# A Data-Efficient Deep Learning Strategy for Tissue Characterization via Quantitative Ultrasound: Zone Training

Ufuk Soylu🄳, *Graduate Student Member, IEEE*, and Michael L. Oelze🄳, *Senior Member, IEEE*

Abstract—**Deep learning (DL) powered biomedical ultrasound imaging is an emerging research field where researchers adapt the image analysis capabilities of DL algorithms to biomedical ultrasound imaging settings. A major roadblock to wider adoption of DL powered biomedical ultrasound imaging is that acquisition of large and diverse datasets is expensive in clinical settings, which is a requirement for successful DL implementation. Hence, there is a constant need for developing data-efficient DL techniques to turn DL powered biomedical ultrasound imaging into reality. In this work, we develop a data-efficient DL training strategy for classifying tissues based on the ultrasonic backscattered RF data, i.e., quantitative ultrasound (QUS), which we named zone training. In zone training, we propose to divide the complete field of view of an ultrasound image into multiple zones associated with different regions of a diffraction pattern and then, train separate DL networks for each zone. The main advantage of zone training is that it requires less training data to achieve high accuracy. In this work, three different tissue-mimicking phantoms were classified by a DL network. The results demonstrated that zone training can require a factor of 2–3 less training data in low data regime to achieve similar classification accuracies compared to a conventional training strategy.**

Index Terms— **Biomedical ultrasound imaging, deep learning (DL), tissue classification.**
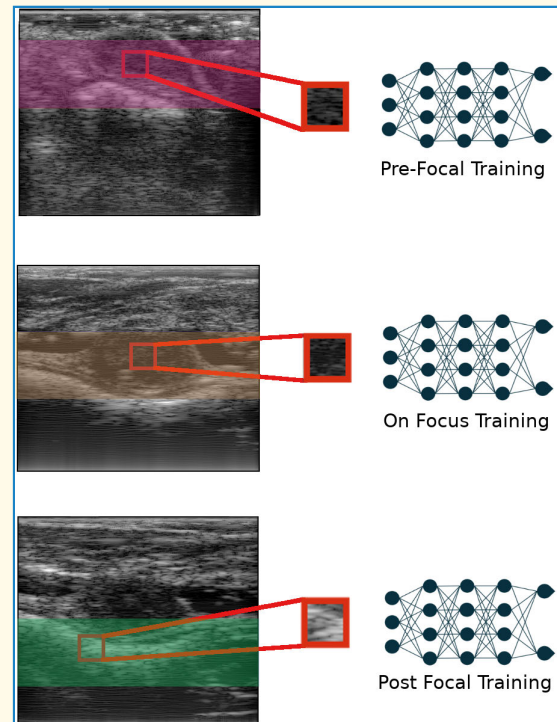
I. INTRODUCTION

DEEP learning (DL) powered biomedical ultrasound imaging is becoming more advanced and coming closer to routine clinical applications in recent years [1]. DL is the process of learning a hierarchy of parameterized nonlinear transformations to perform a desired function. Therefore, DL algorithms extract a hierarchy of features from raw input images and image data automatically rather than extracting features manually. Due to rapid increase in computational power and large datasets, DL and machine learning algorithms have emerged as leading tools and have achieved impressive results in various research fields. Among DL algorithms, convolutional neural networks (CNNs) use convolutional layers to embed structural priors of translational invariance, which make them parameter and data-efficient learners for image analysis tasks. Respectively, CNNs are the most popular and successful DL structure for ultrasound biomedical imaging [2]. Common DL applications that have provided notable results in the context of biomedical ultrasound imaging are classification [2], [3], [4], [5], [6], [7], detection [2], [8], segmentation [2], [9], [10], [11], [12], image reconstruction [13], [14], [15], [16], [17], [18], [19], [20], [21], and ultrasound elastography [22], [23], [24], [25], [26]. Furthermore, DL algorithms have been employed in advanced ultrasound imaging applications such as super-resolution imaging of microvasculature structure via ultrasound localization microscopy [27], [28].

Even though DL is promising for biomedical ultrasound imaging, there are certain roadblocks to wider adoption.

