LESION SIZE ASSESSMENT ON MULTI-FREQUENCY SINGLE TRANSDUCER-HARMONIC MOTION IMAGING USING CONVOLUTIONAL NEURAL NETWORK

Shiqi Hu[1], Md Murad Hossain[1], Xiaoyue Li[1], Elisa E. Konofagou[1]

[1]Department of Biomedical Engineering, Columbia University, New York, NY, USA

ABSTRACT

Medical imaging-derived tumor or lesion size quantification provides clinically relevant information for diagnosis and treatment monitoring. Multi-frequency single transducer harmonic motion imaging (ST-HMI) is an acoustic radiation force (ARF) based ultrasound elastography method that interrogates tissue mechanical properties by transmitting a multi-frequency ARF excitation pulse and estimating the induced oscillatory displacements at 100-1000 Hz frequency simultaneously. In this study, an automated lesion size assessment method is presented using the U2Net-based convolutional neural network (CNN) model. The training-validation set contains 1094 multi-frequency ST-HMI images of heterogeneous phantoms with the inclusion's Young's moduli ranging from 6-70 kPa and diameters ranging from 1.7-10.5 mm. For each of the 1094 images, ST-HMI-derived normalized peak-to-peak displacement (P2PD) images at 100-1000 Hz in the step of 100 Hz were used as the input to the U2Net. Dice score, sensitivity, specificity, and area estimation error were used to evaluate the performance of the model in the test phantom data set, in vivo 4T1 breast cancer mouse tumors, and ex vivo focused ultrasound (FUS)-induced thermal lesions. The average ± standard deviation of (Dice score, sensitivity, and specificity) was (0.92±0.04, 0.92±0.05, 0.99±0.01), and (0.87±0.07, 0.96±0.04, 0.95±0.04) in the phantoms and mouse tumors, respectively. The average absolute error in the model-predicted area of FUS-induced lesion was 9.45±0.77%. These results demonstrated the potential of multi-frequency ST-HMI to determine tumor size change in chemotherapy treatment monitoring and lesion size in ablation therapy guidance using CNN.

Index Terms— Tumor size, ablation margin, single-transducer harmonic motion imaging, neural network

1. INTRODUCTION

The size of medical imaging-derived lesions, including tumors or thermally ablated areas, is clinically relevant both in diagnosis and therapy guidance. Tumor size monitoring could help physicians to assess the anticancer therapy response [1] and determine the appropriate dosage for treatment planning [2]. In addition, intraoperative imaging is used to assist surgeons to delineate tumor boundaries for surgical resection, in applications including liver [3], pancreatic tumor [4], or thermal ablation of liver cancer [5]. An accurate margin covering the whole tumor could prevent local recurrence and improve survival rates. While computed tomography [6] and magnetic resonance imaging [7] are generally used to assess tumor size and location pre-and post-operatively to evaluate treatment efficacy, ultrasound-based methods can not only be used intra-operatively to guide surgery or ablation treatment but can be also used in longitudinal monitoring of treatment response due to its low cost, portability, and non-ionizing characteristics.

While B-mode ultrasound is typically used for tumor size assessment or thermal treatment guidance, ultrasound elastography can reduce the risk of underestimation of tumors or thermal lesions [8]. One subset of ultrasound elastography-based methods is ARF-induced 'on-axis' displacement or 'off-axis' shear wave imaging [9]. In contrast to the 'off-axis' shear wave imaging, on-axis methods, including ARF impulse (ARFI) imaging [10] and harmonic motion imaging (HMI) [11], provide relative stiffness but have higher spatial resolution and are less subjected to tissue heterogeneity. HMI simultaneously uses a focused and an imaging transducer to generate amplitude-modulated ARF and to track on-axis oscillatory motion. The advantage of HMI is that known input oscillations can be easily filtered out from the other motion artifacts. To reduce the complexity of the current HMI data system consisting of two separate transducers and a mechanical positioner, ST-HMI has been proposed [12]. ST-HMI uses a single imaging transducer for both generating and tracking oscillatory motion by interleaving tracking pulses with excitation pulses and generating a 2-D image by electronically steering the excitation and tracking beam. However, the performance of ST-HMI in delineating the inclusion's size depends on the oscillation frequency [13].

