Localizing the Recurrent Laryngeal Nerve via Ultrasound with a Bayesian Shape Framework

Localizing the Recurrent Laryngeal Nerve via Ultrasound with a Bayesian Shape Framework

Haoran Dou[3,4⋆], Luyi Han[5,6⋆], Yushuang He[7], Jun Xu[2], Nishant Ravikumar[3,4], Ritse Mann[5,6], Alejandro F. Frangi[3,4,8,9,10], Pew-Thian Yap[11], and Yunzhi Huang[1(✉)]

[1] Institute for AI in Medicine, School of Automation, Nanjing University of Information Science and Technology, Nanjing, China
[2] Institute for AI in Medicine, School of Artificial Intelligence, Nanjing University of Information Science and Technology, Nanjing, China
[3] Centre for Computational Imaging and Simulation Technologies in Biomedicine (CISTIB), School of Computing, University of Leeds, Leeds, UK
[4] Biomedical Imaging Department, Leeds Institute for Cardiovascular and Metabolic Medicine (LICAMM), School of Medicine, University of Leeds, Leeds, UK
[5] Department of Radiology and Nuclear Medicine, Radboud University Medical Centre, Nijmegen, The Netherlands
[6] Department of Radiology, Netherlands Cancer Institute (NKI), Amsterdam, The Netherlands
[7] West China Hospital of Sichuan University, Chengdu, China
[8] Department of Cardiovascular Sciences, KU Leuven, Leuven, Belgium
[9] Department of Electrical Engineering, KU Leuven, Leuven, Belgium
[10] Alan Turing Institute, London, UK
[11] Department of Radiology and Biomedical Research Imaging Center (BRIC), University of North Carolina, Chapel Hill, USA
{yunzhi.huang.scu}@gmail.com

Abstract. Tumor infiltration of the recurrent laryngeal nerve (RLN) is a contraindication for robotic thyroidectomy and can be difficult to detect via standard laryngoscopy. Ultrasound (US) is a viable alternative for RLN detection due to its safety and ability to provide real-time feedback. However, the tininess of the RLN, with a diameter typically less than 3 mm, poses significant challenges to the accurate localization of the RLN. In this work, we propose a knowledge-driven framework for RLN localization, mimicking the standard approach surgeons take to identify the RLN according to its surrounding organs. We construct a prior anatomical model based on the inherent relative spatial relationships between organs. Through Bayesian shape alignment (BSA), we obtain the candidate coordinates of the center of a region of interest (ROI) that encloses the RLN. The ROI allows a decreased field of view for determining the refined centroid of the RLN using a dual-path identification network, based on multi-scale semantic information. Experimental results indicate that the proposed method achieves superior hit rates and substantially smaller distance errors compared with state-of-the-art methods.

[⋆] Haoran Dou and Luyi Han contributed equally to this work.

Keywords: Bayesian Shape Alignment · Recurrent Laryngeal Nerve · Localization in Ultrasound.

1 Introduction

Robotic thyroidectomy safely removes low-risk tumors and is preferred by patients who want a scarless operation with minimal invasiveness [7,17]. A complete pre-operative assessment of the surroundings of the thyroid is critical for accurate surgical planning to prevent unnecessary harm to the internal jugular vein or the recurrent laryngeal nerve (RLN). Currently, laryngoscopy is the only method surgeons can rely on to detect whether the RLN is tumor-infiltrated. However, laryngoscopy determines the RLN indirectly by assessing the activity of the vocal cords [1]. This approach is therefore highly inaccurate and can only distinguish RLN abnormality in about 1–3% patients [3]. Recent clinical trials have turned to ultrasound (US) as an alternative method for pre-operative inspection due to its safety and ability to provide real-time feedback [5].

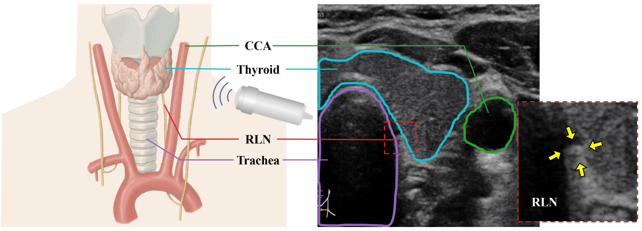


Fig. 1. Ultrasound imaging of bilateral RLNs.

