Information and Computing Technology (ICCICT), February. Mumbai, India.

Kingma DP, Ba J. Adam: A method for stochastic optimization. The 3rd International Conference for Learning Representations. San Diego. arXiv:1412.6980v9 [cs.LG].

Li C, Xu C, Gui C, Fox MD. Distance regularized level set evolution and its application to image segmentation. IEEE Trans Image Process 2010;19:3243–3254.

Lin YH, Hsieh MY, Su FC, Wang SH. Assessment of the kinetic trajectory of the median nerve in the wrist by high-frequency ultrasound. Sensor 2014;14:7738–7752.

Liu S, Wang Y, Yang X, Lei B, Liu L, Li SX, Ni D, Wang T. Deep learning in medical ultrasound analysis: A review. Engineering 2019a;5:261–275.

Liu Z, Song YQ, Sheng VS, Wang L, Jiang R, Zhang X, Yuan D. Liver CT sequence segmentation based with improved U-Net and graph cut. Expert Syst Appl 2019b;126:54–63.

Perazzi F, Khoreva A, Benenson R, Schiele B, Sorkine-Horung A. Learning video object segmentation from static images. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA : IEEE; 2017. p. 2663–2672.

Reed P. Sample topic: Carpal tunnel syndrome. The medical disability advisor: Workplace guidelines for disability duration : Reed Group; 2005.

Romera-Paredes B, Torr PHS. Recurrent instance segmentation. In: Proceedings, the 14th European Conference on Computer Vision (ECCV 2016). Amsterdam, The Netherlands312–329. Part VI.

Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. In: Proceedings, 18th International Conference on Medical Image Computing and Computer-Assisted Intervention. Munich, Germany234–241.

Ruder S. An overview of gradient descent optimization algorithms. arXiv:1609.04747v2, 2017.

Yoshii Y, Villarraga HR, Henderson J, Zhao C, An KN, Amadio PC. Ultrasound assessment of the displacement and deformation of the median nerve in the human carpal tunnel with active finger motion. Journal Bone Joint Surg 2009;91:2911–2930.

Yoshii Y, Ishii T, Tung WL, Sakai S, Amadio PC. Median nerve deformation and displacement in the carpal tunnel during finger motion. J Orthop Res 2013;876–1880.