Anatomical Feature-Based Lung Ultrasound Image Quality Assessment Using Deep Convolutional Neural Network

Surya M. Ravishankar[1], Ryosuke Tsumura[1], John W. Hardin[2], Beatrice Hoffmann[2], Ziming Zhang[1], Haichong K. Zhang[1,*]

[1]Worcester Polytechnic Institute, Worcester, USA

[2]Beth Israel Deaconess Medical Center, Boston, USA

[*]hzhang10@wpi.edu

Abstract— **Lung ultrasound (LUS) has been used for point-of-care diagnosis of respiratory diseases including COVID-19, with advantages such as low cost, safety, absence of radiation, and portability. The scanning procedure and assessment of LUS are highly operator-dependent, and the appearance of LUS images varies with the probe's position, orientation, and contact force. Karamalis et al. introduced the concept of ultrasound confidence maps based on random walks to assess the ultrasound image quality algorithmically by estimating the per-pixel confidence in the image data. However, these confidence maps do not consider the clinical context of an image, such as anatomical feature visibility and diagnosability. This work proposes a deep convolutional network that detects important anatomical features in an LUS image to quantify its clinical context. This work introduces an Anatomical Feature-based Confidence (AFC) Map, quantifying an LUS image's clinical context based on the visible anatomical features. We developed two U-net models, each segmenting one of the two classes crucial for analyzing an LUS image, namely 1) Bright Features: Pleural and Rib Lines and 2) Dark Features: Rib Shadows. Each model takes the LUS image as input and outputs the segmented regions with confidence values for the corresponding class. The evaluation dataset consists of ultrasound images extracted from videos of two sub-regions of the chest above the anterior axial line from three human subjects. The feature segmentation models achieved an average Dice score of 0.72 on the model's output for the testing data. The average of non-zero confidence values in all the pixels was calculated and compared against the image quality scores. The confidence values were different between different image quality scores. The results demonstrated the relevance of using an AFC Map to quantify the clinical context of an LUS image.**

Keywords—Lung Ultrasound; Confidence Map; Image Quality; Deep Learning

I. INTRODUCTION

Respiratory symptoms are the primary manifestation of infectious respiratory diseases, which in turn are frequently caused by respiratory viruses that infect the cells in the airways of the nose, throat, and lungs. These infections can cause myriad illnesses such as flu (influenza), severe acute respiratory syndrome (SARS), and COVID-19. Novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has already caused a pandemic of epic proportions, affecting over 8 million humans in an estimated 100 countries [1]. Ultrasound imaging is used to diagnose various lung diseases such as pulmonary edema, pulmonary embolism, pneumonia, asthma, and pneumothorax [2]. Lung ultrasound (LUS) has emerged as a point-of-care alternative to chest CT and x-ray for rapid point-of-care diagnosis with its advantage of low cost, enhanced safety, absence of radiation, and portability for bedside diagnosis [3-5]. It has been widely adopted for imaging COVID-19 patients with clinical guidelines already having been proposed [6-9]. However, LUS diagnostic accuracy is highly operator-dependent, affected by the diagnosability of the acquired LUS images.

The LUS is acquired from the anterior, lateral, and posterior regions of the chest. An LUS image does not contain structured organs since the lung is filled with air that reflects the ultrasound wave. We will be able to see the lung only when there is consolidation or fluid in the lung. The presence of A-lines, which appear as horizontal lines, indicate dry interlobular septa and removes the concerns for pulmonary edema. The presence of B-Lines, which appear as white lines from the pleura to the bottom of the image, means abnormal aeration, suggests pulmonary edema, and excludes pneumothorax. Lung sliding is defined by a subtle shimmering seen at the pleural line and needs to be observed to exclude pneumothorax with certainty. Other artifacts can be seen in a lung ultrasound, such as lung point, comet tail, ring down, and mirror image, etc. Though these different features are required for the diagnosis of LUS, the most important features that are always essential and present in any of these LUS images are the bright feature formed by pleural lines and rib lines, and dark feature represents rib shadows because the rib does not allow ultrasound to pass through. Thus, this paper aims to recognize these features to determine image diagnosability.