The RLN in the US image is rather tiny relative to organs like the trachea and thyroid (Fig. 1). Therefore, automatic detection of the RLN from an US image is a challenging task. Recent studies [6,15,18] have demonstrated the feasibility of segmenting large nerves from US images, i.e., the sciatic nerve (diameter ranging from 16 to 20 mm [13]) and the median nerve (cross-sectional area ranging from 6.1 to 10.4 mm^2 [10]). These studies propose modifications to basic segmentation models [12,4] to improve segmentation accuracy. For example, van Boxtel et al. [15] investigated the efficacy of a hybrid model on nerve segmentation in US images. Horng et el. [6] integrated a ConvLSTM block [11] at the bottom layer of the U-Net to capture long-term spatial dependencies. Wu et al. [18] employed multi-size kernels and a pyramid architecture to aggregate features for segmentation. Despite these advances, localization of the RLN remains challenging since it is tiny with mean diameter ranging from 1 to 3 mm [16], significantly smaller than the larger nerves mentioned above.

In this work, we propose a knowledge-driven framework for RLN localization, mimicking the standard approach surgeons take to identify the RLN according to its surrounding organs. Our primary contributions are as follows: (1) We

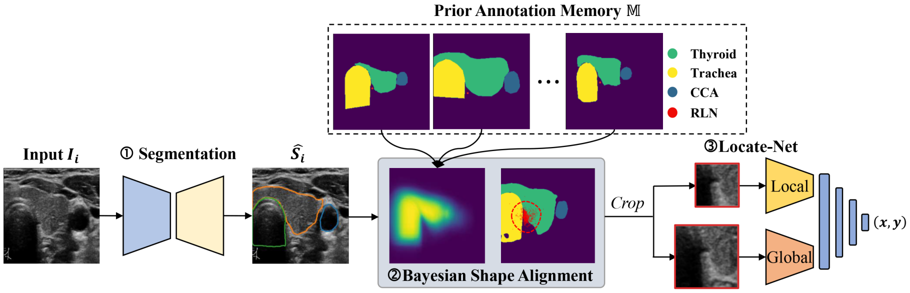


Fig. 2. Overview of the proposed framework.

propose the first learning-based framework to identify the RLN from a US image for pre-operative assessment of contraindication for robotic thyroidectomy; (2) We introduce Bayesian shape alignment for geometrical constraints, allowing the utilization of spatial prior knowledge in determining an ROI enclosing the RLN; (3) We introduce Locate-Net, a dual-path network that uses both local and global information to refine the localization of the RLN centroid.

2 Methods

Fig. 2 illustrates the proposed framework for identifying the RLN from a US image using anatomical prior knowledge. Our framework includes three coarse-to-fine sequential modules: (1) Segmentation module; (2) Bayesian shape alignment (BSA) module; and (3) Locate-Net module. The segmentation module obtains the segmentations $\hat{S}$ of organs surrounding the RLN, including the common carotid arteries (CCA), thyroid, and trachea. These segmentations form posteriors for the BSA module to infer the candidate coordinates of the RLN. Finally, Locate-Net refines the RLN centroid using local details and global contexts based on the patch centered at the inferred candidate coordinates. Details on the three modules are described in the following sections.