Instead of separately collecting motion data at multiple frequencies to improve performance, ST-HMI with a multi-frequency excitation pulse has been developed to induce and estimate oscillatory motion at 100-1000 Hz in a single acquisition [9]. It has been shown that the oscillation frequency can be exploited to optimize the inclusion image quality without prior knowledge of the inclusion's stiffness or size [9]. For example, 1000 and 700 Hz provided the highest CNR for 2.5 and 10.4 mm diameter inclusions with Young's modulus of 36 kPa [9]. However, assessing the size of the inclusion by sorting out the best frequency is manually

cumbersome and may have subjective variability. In addition, the dynamic color range in the image can also impact the perceived lesion size. Therefore, there is a need for an accurate and less manual input-based method to assess lesion size.

Convolutional neural networks (CNN) have been widely used to evaluate tumor size on B-mode ultrasound images. U-shaped CNN-based models (e.g. UNet) have shown promising results in breast, prostate, and liver tumor detection using B-mode, MRI, and CT images [13]–[15]. However, to the authors' knowledge, the CNN network was never used to assess tumor size and thermal lesions using ultrasound elastography images.

The objective of this study is to develop an accurate and automated tumor and FUS-induced lesion size assessment tool using U2Net [16] by exploiting ST-HMI-derived P2PD images at 100-1000 Hz. U2Net was chosen due to its multi-scale feature extraction capability and strength in dealing with images of multiple resolutions and segmentation conditions [16]. The performance of U2Net was the best in segmenting median nerve on the B-mode images compared to the UNet and ResUNET [17]. The output probability map from the network is considered the predicted boundary of the input. The U2Net model was trained on a phantom data set [9]. Data augmentation (adapted from Pytorch RandomCrop) was performed to increase the dataset size to 1096. The robustness of the model was tested on the phantom inclusion with stiffness and size different from the training data, with the presence of different aberrating layers, FUS-induced lesions in pork muscle, and in vivo 4T1 mouse dataset [9]. The main contribution of this study is two-fold. First, we adapted U2Net to accommodate ST-HMI-derived multi-frequency displacement images at 100 to 1000 Hz in steps of 100 to assess lesion size automatically. Second, the robustness of the model was tested in the in vivo and ex vivo tissues with lesion shapes different from the training shapes. Note, the model was trained in the phantom dataset with circular inclusion.

2. METHOD

2.1. Dataset description and preprocessing

A total of 548 multi-frequency P2PD images at 100—1000Hz were split into training (431) and validation (117) datasets. They were acquired using an L7-4 transducer (Philips Healthcare, Andover, MA, USA) operated by a Verasonics research system (Vantage 256, Verasonics Inc., Kirkland, WA, USA) with and without aberration noise (i.e., wire mesh was placed between the transducer and phantom surface). Multiple acquisitions were collected by moving the transducer in elevational directions. A detailed description of data collection can be found in [9]. In addition to the original images, images were cropped in the background to increase the dataset size. Cropping was performed in such a way that the background area and background-to-inclusion area ratio was different between different cropped images.

We tested our model in four different datasets: (1) cylindrical inclusions with identical stiffness and size but at different elevational locations compared to the training-validation set. (2) spherical inclusions with Young's moduli of 40 and 160 kPa embedded in a 10 kPa background with and without aberration noise which represents different stiffness and boundary configurations compared to the training-validation set (3) in vivo mouse tumor at 6 and 19 days post-injection of 4T1 tumor cells (4) focused ultrasound (FUS)-induced thermally ablated lesions in the ex vivo pork muscles [9]. Mouse tumors were imaged using an L11-5 linear array transducer. Therefore, this dataset tested the model robustness on in vivo data and different transducers. The FUS lesions were ablated for 40 seconds and 3 minutes at a peak positive pressure of 8.5 MPa.

Histogram equalization (HE) proved to be an effective image enhancement technique for kidney ultrasound image segmentation if object edge detection is the focus [18]. An adaptive HE (CLAHE) was used for the phantom dataset (patch size = (4, 4), limit = 2). HE (equalizeHist) was performed on in vivo and ex vivo datasets because the large range of variation in the size and intensity of the noise pattern prevented setting a reasonable limit and tile size.

2.2. Architecture

The CNN for this study is adapted from the structure of U2Net [16]. Within each layer of the network, the conventional Conv blocks were replaced by Residual U Blocks (RSU). A side output probability map was generated to calculate the loss at this layer, and then the loss of all layers was combined with the loss of the output layer to form the loss function. The input was normalized P2PD images at 10 frequencies with a dimension of (10, 144, 144) and the output was (1, 144, 144), which represented the probability of the pixels classified as inclusion.

Based on the characteristics of the multi-frequency ST-HMI data, two modifications were made to the original structure. First, instead of using binary cross entropy as the loss function, Dice coefficient loss [19] was used in this study because DICE exhibits robustness in dealing with imbalanced classes [19]. Second, instead of the five-layer structure in the original U^2Net, a lightweight four-layer structure is used due to the amount of training data and the complexity of the task.