Highlights

- **The most significant element of novelty: The proposed method trains separate deep learning networks for different zones of the ultrasound images by dividing the complete field of view into regions with varying diffraction patterns.**

- **An executive summary on the main results: The results showed that the proposed method requires 2–3 times less training data in low data regimes while achieving similar accuracy to conventional training strategies.**

- **Highlight the implications and importance of the reported findings to the research field: The paper developed a data-efficient deep learning training strategy, a step towards the development of deep learning-based ultrasound imaging for wider clinical adoption.**

A major roadblock is that acquiring large and diverse datasets is expensive. Hence, it is important to develop data-efficient DL algorithms to overcome this limitation. Another roadblock is that there are large variations in ultrasound images due to operator, patient or machine dependent factors. Therefore, improving the robustness of DL algorithms against variations in ultrasound images is necessary. Overall, to turn DL powered biomedical ultrasound imaging into reality, there is a constant need for developing DL algorithms, which are data efficient and more robust against variations in ultrasound images.

In this paper, we examine DL techniques for classifying samples based on ultrasonic backscattered RF data similar to the work of Nguyen et al. [5]. Classifying tissues has recently evolved from model-based approaches such as quantitative ultrasound (QUS) techniques to model-free, DL-based techniques. Nguyen et al. [29] demonstrated that QUS techniques are able to detect the presence of steatosis in a rabbit model of fatty liver with a classification accuracy of 84.11%. In a later study, Nguyen et al. [5] compared a DL-based classifier to a QUS-based classifier for the problem of fatty liver classifier and found that the DL-based classifier outperformed the QUS-based approach with the accuracy of 74% versus 59%. While the traditional spectral-based QUS approach does not utilize the phase information in the RF signal, DL-based approaches can extract additional classification power from the lost phase information from the RF data. Furthermore, the DL-based approach does not require a model like the QUS approach, which means that features of the backscattered signal that are missed by the QUS approach can be picked up by the DL approach. Subsequently, the DL approach performs feature extraction and classification simultaneously.

To improve classification, we consider the diffraction patterns associated with ultrasonic transducers and how they result in different regions or "zones" that must also be learned to separate the system signal from the sample signal. We propose a training strategy, which we call zone training. In zone training, we propose to divide the complete field of view of an ultrasound image into multiple zones such as pre-focal, on-focus, and post-focal zones. Then, we train separate neural networks for each zone using the data belonging to the corresponding zone. In a sense, we train expert neural networks for each zone as opposed to regular training, which uses all data coming from the complete field of view to train a single neural network. The main intuition is that at each

zone, there are different diffraction patterns and learning all the patterns by a single network is harder than learning a single diffraction pattern by a single expert network. The main advantage of zone training is that it requires less data to achieve similar classification performance in comparison to regular training in low data regime.

II. BACKGROUND AND MOTIVATION

Zone training is similar to applying an attention mask to the input manually and training separate networks for each mask to learn dedicated convolutional filters per zone. In this sense, zone training applies attention in a simple and direct way to incorporate the physics of diffraction into DL training. There are methods in the literature that enable learning of varying convolution kernels over the complete field of view, e.g., pixel-adaptive convolution [30]. In general, attention mechanisms, e.g. self-attention mechanisms, Guo et al. [31] were invented initially for computer vision tasks, where the data are abundant and they apply attention by altering network architecture and hence model complexity to improve classification accuracy. However, in the context of biomedical imaging, we are in a different regime where the data are often scarce. Therefore, we favor utilizing a smaller training set to achieve a desired classification accuracy. Overall, zone training provides us a method to reduce training set size by modifying data distribution without altering model complexity. Furthermore, zone training can be perceived as utilizing a symbolistic approach, in the form of a simple if-else structure (if data are from a certain zone, train a specific network), to transfer physics knowledge into DL training. Combining DL and symbolic reasoning is known in the literature as neural-symbolic computing (NSC), which can lead to data-efficient AI [32].