2.1 Bayesian Shape Alignment

To avoid missing the tiny RLN [16] in the midst of significantly larger structures, we introduce a method based on anatomical prior knowledge for RLN detection. Clinically, surgeons recognize the RLN based on its surrounding CCA, thyroid, and trachea [17]. The spatial relationships of these anatomical structures are typically consistent, but not entirely identical, across individuals. Here, we incorporate the spatial relationship into the Bayesian inference with the following mathematical model for a given image I:

$$q(\mathrm{RLN}, \mathrm{SO}|I) \propto p(I|\mathrm{RLN}, \mathrm{SO}) \times p(\mathrm{RLN}|\mathrm{SO}) \times p(\mathrm{SO}) \tag{1}$$

where SO is a matrix of pixels of a segmentation image, classifying whether each pixel belongs to the surrounding organs (CCA, trachea, thyroid); RLN refers to a set of center points, where each point is specified by a location vector (x, y) for the image matrix. $p(\mathrm{SO})$ is the prior distribution of the segmentation maps of the surrounding organs; and $p(\mathrm{RLN}|\mathrm{SO})$ is the likelihood of the RLN centroid given the segmented matrix of its surrounding organs. $p(\mathrm{SO})$ and $p(\mathrm{RLN}|\mathrm{SO})$ depend on the observed cohort and can be taken as prior knowledge. $p(I|\mathrm{RLN}, \mathrm{SO})$ represents the joint likelihood for the surrounding organs and the RLN's centroid, and is treated as a constant; and $q(\mathrm{RLN}, \mathrm{SO}|I)$ is the likelihood of the RLN centroid and the segmented matrix of its surrounding organs given a particular image I.

In Eq. 1, we aim to predict the RLN's centroid (x, y) given a image I by obtaining the maximum likelihood probability from the priors. The prior distributions for both the RLN centroids and its surrounding segmentation dependent on the given cohort. For each sample I_j from training set, the approximate probability $p_j(\mathrm{RLN}|\mathrm{SO})$ and $p_j(SO)$ attain the maximum value when I_j is most similar to the given sample I. Based on this, we can infer the likelihood RLN's centroids from the samples with similar surrounding segmentation matrix.

Prior Distribution for RLN's Surroundings Segmenting the organs surrounding the RLN is a prerequisite to determining $p(\mathrm{SO})$. Here, the widely adopted segmentation model, U-Net [12], is employed to segment the CCA, thyroid, and trachea from a US image. The segmentation network takes an US image and outputs the corresponding segmentation maps of the three organs. It comprises an encoder and decoder with a skip-connection to forward the feature representations from each stage of the encoder to the corresponding stage in the decoder. The numbers of feature maps in the encoder are 64, 128, 256, 512, 1024, and similarly in the decoder. Each stage in the segmentation network contains two convolution layers followed by instance normalization [14] and ReLU functions [19]. The training loss function is composed with a cross-entropy loss and a dice similarity coefficient (DSC) loss:

$$\mathcal{L}_{\mathrm{seg}} = \mathcal{L}_{\mathrm{ce}}(\hat{S}, M) + \mathcal{L}_{\mathrm{dsc}}(\hat{S}, M) \tag{2}$$

where $\mathcal{L}_{\mathrm{ce}}$ and $\mathcal{L}_{\mathrm{dsc}}$ refer to cross-entropy loss and DSC loss, respectively.

Alignment-based Likelihood Employing the prior sample I_j that is similar in the surrounding masks with the given image I_i to derive the RLN's centroid can derive higher likelihood $p(\mathrm{RLN}|\mathrm{SO})$. However, affected by the probe scanning angles θ, the observations belong to different angle distributions and can not be directly used to construct the priors of RLN's surrounding segmentation, hence, we embed a pre-alignment module to eliminate the influence of θ. The detailed implementation of the proposed alignment-based likelihood infer for the RLN is described in Algorithm 1.

