DeepEdge: automatic ultrasound tongue contouring combining a deep neural network and an edge detection algorithm

Wei-Rong Chen[1], Mark Tiede[1], D.H. Whalen[1,2,3]

[1]Haskins Laboratories, [2]City University of New York, [3]Yale University

Ultrasound imaging (USI) of the tongue has been widely used for research purpose as well as in clinical trials (e.g., visual biofeedback). However, USI contouring is a labor-intensive task. Recent improvements in automatic methods of USI contouring have focused on two directions: 1) edge detection/tracking algorithms, such as EdgeTrak (SNAKE-based edge detection) [1], and SLURP [2]; and 2) deep neural network models [e.g., 3, 4, 5]. Improved edge detection algorithms can yield accurate contours, but only when (semi-) manually-generated "seeds" provide the (nearly) true tongue edge. They are also usually vulnerable to noise (e.g., ultrasound speckles). A deep neural network, on the other hand, is a powerful tool for recognizing broad region of interests (ROI) surrounding the tongue edge and robust to noise, but current models have much lower spatial resolution (e.g., 16x32 pixels in [3]; 128x128 pixels in[4]) than edge detection algorithms, due to a trade-off between image size and network robustness: The larger the image, the less robust the model. Here we present a hybrid method which uses a deep neural network to determine an ROI and then detects the tongue surface within the ROI using a SNAKE-based algorithm. A MATLAB-based graphical user interface (GUI) of our proposed method, dubbed 'DeepEdge', is provided through Github.

We acquired 2189 frames of manually labelled USI [6] as our training dataset, and another set of 720 frames [7] as test data. Images were down-sampled to 64x64 pixels before training. We adopted U-Net (a convolution neural network-based architecture commonly used in medical image contouring) [8] as the architecture of our neural network (NN) model. The trained NN model predicts the probability of each pixel that contains the tongue edge, shown as the white 'blobs' in Fig. 1b. Previous studies only considered the largest blob to be the one containing the tongue edge; this ignores the common case of discontinuous tongue edge due to noise. Thus, we retain blobs larger than 20% of the largest blob and remove any blob that is below or above a larger blob (Fig. 1c). The output of blob analysis is "skeletonized" (Fig. 1d) and up-sampled to the original image resolution; then a SNAKE-base edge detection snaps the skeletonized thin line to the nearby tongue surface (red dots in Fig. 1e).

Fig. 2 demonstrates the GUI of our method. A video file of USI is first loaded and then cropped to exclude excessive meta-information. The final output of the GUI can be exported to either a text file or MATLAB data structure that is ready to be further edited or manually-corrected in other tools, such as GetContours [9]. We evaluated the accuracy of our method by calculating the mean sum of distances (MSD) [2, 3] with the test dataset. The average error (against human annotators) of our method was 1.44 mm (SD=1.42mm) across 720 images. Edge detection combined with the results of the neural network model showed improvements (over neural network model alone) only qualitatively by visual inspection, but not in the calculated error measurements; improved error metrics for USI [e.g., 7] are needed in future studies.

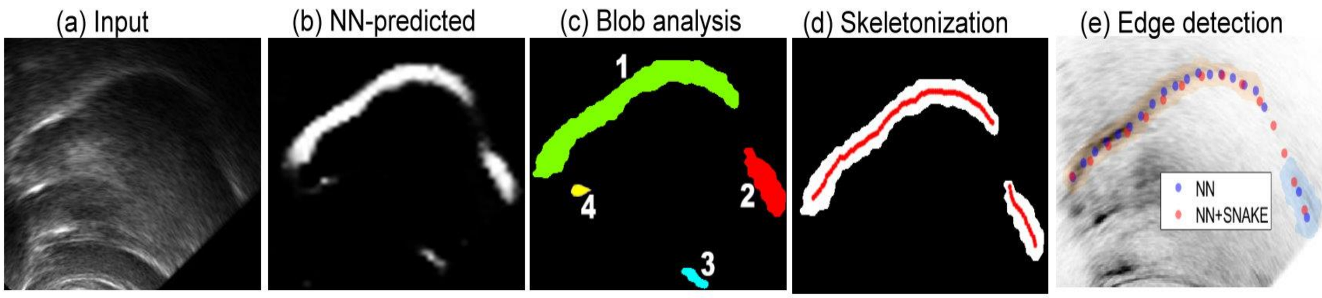


Figure 1. (a) input image; (b) NN-model predicts "blobs" of ROI containing tongue edge; (c) blob analysis retains the first two largest blobs; (d) Skeletonization; and (e) SNAKE edge detection refines the output of the deep neural network model.