In our experiments, we chose tissue classification as the primary application and tested our proposed method to classify three distinct tissue-mimicking phantoms. To further motivate zone training, we describe a clinical scenario when it is the most relevant and advantageous. For instance, ultrasound imaging can be used to examine and characterize tumors, whether benign or malignant, which can exist at different depths within a body. When using QUS approaches for tumor characterization, a region of interest (ROI) inside the tumor is selected to examine the signals from the tumor. We show two tumor image examples where the tumors are at different depths in Fig. 1, but the same probe is used. Different depths correspond to different zones and red rectangles are
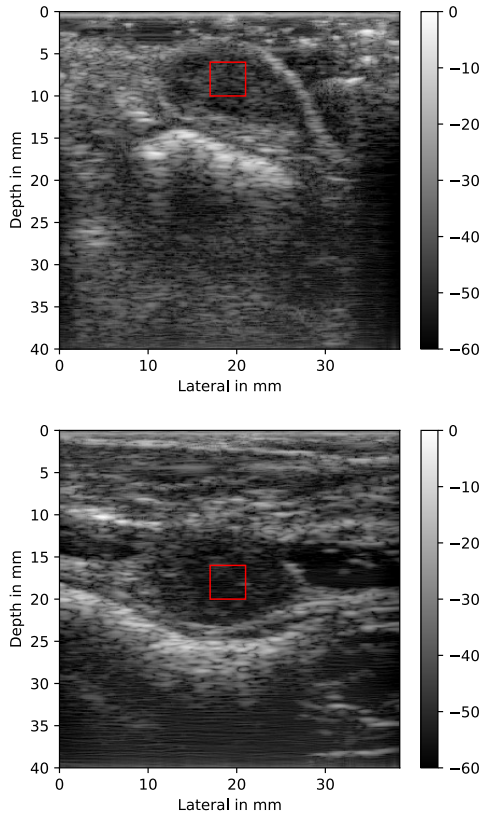
Fig. 1. Tumor examples, in dB scale, whose scan depths vary in the field of a transducer probe. The tumor images were acquired from rabbits having mammary VX2 tumors. All animal experiments were approved by the Institutional Animal Care and Use Committee at the University of Illinois at Urbana–Champaign.

sampled from the tumors in those ultrasound images. In zone training, we have separately trained DL algorithms for each zone. Experimental studies for zone training in this work were conducted under two assumptions following the clinical scenario. First, we assume that we are not trying to detect the ROI. In other words, we are given rectangular patches of data to classify a tissue state. The ROI can be detected by another algorithm or by the operator. The operator in the clinic can adjust imaging settings to obtain the best imaging quality, and then, select the ROI which should be considered as Human-centered AI, whose aim is to amplify and augment rather than displace human abilities [33]. The second assumption is that the ROI is larger than the rectangular patches of data so that the classification networks in this work, take uniform rectangular patches as their input.

III. METHODS

A. Phantoms

Three different tissue-mimicking phantoms were used in the experiments, which we designated as Phantom1, Phantom2, and Phantom3. They are cylindrically shaped as shown in Fig. 2 and their properties are summarized in Table I.

Phantom1, which mimics human liver, has been described by Wear et al. [34]. Phantom1 had a measured attenuation coefficient slope of approximately 0.4 dB $\times$ cm^{-1} $\times$ MHz^{-1}. Its materials were produced based on the method of
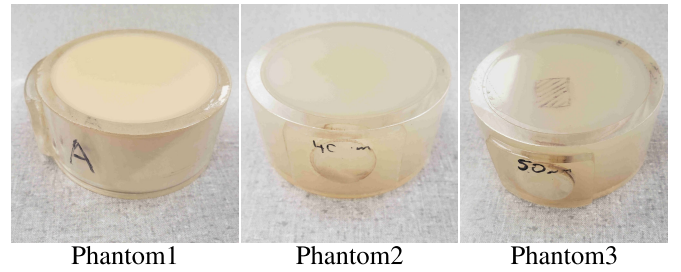

Fig. 2. Tissue-mimicking Phantoms.

TABLE I
PHANTOM PROPERTIES

	Phantom1	Phantom2	Phantom3
Sphere diameter (μm)	75-90	41 $\pm$ 2	50 $\pm$ 2.4
Background material	3.5% agar	2% agar	2% agar
Sound speed (m/s)	1540	1539	1539
Attenuation (dB/cm/MHz)	0.4	0.1	0.1

Madsen et al. [35] and they are macroscopically uniform. The only nonuniformity in Phantom1 results from the random positioning of microscopic glass bead scatterers. The component materials and their relative amounts by weight for Phantom1 are agarose (3.5%), n-propanol (3.4%), 75–90-μm-diameter glass beads (0.38%), bovine milk concentrated three times by reverse osmosis (24.5%), liquid Germall Plus preservative (International Specialty Products, Wayne, NJ) (1.88%), and 18-$M\Omega$-cm deionized water (66.3%)."