Karamalis et al. introduced the concept of ultrasound confidence maps based on random walks to assess the ultrasound image quality algorithmically by estimating the per-pixel confidence in the image data [10]. Yet, these confidence maps do not consider the clinical context, such as anatomical feature visibility and diagnosability. Toward recognizing the anatomical landmarks, Roy et al. explored deep learning for classification and localization of COVID-19 markers in point-

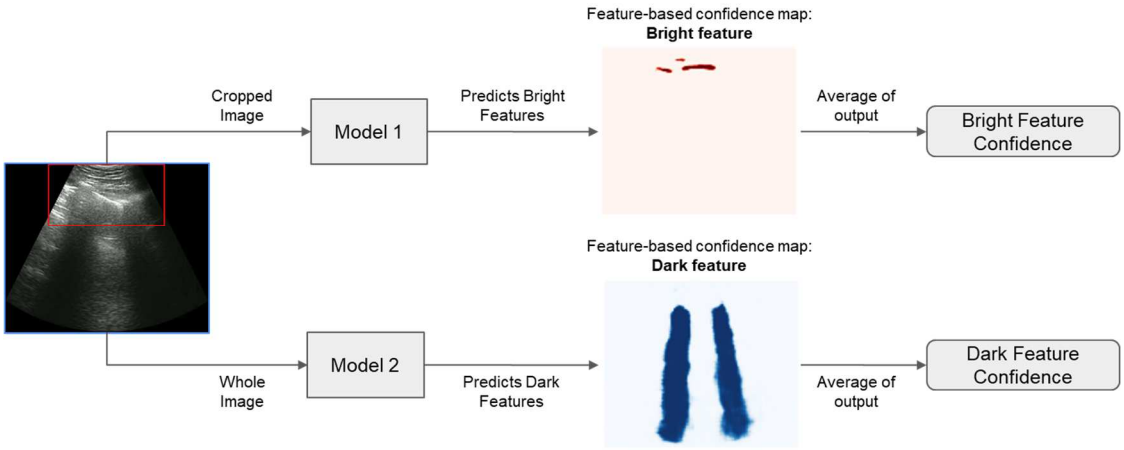

Fig. 1 The flowchart presents the cropped input image as input to Model 1 and the whole image as input to Model 2 to provide a segmentation output of bright (pleural and rib lines) and dark features (rib shadows) as the Anatomical Feature-based Confidence (AFC) map.

of-care LUS, presenting a deep learning architecture to predict pathological scores from an LUS image sequence and optimally fuses them to produce a disease severity score for each video [11]. In another work, Smistad et al. introduced highlighting nerves and blood vessels for ultrasound-guided axillary nerve block procedures using neural networks [12] by using a convolutional U-net neural network to identify the musculocutaneous, median, ulnar, radial nerves, and blood vessels in ultrasound images.

In this paper, we introduce a deep convolutional neural network that detects critical anatomical features in an LUS image to quantify its clinical context. This work introduces an Anatomical Feature-based Confidence (AFC) Map, quantifying an LUS image's clinical context based on the visible anatomical features.

II. Methods

A. Model Definition

We developed two deep learning models, each segmenting one of the two features crucial for analyzing an LUS image, namely 1) Bright features: pleural and rib lines, and 2) Dark features: rib shadows. Each model takes a single LUS image as input and outputs the corresponding segmented regions. Each pixel in the output can be taken as the confidence in the original image belonging to the corresponding class. **Figure 1** shows the flow of the data from input to output the AFC values. Two independent models taking varied image size data are used. When the whole image is resized to be fed to the model, bright features are almost invisible, making it hard for the model to recognize it. The input to the bright feature model is cropped out from the whole image and focuses on the top part of the LUS to help recognize the bright features. The dark feature model takes in the whole image as input and outputs the segmentation for dark features.