Algorithm 1 Bayesian shape alignment during inference

```
Input: The predicted segmentation mask $\hat{S}_i$ for a inference sample, a set of mask
  labels $\mathbb{M} = \{M_j, j \in \mathbb{N}^*\}$ and RLN labels $\mathbb{C} = \{c_j, j \in \mathbb{N}^*\}$ for the training samples
Output: Candidate coordinate $\mathcal{C}_i$ of RLN for $\hat{S}_i$
  $D[*] = \{d_j, j \in \mathbb{N}^*\}$
  $C[*] = \{p_j, j \in \mathbb{N}^*\}$
  for corresponding label $\{M_j,c_j\}$ in $\{\mathbb{M}, \mathbb{C}\}$ do
    $\phi_j \leftarrow \mathrm{Affine}(M_j, \hat{S}_i)$                ▷ shape analysis
    $D[j] \leftarrow \mathrm{Dice}(M_j \circ \phi_j, \hat{S}_i)$        ▷ calculate dice metric
    $C[j] \leftarrow c_j \circ \phi_j$                       ▷ transform centroid of RLN
  end for
  $index \leftarrow \mathrm{Rank}(D[*])$                    ▷ rank dice in the descending order
  $C_{sort} \leftarrow \mathrm{Sort}(C[*], index)$            ▷ sort coordinates with dice rank
  $\mathcal{C}_i \leftarrow \mathrm{AverageTopK}(C_{sort})$   ▷ average the Top-$k$ candidate coordinates
```

Algorithm 1 estimates the likelihood RLN's centroid by searching the similar samples from the priors. We first align the priors to the common angle distribution. Taking each predicted mask $\hat{S}_i$ as the target segmentation and every transversed annotation M_j in the training dataset $\mathbb{M}$ as the moving segmentation, the affine transformation matrix ϕ_j with 6 degrees of freedom (DOF) can be conducted with the mutual information as the similarity metric. After such alignment, the Dice ratio (DSC) between the predicted mask $\hat{S}_i$ and each affined segmentation $M_j \circ \phi_j$ in the training dataset is calculated as the shape similarity score. More similar in the surrounding segmentation correspond to higher likelihood estimation for the entire image and the RLN's centroid. Subsequently, we descend the priors depending on the DSC, and average the RLNs' candidates with the Top-k DSCs to infer the position of RLN for the given image I_i.

2.2 Locate-Net

The BSA module is followed by a dual-path neural network to further refine the centroid of the RLN. Based on the candidate center coordinates of the RLN determined by the BSA module, a global patch with the size of 64×64 and a local patch with the size of 24×24 are cropped from the US image and jointly fed to the refinement network to provide global and local information. The features from the global patch contain global semantics to capture the potential outlier unobserved in the priors. The features from the local patch provide details to help refine the centroid.

In the refinement network, features from the local and global patches are separately extracted with two weight-shared sub-networks. Each sub-network contains three convolutional blocks. After each convolution block, the feature maps are down-sampled with a factor of 2 via the max-pooling layer. The detailed composition of each convolutional block is the same with the encoder in the segmentation network (seen in Section 2.1). The outputs of the two sub-networks are then re-scaled to the same size with a adaptive pyramid pooling layer [20], and followed with a concatenation layer and two convolution blocks. The top of the refinement network is three fully connected layers with the hidden units of 512, 64, 2, so that to predict the refined centroid of RLN. The regression loss function for training the refinement network is defined as:

$$\mathcal{L}_{\text{reg}} = \sum_{i} \mathcal{L}_{s1}(\hat{c}_i - c_i), \tag{3}$$

where $\mathcal{L}_{s1}$ is the smoothed L_1 loss with β of 1.0 indexed for the difference of image pixel $(\Delta x, \Delta y)$, $\hat{c}$ and c denote as the predicted RLN centroid and the ground truth, respectively.

$$\mathcal{L}_{s1}(\Delta x, \Delta y) = \begin{cases} \frac{1}{2\beta}(\Delta x^2 + \Delta y^2) & |\Delta x| + |\Delta y| < 2\beta, \\ |\Delta x| + |\Delta y| - \beta & \text{others.} \end{cases} \tag{4}$$

3 Experiments

3.1 Dataset and Evaluation Metrics

2D ultrasound images were collected with an Aixplorer color Doppler ultrasound device (Hologic Supersonic imagine, AIX en Provence), equipped with a linear array probe with a frequency of 4–15 MHz. A total of 465 patients diagnosed with thyroid cancer by preoperative biopsy and enrolled for thyroidectomy participated in this study. Each patient has both left and right scans of the RLN. Each scan contains a variable number of qualified US frames, ranging from 1 to 4. 325, 46 and 94 subjects were randomly selected for training, validation, and testing, respectively. All images were resampled to a common size of 256 × 256. Manual annotation was contented by three clinical experts and passed through strict quality control from a senior expert.