Phantom2 and Phantom3 are both low attenuation phantoms, whose properties have been described by Anderson et al. [36] and constructions have been described Madsen et al. [37]. Both phantoms were made with the same weakly scattering agar background material but contained different sizes of scatterers. They have an attenuation coefficient slope of approximately equal to 0.1 dB $\times$ cm^{-1} $\times$ MHz^{-1}. Glass-sphere scatterers (Potters Industries, Inc., Valley Forge, PA; Thermo Fisher Scientific (formerly Duke Scientific), Inc., Waltham, MA) were used in both phantoms with weakly scattering 2% agar background. The only difference in the phantoms was the size distribution of the glass bead scatterers, i.e., Phantom2 had a mean diameter of 41 μm and Phantom3 had a mean diameter of 50 μm.

B. Ultrasound Scanning Procedures

Ultrasound gel was placed on the surfaces of the phantoms and then the phantoms were scanned with an L9-4/38 transducer using a SonixOne system (Analogical Corporation, Boston, MA, USA) providing an analysis bandwidth of 2-7.5 MHz. 1007 frames of post-beamformed RF data sampled at 40 MHz were acquired from each phantom and saved for offline processing.

The imaging array had a center frequency of approximately 5.5 MHz and was operated with a single axial focus at 2 cm depth and a fixed elevational focus of 1.9 cm. The center frequency of the pulse was chosen as 9 MHz to provide higher bandwidth (resolution) for the transducer. The total
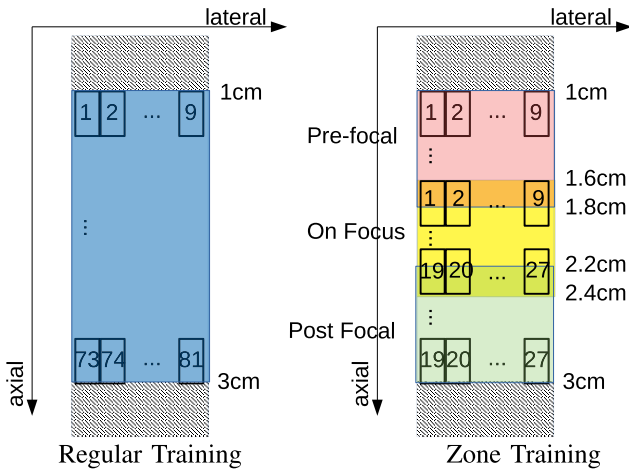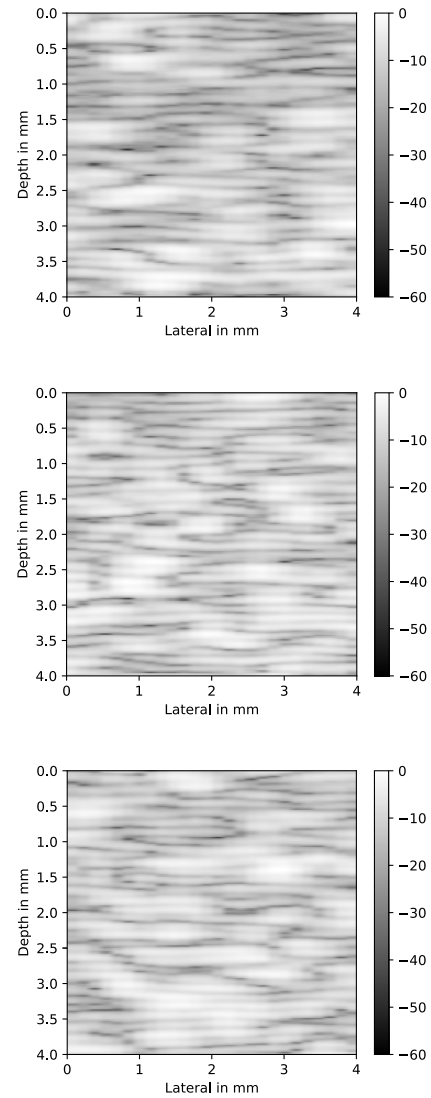
Fig. 3. Patch extraction.

Fig. 4. Example B-mode images of extracted patches in dB scale from the pre-focal zone centered at 1.4 cm. Classifying these patches is difficult by visual inspection. Top row: Phantom1; Middle row: Phantom2 and Bottom row: Phantom3.

imaging depth was chosen as 4 cm, which is equal to the height of the phantoms. Output power was chosen as -5 dB, which corresponds to -5 dB lower power level with respect to maximum output power of the system.