Both the model architectures are identical and consist of an encoder stage and a decoder stage that finally outputs a segmentation of two classes: the corresponding feature and background. As shown in **Fig. 2**, the encoder stage performs two 3x3 convolutions in each layer followed by 2x2 max pooling. The decoder stage contains a 2x2 up-convolution layer followed by two 3x3 convolutions. The number of filters in convolution layers were doubled after each stage, starting from 8 filters in the encoding stage and ending at 128 (5 stages). The filters were halved after each stage in the decoding stage and eventually a 1x1 convolution to output the segmentation. High-resolution features from the downsampling path are combined with the upsampled images to localize. Successive convolution layers will learn to assemble more precise output based on this information [13]. The cropped image input to the first model lets the bright features be visible even after resizing and thus improves the prediction of these features

The output of segmentation consists of two classes for each feature: (1) Bright features and (2) background for the first feature, and (1) dark features and (2) background for the second feature. The output values in the first layer are in the range 0 to 1, which could also mean the AFC of a pixel to be a bright/dark feature. The average of all the values above 0.01 is taken and considered the output AFC for the corresponding feature. The output is multiplied by 255 to visualize the AFC maps.

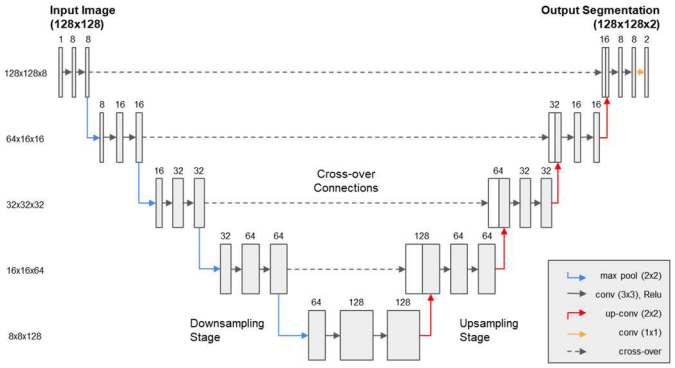

Fig. 2 The U-net architecture. Each grey box corresponds to a feature map. The number of channels is specified on top of each box—the white boxes denote the copied features. The operations denoted by each arrow are mentioned in the image.

B. Loss Definition

There is a class imbalance problem because most of the pixels belong to the background class. We used the Dice loss

function to resolve it. The models have Dice loss for the first class (bright/dark features) and categorical cross-entropy loss for the background.

C. Data Collection and Annotation

The dataset consists of ultrasound collected from three human subjects in two sub-regions of the chest above the anterior axial line on the left and right sides [14]. A total of 312 ultrasound images were extracted from these videos that were used for diagnosing the lung. The bright and dark features were hand-labeled on 312 images as ground truth for the model. In addition, the LUS videos were scored from 1 to 5 to quantify the quality of clinical context appearance for each feature class. The data was split into 66% for training and 34% for testing.

D. Augmentation, Flipping, Rotation, and Elastic Deformation

The data is augmented to avoid overfitting the training data and allow the model to generalize better for unseen data. This also compensates for the few training samples available. The ground truth for segmentation was transformed with the same augmentation along with the images. The images were flipped horizontally with a probability of 0.5. Additionally, the images were rotated from a random angle between -10 ~ 10 degrees. This would let the model learn rotational invariance and help the model learn to recognize features from different probe orientations. Finally, elastic deformation was applied to the images with a random 3x3 grid and randomly selected sigma value from 0 to 5.

E. Training and Evaluation

The model was trained with Adam optimizer parameters on Tensorflow version 1.14, Keras version 2.3.1, Ubuntu 18.04, NVIDIA RTX 2060 for 20000 epochs, with a batch size of 4. Training time was 16 hours with 5 networks training in parallel.

Dice score was used to evaluate the performance of the segmentation model. The correlation of the output AFC of the models with the labeled scores was used to determine the model's performance in predicting the visibility of the bright and dark features.