Performance was quantified using the absolute distance error and the hit rate between the predicted RLN centroid and the ground truth:

$$\texttt{Dis}(\hat{c}, c) = \|\hat{c} - c\|_1 \tag{5}$$

$$\texttt{Hit}(\hat{c}, c) = \begin{cases} 1 & \hat{c} \subseteq N_c, \\ 0 & \text{others,} \end{cases} \tag{6}$$

where $\hat{c}$ and c denote as the predicted RLN centroid and the ground truth, respectively. N_c indicates the neighborhood of c with radius r_θ, which is set to be 15 pixels in this work. Lower distance error and higher hit rate correspond to better identification performance.

3.2 Implementation Details

We implemented our method using PyTorch on the Google Colab platform with an NVIDIA Tesla P100 GPU. We trained the segmentation network using the

Table 1. Statistics of competing methods for the testing dataset.

	Methods	Left RLN		Right RLN	
		Distance (*pix*)	Hit Rate (%)	Distance (*pix*)	Hit Rate (%)
Coord-based	ResNet-50 [4]	10.9 ± 9.7	77.5	12.3 ± 8.4	70.0
	SwinT-C [8]	19.7 ± 11.8	42.3	14.4± 9.6	59.4
	ConvNeXt-C [9]	16.5 ± 8.2	47.3	14.0 ± 9.5	62.5
Heatmap-based	U-Net [12]	29.3 ± 12.8	11.5	20.9 ± 10.5	31.9
	DeepLab [2]	17.5 ± 8.0	41.2	11.9 ± 7.1	71.3
	SwinT-H [8]	22.7 ± 13.6	30.8	20.2 ± 10.6	35.6
	ConvNeXt-H [9]	12.7 ± 12.1	73.1	13.1 ± 8.7	64.4
	Proposed	**3.49±7.53**	**95.6**	**4.55 ±7.61**	**92.5**

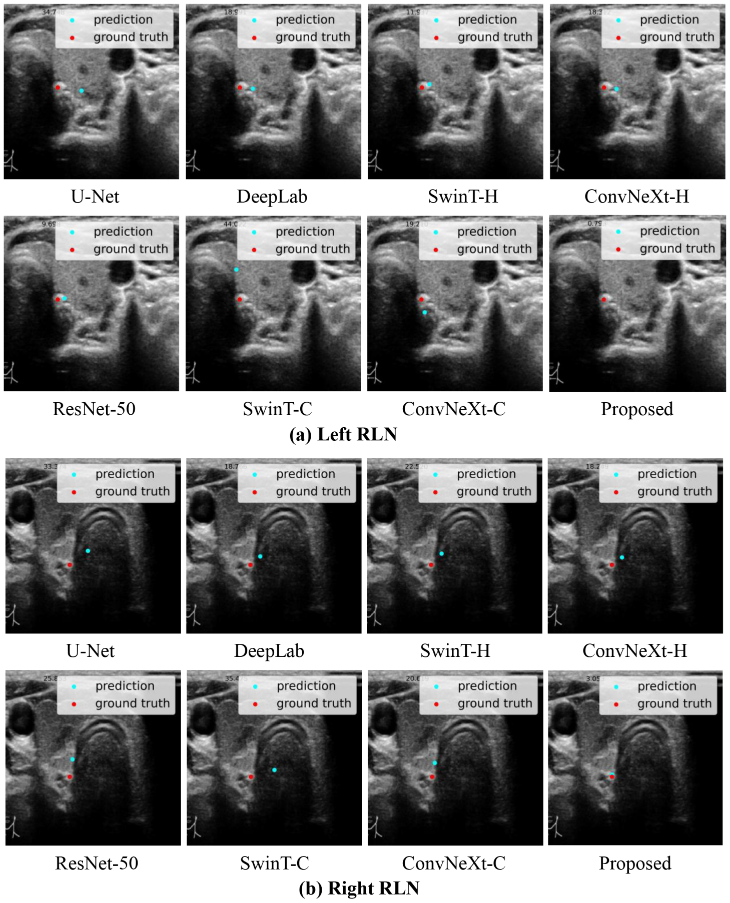
Adam optimizer with an initial learning rate of 3×10^{-4} and a batch size of 16 for 100 epochs, taking about 3 hours. The learning rate was decayed every epoch with a factor of 0.9. The affine matrix in Bayesian shape alignment was initially computed with the center of mass of the masks and iteratively refined with the mutual information metric. For the training of the dual-path refinement network, the learning rate and batch size were set to 1×10^{-3} and 16, respectively. The code for the techniques presented in this study can be found at: https://github.com/wulalago/RLNLocalization