C. Dataset

We acquired 1007 ultrasound images per phantom by free-hand motion. In total, we acquired 3021 ultrasound frames. The size of an ultrasound image frame was 2080 pixels $\times$ 256 pixels. There were 2080 samples along the axial direction that corresponded to 4 cm depth. Even though the L9-4/38 transducer has 128 channels, the SonixOne system interpolates to 256 channels that correspond to 256 lateral pixels. The dataset of ultrasound images is also publicly available at https://osf.io/7ztg3/ (DOI 10.17605/OSF.IO/7ZTG3). After acquiring the ultrasound images per phantom, we extracted rectangular image patches to be used in training and testing.

In patch extraction, which is depicted in Fig. 3, we extracted rectangular image patches whose sizes were 200 pixels $\times$ 26 pixels that correspond to square image patches whose size were 4 mm $\times$ 4 mm in physical dimensions. From one ultrasound image, we could extract 81 (9 lateral $\times$ 9 axial) image patches when we used the complete field of view as in regular training. While extracting image patches, we did not use the first 540 pixels in the ultrasound image. Axially, we obtained the next line of individual patches by translating the start of the next patch by 100 pixels along the axial depth. Laterally, we obtained the next line of individual patches by translating the start of the next patch by 26 pixels along the axial depth. Overall, in patch extraction for regular training, there were nine axial lines and nine lateral lines to extract individual patches that lead to extracting 81 image patches per ultrasound image.

For zone training, we first developed definitions for the zones based on the diffraction pattern for a single focused transducer. In this work, we broke the complete field of view into three zones axially: a pre-focal zone which is centered at 1.4 cm, an on-focus zone which is centered at 2 cm, and a post-focal zone which is centered at 2.6 cm. Then, each zone coincides with three axial lines of the complete field of view

in patch extraction. Therefore, three zones together use the same data as in regular training. For each zone, we extracted 27 (9 lateral $\times$ 3 axial) image patches whose sizes were 200 pixels $\times$ 26 pixels corresponding to 4 mm $\times$ 4 mm in physical dimensions. Example B-mode images of the patches corresponding to each phantom are provided in Fig. 4.

D. Training

DL training was done using two machines each with a single GPU. One machine had TITAN RTX and the other machine had RTX A5000. All implementations were done with the PyTorch library [38]. As a data preprocessing step, we applied z-score normalization at the patch level, i.e., the mean intensity value of patches was subtracted from each patch, and then, each pixel in a patch was divided by the standard deviation of the intensity of the patches. The batch number was chosen as 128 throughout all experiments. Horizontal flip with 0.5 probability was implemented as a data augmentation step

TABLE II
NETWORK ARCHITECTURE

Layer Name	Output Size	Regular & Zone Training
conv1&relu	$48 \times 4 \times 96$	11×11, stride4
conv2&relu	$48 \times 4 \times 256$	5×5, pad2
conv3&relu	$48 \times 4 \times 384$	3×3, pad1
conv4&relu	$48 \times 4 \times 384$	3×3, pad1
conv5&relu	$48 \times 4 \times 256$	3×3, pad1
maxpool1	$23 \times 1 \times 256$	3×3, stride2
fc1&relu	4096	5888×4096 connections
fc2&relu	4096	4096×4096 connections
fc3	3	4096×3 connections

Fig. 5. Depth aware training: two-layered input, one with depth information and the other one with ultrasound RF data.

in the training process. We used the Adam algorithm [39] as the optimizer in all experiments. Additionally, the models were trained using cross entropy loss with uniform class weights, which includes built-in softmax function in PyTorch implementation [38].

In this work, we used CNN architectures consisting of two parts: feature extractors that consist of convolution layers, max-pooling layers, and nonlinear activation functions, and a classifier that consists of fully connected layers and nonlinear activation functions. They also have significantly fewer parameters and so they can be trained more efficiently than fully connected networks [40]. We used a slightly modified CNN architecture, which is derived from AlexNet [41] and is shown in Table II. In the training, dropout layers with 0.5 probabilities were added to improve the regularization and deal with over-fitting, before fully connected1 and fully connected2 layers. Initial weights for the network were chosen based on the original paper [41].