III. Results

The feature segmentation models achieved an average Dice score of 0.72 on the testing data output. The example outputs of AFC maps are shown in **Fig. 3**.

The average of non-zero AFC values in all the pixels was calculated and plotted against the image quality scores in **Figs. 4(a-b)**. The confidence values were different between scores with correlation, especially in the rib shadow detection. The results demonstrated the relevance of using an AFC Map to quantify the clinical context of an LUS image. A similar analysis is performed using the conventional confidence map [10]. The results are plotted in **Fig. 4(c-d)**. While the variations of confidence values between different visibility scores are observed, the trend is less obvious compared to the AFC map.

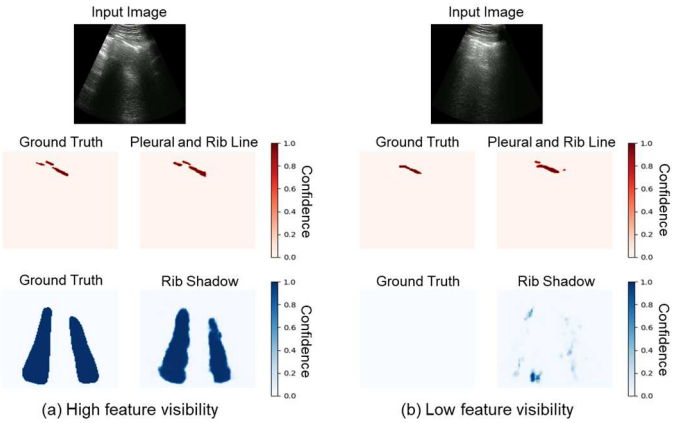

Fig. 3 Example outputs of the models: **(a)** An image with high feature visibility, and **(b)** an image with low feature visibility.

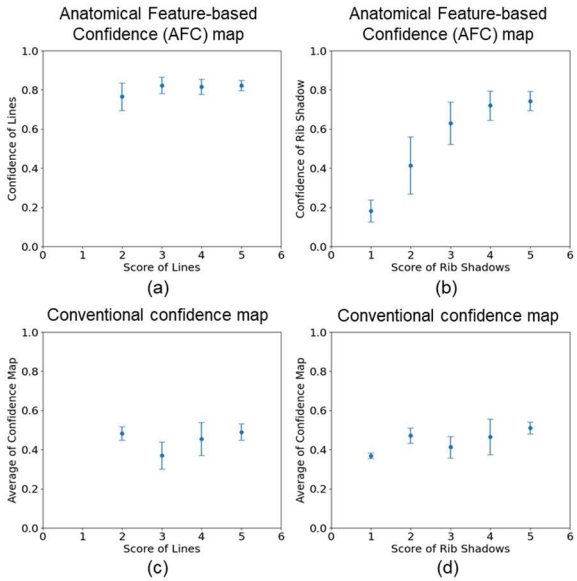

Fig. 4 (a) Average AFC values using Model 1 with respect to the ground truth bright feature visibility scores, and **(b)** Average AFC values using Model 2 with respect to the ground truth dark feature visibility scores. Corresponding comparison results using the conventional confidence map [10] for bright and dark features are shown in **(c-d)**, respectively.

IV. Discussion and Conclusions

The AFC map based on the U-net architecture performs well on the given dataset and provides valuable information about the diagnosability of lung ultrasound images. The results show that the model could recognize the critical features from a lung ultrasound image. The AFC map could predict the visibility of a given feature in a given image.

One potential use of the AFC map is to guide a human operator by providing feedback to optimize the probe position and orientation. For example, if bright features are visible in the center of an ultrasound image between two rib shadows, the probe position likely is at the right location to observe the pleural movements. Recognizing such features could aid in identifying which direction the probe should go towards to acquire good images. If the position and orientation are proper while the confidence value is still low, it may suggest increasing

the contact force to improve the image quality. This model could also provide feedback and guide a robot arm to acquire lung ultrasound images. This could be especially handy when diagnosing diseases like COVID-19 that have a high risk for the operator. It could be used to provide valuable insights in a teleoperation setting [14]. Conventional confidence maps have been used for robot control [10]. Similarly, the proposed AFC map could be used to attain better quality images, especially ensuring the inclusion of clinically relevant anatomical features.