3.3 Comparison Baselines

We compared our method with coordinate and heatmap regression methods. Compared heatmap regression methods include U-Net [12], DeepLab [2], SwinT-H [8], and ConvNeXt-H [9]. Coordinate regression methods include ResNet-50 [4], SwinT-C [8], and ConvNeXt-C [9]. The optimal hyper-parameters of all the baseline methods are obtained based on the grid search strategy. Fig. 3 shows example cases of the bilateral RLNs given by the the baseline methods. The red and cyan circles mark the annotated ground truth and the predicted centroid of the RLN, respectively. Our method predicts the centroids of the bilateral RLNs with higher accuracy than the baseline methods. Table 1 reports the statistics of the results given by the methods on the testing dataset. The proposed method achieves the lowest distance error with the highest hit rate.

3.4 Ablation Study

We compared three types of centroid refinement methodes, including (1) refinement with local information; (2) refinement with global contextual information; and (3) refinement with both local and global features. Table 2 reports the distance errors and hit rates for these settings based on the proposed method. Refinement with local and global information yields the lowest distance error with the highest hit rate.



(a) Left RLN

(b) Right RLN

Fig. 3. Example results from competing methods.

4 Conclusion

Inspired by the way surgeons to recognize RLN, we developed a prior knowledge driven framework to automatically identify the tiny RLN from US images. In the proposed pipeline, we first segment the large organs surrounding RLN as the conditional prior, and then using the Bayesian shape alignment model to determine the candidate coordinate close to RLN. Then following the Locate-Net to refine the centriod of RLN with multi-scales patches to extract the local information and global context around the RLN. Leveraging the spatial relationship between RLN and its surrounding organs as the prior constraint, our model can avoid the tiny RLN being submerged the background. From Tabel 1 and Tabel 2, we can conclude that, any combination of our framework achieves the superiority in the distance error and hit rate as compared to the recent coordinate or heatmap regression models.

Table 2. Statistics for different settings based on the testing dataset.

Methods	Left RLN		Right RLN	
	Distance (*pix*)	Hit Rate (%)	Distance (*pix*)	Hit Rate (%)
Initialization	7.52±7.65	89.0	9.70±7.03	84.9
+ Local information	4.45±8.79	91.8	5.47±7.90	88.7
+ Global context	3.58±7.51	94.5	4.42±7.48	91.8
+ Local & global features	**3.49±7.53**	**95.6**	4.55 ±7.61	**92.5**

Acknowledgement This work was supported by the National Natural Science Foundation of China under Grant No. 62101365 and the startup foundation of Nanjing University of Information Science and Technology. Haoran Dou was funded by the Royal Academy of Engineering Chair in Emerging Technologies Scheme (CiET1819/19). Luyi Han was funded by Chinese Scholarship Council (CSC) scholarship. Alejandro F. Frangi was funded by the Royal Academy of Engineering (INSILEX CiET1819/19), Engineering and Physical Sciences Research Council (EPSRC) programs TUSCA EP/V04799X/1, and the Royal Society Exchange Programme CROSSLINK IES\NSFC\201380. Jun Xu was funded by the National Natural Science Foundation of China (Nos. U1809205, 62171230, 92159301, 61771249, 91959207, 81871352).