2.3. Implementation details and evaluation metrics

The network was trained for 80 epochs, using a learning rate scheduler with default hyper-parameters (lr=0.001, betas=(0.9, 0.999), eps=1e-8, weight decay=0). The learning curve reached a plateau at the end of training, which verified that the learning was completed after 80 epochs of training. The network was performed on GPU NVIDIA Quadro P400 with a total time of 6 hours. The best epoch (56) was selected based on the best Dice score (0.83) on the validation dataset. To evaluate the accuracy of the model in predicting

boundaries compared to manually drawn boundaries, the Dice similarity score, sensitivity, and specificity were calculated for phantom and in vivo tissue data, and area estimation errors were calculated for FUS-induced lesions.

3. RESULT AND DISCUSSION

3.1. Phantom with cylindrical and spherical inclusions

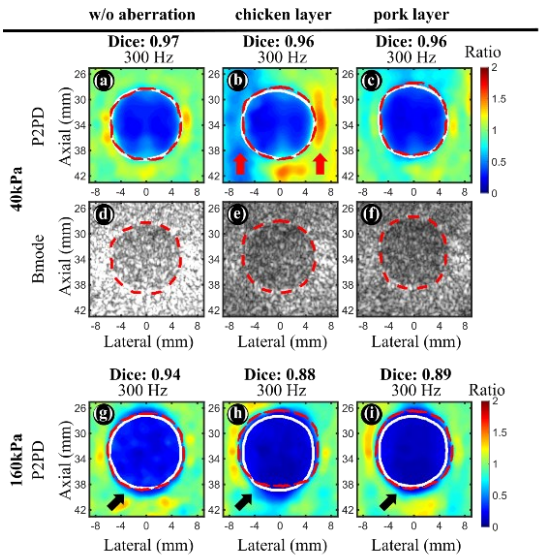

Figure 1. Predicted (white) and true boundary (red) of spherical inclusions. 1st and 2nd row: 40 kPa without (a, d) and with chicken (b, e) and pork (c, f) layer. The red arrows mark the difference in acoustic force between the left and right sides of the inclusion. 3rd row: 160 kPa with wire mesh (g), with chicken (h), and pork (i) layer. The black arrows (g, h, i) mark the boundary effects.

First, for the phantom with cylindrical inclusions, the network can predict boundaries regardless of the stiffness of inclusions with a Dice score (mean ± standard deviation) of 0.91 ±0.03, 0.82± 0.06, 0.91± 0.06, 0.95± 0.03, 0.95± 0.03 for inclusions of 1.7 mm, 2.5 mm, 4.1 mm, 6.5 mm, and 10.5 mm, respectively. The prediction accuracy is higher for larger inclusions. But the performance of 2.5 mm is relatively low compared to other sizes. This may be due to the low contrast of the 2.5 mm B-mode images, which impacted manual boundary delineation. However, more investigations are needed to find the source of the lower accuracy for 2.5 mm. **Figure 1.** shows the P2PD images at 300 Hz and B-mode images for spherical inclusions. The Dice score over 4 positions are 0.97±0.01, 0.97±0.01, 0.95±0.01, 0.94±0.001, 0.90±0.01, 0.89±0.004, for 40kPa, 40kPa with chicken layer, 160 kPa with wire mesh, 160 kPa with chicken layer, and 160 kPa with pork layer. When the imaging field of view was subjected to laterally different acoustic forces due to the inhomogeneous attenuation (panel (b) denoted by red arrows on the left and right side of the inclusion), the network was still able to delineate the boundary. The prediction accuracy was worse for stiffer inclusion because higher stiffness differences between background and inclusion induced a higher boundary effect (panels (g-i), denoted by black arrows). In addition, 160 kPa has a 4 times higher stiffness difference from that of the largest difference in the training data set.

3.2. In vivo 4T1 breast cancer mouse model

The (Dice score, sensitivity, and specificity) for mouse 1 on Day 6 and Day 19 are (0.93, 0.92, 0.99) and (0.88, 0.99, 0.88), respectively. The (Dice score, sensitivity, and specificity) for mouse 2 on Day 6 and Day 19 are (0.76, 0.99, 0.96) and (0. 92, 0.94, 0.9), respectively. The P2PD and B-mode images with predicted and true boundaries are shown in **Figure 2.** The tumor boundary was not always clear on B-mode, but an approximation can be made based on the slight echogenicity difference to provide the comparative benchmark for the size assessment.