The future scopes could be studying how the change in the probe's position, orientation, and force would affect the output AFC of the features from the models. With more data from many human subjects, the model could be improved in both training and testing. The current dataset consists only of images of healthy lungs, and most of these images contain good quality images with the features visible. Hence collecting more diverse data could improve the model. This paper has considered only two anterior regions in the chest for training the model. This work could be extended to other regions of the chest to observe its performance variation.

ACKNOWLEDGMENTS

This work was supported by the Worcester Polytechnic Institute internal fund and the National Institutes of Health funding [grant number: DP5 OD028162].

REFERENCES

[1] "Home - Johns Hopkins Coronavirus Resource Center." https://coronavirus.jhu.edu/ (accessed Feb. 12, 2021).

[2] L. Gargani and G. Volpicelli, "How I do it: lung ultrasound," *Cardiovascular Ultrasound*, vol. 12, no. 1, pp. 1-10, 2014.

[2] Y. Xia, Y. Ying, S. Wang, W. Li, and H. Shen, "Effectiveness of lung ultrasonography for diagnosis of pneumonia in adults: A systematic review and meta-analysis," *J. Thorac. Dis.*, vol. 8, no. 10, pp. 2822–2831, 2016.

[3] D. A. Lichtenstein and G. A. Mezière, "Relevance of lung ultrasound in the diagnosis of acute respiratory failure the BLUE protocol," *Chest*, vol. 134, no. 1, pp. 117–125, 2008.

[4] Q. Y. Peng, X. T. Wang, and L. N. Zhang, "Findings of lung ultrasonography of novel corona virus pneumonia during the 2019–2020 epidemic," *Intensive Care Medicine*, vol. 46, no. 5, pp. 849–850, 2020.

[5] V. Manivel, A. Lesnewski, S. Shamim, G. Carbonatto, and T. Govindan, "CLUE: COVID-19 lung ultrasound in emergency department," *EMA - Emerg. Med. Australas.*, vol. 32, no. 4, pp. 694–696, 2020.

[6] S. Moore and E. Gardiner, "Point of care and intensive care lung ultrasound: A reference guide for practitioners during COVID-19," *Radiography*, vol. 26, no. 4, pp. e297-e302, 2020.

[7] G. Soldati, *et al.*, "Proposal for International Standardization of the Use of Lung Ultrasound for Patients With COVID-19," *J. Ultrasound Med.*, vol. 39, no. 7, pp. 1413–1419, 2020.

[8] F. Moro, *et al.*, "How to perform lung ultrasound in pregnant women with suspected COVID-19," *Ultrasound Obstet. Gynecol.*, vol. 55, no. 5, pp. 593–598, 2020.

[9] A. Karamalis, W. Wein, T. Klein, and N. Navab, "Ultrasoundconfidence maps using random walks," *Medical Image Analysis*, vol. 16, no. 6, pp. 1101–1112, 2012.

[10] S. Roy, *et al.*, "Deep learningfor classification and localization of covid-19 markers in point-of-carelung ultrasound," *IEEE Transactions on Medical Imaging*, vol. 39, no. 8, pp. 2676–2687, 2020.

[11] E. Smistad, K. Johansen, D. Iversen, and I. Reinertsen, "Highlightingnerves and blood vessels for ultrasound-guided axillary nerve blockprocedures using neural networks," *Journal of Medical Imaging*, vol. 5, no. 4, pp. 044004, 2018.

[12] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Springer, pp. 234-241, 2015..

[13] R. Tsumura, *et al.*, "Tele-Operative Low-Cost Robotic Lung Ultrasound Scanning Platform for Triage of COVID-19 Patients," IEEE Robotics and Automation Letters, vol. 6, no. 3, pp. 4664-4671, 2021.