References

1. Chandrasekhar, S.S., Randolph, G.W., Seidman, M.D., Rosenfeld, R.M., Angelos, P., Barkmeier-Kraemer, J., Benninger, M.S., Blumin, J.H., Dennis, G., Hanks, J.B., Haymart, M.R., Kloos, R.T., Seals, B., Schreibstein, J.M., Thomas, M.A., Waddington, C., Warren, B., Robertson, P.J.: Clinical practice guideline: Improving voice outcomes after thyroid surgery. Otolaryngology-Head and Neck Surgery **148** (2013)
2. Cheng, B., Collins, M.D., Zhu, Y., Liu, T., Huang, T.S., Adam, H., Chen, L.C.: Panoptic-deeplab: A simple, strong, and fast baseline for bottom-up panoptic segmentation. In: Computer Vision and Pattern Recognition (2020)
3. Dionigi, G., Boni, L., Rovera, F., Rausei, S., Castelnuovo, P., Dionigi, R.: Postoperative laryngoscopy in thyroid surgery: proper timing to detect recurrent laryngeal nerve injury. Langenbeck's Archives of Surgery **395**, 327–331 (2010)
4. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Computer Vision and Pattern Recognition (2016)
5. He, Y., Li, Z., Yang, Y., Lei, J., Peng, Y.L.: Preoperative visualized ultrasound assessment of the recurrent laryngeal nerve in thyroid cancer surgery: Reliability and risk features by imaging. Cancer management and research **13**, 7057–7066 (2021)
6. Horng, M.H., Yang, C.W., Sun, Y.N., Yang, T.H.: Deepnerve: A new convolutional neural network for the localization and segmentation of the median nerve in ultrasound image sequences. Ultrasound in Medicine and Biology **46**, 2439–2452 (2020)

7. Lee, J., Chung, W.Y.: Robotic surgery for thyroid disease. European thyroid journal **2**, 93–101 (2013)
8. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin transformer: Hierarchical vision transformer using shifted windows. arXiv: Computer Vision and Pattern Recognition (2021)
9. Liu, Z., Mao, H., Wu, C.Y., Feichtenhofer, C., Darrell, T., Xie, S.: A convnet for the 2020s. arXiv preprint arXiv:2201.03545 (2022)
10. Meyer, P., Lintingre, P.F., Pesquer, L., Poussange, N., Silvestre, A., Dallaudière, B.: The median nerve at the carpal tunnel ... and elsewhere. Journal of the Belgian Society of Radiology **102**, 17–17 (2018)
11. Romera-Paredes, B., Torr, P.H.S.: Recurrent instance segmentation (2015)
12. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
13. Tiel, R., Filler, A.G.: Nerve injuries of the lower extremity (2011)
14. Ulyanov, D., Vedaldi, A., Lempitsky, V.: Instance normalization: The missing ingredient for fast stylization. arXiv preprint arXiv:1607.08022 (2016)
15. Van Boxtel, J., Vousten, V., Pluim, J., Rad, N.M.: Hybrid deep neural network for brachial plexus nerve segmentation in ultrasound images. In: 2021 29th European Signal Processing Conference (EUSIPCO). pp. 1246–1250. IEEE (2021)
16. Wojtczak, B., Kaliszewski, K., Sutkowski, K., Bolanowski, M., Barczyński, M.: A functional assessment of anatomical variants of the recurrent laryngeal nerve during thyroidectomies using neuromonitoring. Endocrine **59**, 82–89 (2018)
17. Wong, K.P., Lang, B.H.H.: Endoscopic thyroidectomy: A literature review and update. Current Surgery Reports **1**, 7–15 (2013)
18. Wu, H., Liu, J., Wang, W., Wen, Z., Qin, J.: Region-aware global context modeling for automatic nerve segmentation from ultrasound images. In: National Conference on Artificial Intelligence (2021)
19. Xu, B., Wang, N., Chen, T., Li, M.: Empirical evaluation of rectified activations in convolutional network. arXiv preprint arXiv:1505.00853 (2015)
20. Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J.: Pyramid scene parsing network. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2881–2890 (2017)