3.3. Ex vivo FUS-induced lesion in pork muscles3.4.

Figure 3. shows the P2PD, B-mode, and gross pathology for the two lesions. The P2PD images are presented at 200, 600, and 1000 Hz to show the variation in perceived lesion shape across frequencies. The predicted boundaries are consistent with the gross pathology images, with an absolute area error of approximately 10% (3 min) and 8.9% (40 s). The network predicted ellipse boundaries, despite the perceived shape of lesions in P2PD images being roughly rectangular, especially for 3 min. This may be because the network was trained purely on circular inclusions.

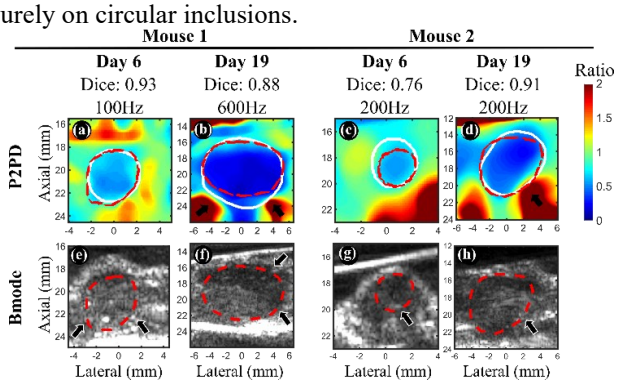

Figure 2. Predicted (white) and true boundaries (red) of mouse 1 and mouse 2 on Day 6 and Day 19. (a, b): P2PD images at 100Hz and 600Hz. (c, d): B-mode images with manually drawn boundaries. The black arrows in (e-h) denote the echogenicity differences used to infer the tumor boundaries.

3.4. Limitations and Future Studies

There are mainly three limitations in this study. First, Bmode-derived boundaries were used as comparative benchmarks. Future studies will evaluate the model in digital phantoms with known boundaries. Second, only circular-shaped inclusions were used in the training validation sets which made it difficult for the model to predict the tumor and FUS-

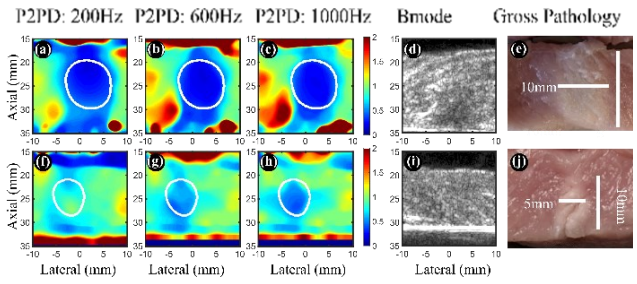

Figure 3. Predicted (white) overlayed with P2PD images of pork muscles after HIFU experiments of 3 min (1st row) and 40 s (2nd row), B-mode images and photographs.

lesion shape correctly (**Figures 2** and **3**). In future studies, we will train the network on the inclusions of different shapes and incorporate average distance loss and B-mode to improve its performance in predicting irregular shapes. Note that, B-mode is collected as a part of the ST-HMI sequence. Third, while the advantage of U2Net over UNet and ResUNET was shown in the segmentation of Bmode ultrasound median nerve images [17], no comparison was performed in this study for ST-HMI imaging. Future studies will compare different CNN structures using ST-HMI imaging.

4. CONCLUSION

An automated tumor size assessment tool using CNN on ST-HMI-derived P2PD images is presented in this study. The network was trained on normalized P2PD images of phantoms with different size and stiffness inclusions and can predict lesion boundaries in tissue-mimicking phantoms, in vivo and ex vivo tissues with different sizes, stiffness, and aberrating conditions with an average accuracy of 0.90. Our future work is to apply it to more irregularly shaped lesions and breast tumors in patients.

5. Acknowledgments

This work was supported by the NIH under Grant R01 CA228275. The authors thank Drs. Saurabh Singh, Indranil Basu, and Chandan Guha from the Albert Einstein College of Medicine and Montefiore Medical Center, Bronx, NY, USA, for providing the cancer cell for the mouse study.

6. Compliance with Ethical Standards

Cancer induction and imaging protocols for mice were reviewed and approved by the Columbia University Irving Medical Institutional Animal Care and Use Committee.

REFERENCES

[1] E. A. Eisenhauer *et al.*, "New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1)," *Eur J Cancer*, vol. 45, no. 2, pp. 228–247, 2009.

[2] S. Mallidi, K. Watanabe, D. Timerman, D. Schoenfeld, and T. Hasan, "Prediction of tumor recurrence and therapy monitoring using ultrasound-guided photoacoustic imaging," *Theranostics*, vol. 5, no. 3, p. 289, 2015.