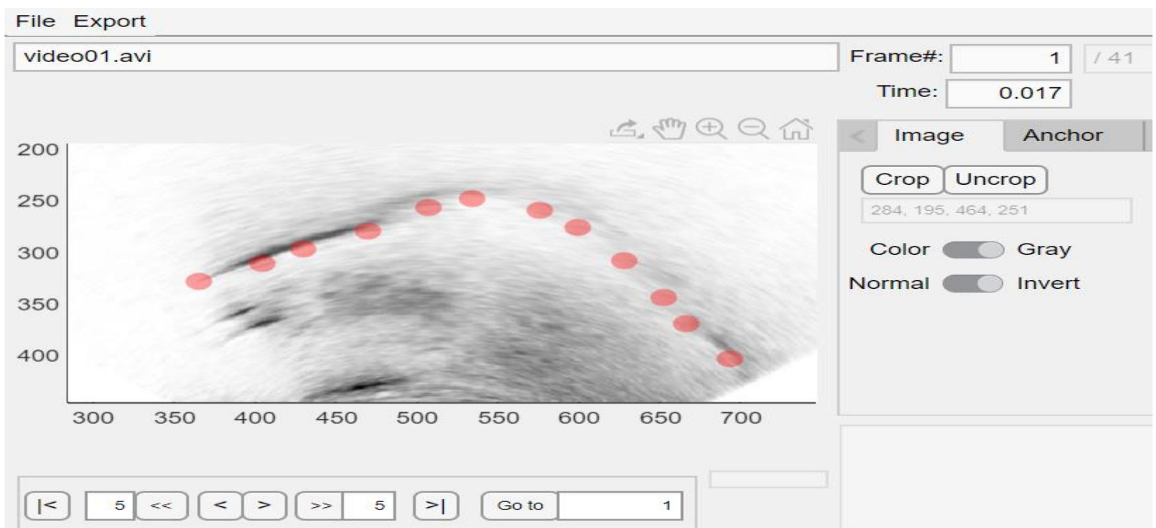


Figure 2. A MATLAB-based GUI of the proposed hybrid method.

## References


1. Li, M., C. Kambhamettu, and M. Stone, *Automatic contour tracking in ultrasound images.* Clinical Linguistics & Phonetics, 2005. **19**(6-7): p. 545-554.
2. Laporte, C. and L. Ménard, *Multi-hypothesis tracking of the tongue surface in ultrasound video recordings of normal and impaired speech.* Medical Image Analysis, 2018. **44**: p. 98-114.
3. Fasel, I. and J. Berry. *Deep Belief Networks for Real-Time Extraction of Tongue Contours from Ultrasound During Speech*. in *20th ICPR*. 2010.
4. Zhu, J., W. Styler, and I.C. Calloway, *Automatic tongue contour extraction in ultrasound images with convolutional neural networks.* JASA, 2018. **143**(3): p. 1966-1966.
5. Mozaffari, M.H. and W.-S. Lee, *BowNet: Dilated Convolution Neural Network for Ultrasound Tongue Contour Extraction.* arXiv preprint, 2019.
6. Kim, B., M.K. Tiede, and D.H. Whalen. *Evidence for pivots in tongue movement for diphthongs*. in *ICPhS*. 2019. Melbourne, Australia.
7. Whalen, D.H., et al. *Accuracy assessments of hand and automatic measurements of ultrasound images of the tongue*. in *ICPhS*. 2019. Melbourne, Australia.
8. Ronneberger, O., P. Fischer, and T. Brox. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. 2015. Cham: Springer International Publishing.
9. Tiede, M. K. (2018). GetContours (Version 1.3). Retrieved from https://github.com/mktiede/GetContours.