[3] A. Teatini *et al.*, "The effect of intraoperative imaging on surgical navigation for laparoscopic liver resection surgery," *Sci Rep*, vol. 9, no. 1, pp. 1–11, 2019.

[4] S. Kaushal *et al.*, "Fluorophore-conjugated anti-CEA antibody for the intraoperative imaging of pancreatic and colorectal cancer," *Journal of gastrointestinal surgery*, vol. 12, no. 11, pp. 1938–1950, 2008.

[5] M. G. van Vledder *et al.*, "Intra-operative ultrasound elasticity imaging for monitoring of hepatic tumour thermal ablation," *Hpb*, vol. 12, no. 10, pp. 717–723, 2010.

[6] B. J. Park *et al.*, "CT-guided radiofrequency ablation for hepatocellular carcinomas that were undetectable at US: therapeutic effectiveness and safety," *Journal of Vascular and Interventional Radiology*, vol. 20, no. 4, pp. 490–499, 2009.

[7] B. D. de Senneville, C. Moonen, and M. Ries, "MRI-guided HIFU methods for the ablation of liver and renal cancers," *Therapeutic Ultrasound*, pp. 43–63, 2016.

[8] A. Farrokh, N. Maass, L. Treu, T. Heilmann, and F. K. W. Schäfer, "Accuracy of tumor size measurement: comparison of B-mode ultrasound, strain elastography, and 2D and 3D shear wave elastography with histopathological lesion size," *Acta radiol*, vol. 60, no. 4, pp. 451–458, Jul. 2018, doi: 10.1177/0284185118787354.

[9] M. Hossain and E. E. Konofagou, "Imaging of Single Transducer-Harmonic Motion Imaging-derived Displacements at Several Oscillation Frequencies Simultaneously," *IEEE Trans Med Imaging*, vol. xx, p. 2022.

[10] P.-C. Lee *et al.*, "Liver stiffness measured by acoustic radiation force impulse elastography predicted prognoses of hepatocellular carcinoma after radiofrequency ablation," *Sci Rep*, vol. 10, no. 1, pp. 1–13, 2020.

[11] N. Saharkhiz *et al.*, "Harmonic motion imaging of human breast masses: an in vivo clinical feasibility," *Sci Rep*, vol. 10, no. 1, pp. 1–13, 2020.

[12] M. M. Hossain, N. Saharkhiz, and E. E. Konofagou, "Feasibility of Harmonic Motion Imaging Using a Single Transducer: In Vivo Imaging of Breast Cancer in a Mouse Model and Human Subjects," *IEEE Trans Med Imaging*, vol. 40, no. 5, pp. 1390–1404, May 2021, doi: 10.1109/TMI.2021.3055779.

[13] X. Zhuang *et al.*, "Residual Swin Transformer Unet with Consistency Regularization for Automatic Breast Ultrasound Tumor Segmentation," in *2022 IEEE International Conference on Image Processing (ICIP)*, 2022, pp. 3071–3075.

[14] S. Li, Y. Chen, S. Yang, and W. Luo, "Cascade dense-unet for prostate segmentation in MR images," in *International Conference on Intelligent Computing*, 2019, pp. 481–490.

[15] Q. Jin, Z. Meng, C. Sun, H. Cui, and R. Su, "RA-UNet: A hybrid deep attention-aware network to extract liver and tumor in CT scans," *Front Bioeng Biotechnol*, p. 1471, 2020.

[16] X. Qin, Z. Zhang, C. Huang, M. Dehghan, O. R. Zaiane, and M. Jagersand, "U2-Net: Going deeper with nested U-structure for salient object detection," *Pattern Recognit*, vol. 106, Oct. 2020, doi: 10.1016/j.patcog.2020.107404.

[17] J. Shao, K. Zhou, Y. H. Cai, and D. Y. Geng, "Application of an Improved U2-Net Model in Ultrasound Median Neural Image Segmentation," *Ultrasound Med Biol*, vol. 48, no. 12, pp. 2512–2520, Dec. 2022, doi: 10.1016/j.ultrasmedbio.2022.08.003.

[18] W. M. Hafizah and E. Supriyanto, "Comparative Evaluation of Ultrasound Kidney Image Enhancement Techniques," 2011.

[19] F. Milletari, N. Navab, and S. A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proceedings - 2016 4th International Conference on 3D Vision, 3DV 2016*, Dec. 2016, pp. 565–571. doi: 10.1109/3DV.2016.79.