In the experiments, we searched the learning rate and the epoch number using a validation set. More specifically, the learning rate and the epoch number were determined to achieve "asymptotic test accuracy," which ideally is defined as the number of epochs of training required such that any further training provides no improvement in test accuracy. The process of forming training, testing, and validation sets started with randomly selecting the desired number of ultrasound images per phantom. The same number of ultrasound frames were set apart for validation, training and testing sets. Then, we extracted patches, as described in Section III-C, to form the training, testing and validation sets. After adjusting the learning rate and epoch number using the validation set, we trained neural networks in the training sets and obtained classification accuracies in the test sets. We repeated each experiment ten times starting from random ultrasound frame selection for training and testing sets. In Section IV, we report the learning rate, epoch number, mean classification accuracy, and standard deviation for each experiment.

E. Depth-Aware Training

In patch extraction, global coordinates are lost. Therefore, in addition to regular training, zone training is also compared against *depth aware training*, which utilizes global coordinates in the training. In *depth aware taining*, we input the
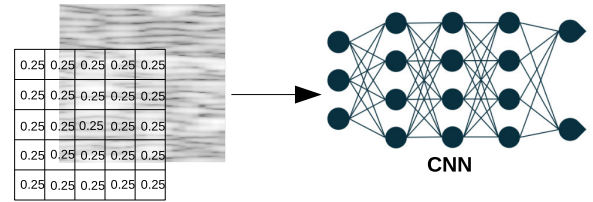
depth as an additional feature. Specifically, the CNN now takes a two-layered input, one layer is the image patch of 200 pixels × 26 pixels and the other one is a constant array of 200 pixels × 26 pixels whose values correspond to the relative depth, as shown simplistically in Fig. 5. The depth information is normalized between 0 and 1 where 0 is the depth of the nearest patch and 1 is the depth of the farthest patch. Overall, *depth aware training* is designed to consider the global location of the input patch during both training and testing so that the DL network adapts itself based on the relative depth being near 0 or 1.

IV. RESULTS

Results are organized into two parts. In the first part, we present results that help us to determine if our zone definitions are favorable by experimenting with axial zone widths, axial zone locations, and by sweeping testing zone centers around training zone centers. Our purpose in this part is to determine a reasonable way to divide the field of view into multiple zones, which is required for zone training. In the second part, we investigate the relationship between training set size and classification accuracy for zone training, regular training, and *depth aware training*.

A. Examination of Zone Definitions

We now present four results that are helpful in determining zone definitions. In the first result, we investigate how much classification accuracy drops as we shift the testing zone away from the training zone. Specifically, we train a neural network using patches from the on-focus zone, and then, we test the neural network with patches from nearby zones. This result shows us how much the diffraction patterns change around the focal zone. In the second result, we repeat the same experiment for the pre-focal zone and the post-focal zone to investigate how much the diffraction patterns change around these zones. In the third result, we experiment with axial zone width in terms of number of overlapping patches per zone. We plot classification accuracy for the on-focus zone when we increase the number of overlapping patches used in patch extraction. In the fourth result, we experiment with axial zone locations and we plot classification accuracy at different zone centers.

In Fig. 6, classification accuracy is plotted as the testing zone center is swept by 0.8 cm toward and away from the transducer around the training zone center. We trained a neural network using patches from the on-focus zone centered at 2 cm depth, and then we tested the neural network with patches from zones centered at 1.2, 1.4, 1.6, 1.8, 2, 2.2, 2.4,
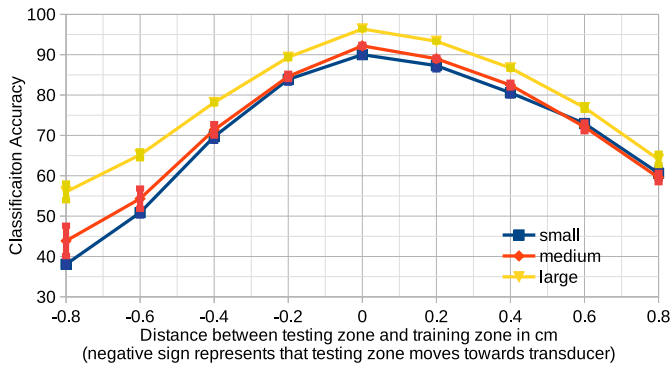
Fig. 6. Sweeping testing zone center for a network trained using the on-focus zone: classification accuracy versus distance between the training and testing zones for different dataset sizes. The colors indicate the size of the training set. The blue color is for 675 patches which is labeled as small, the red color is for 2700 patches which is labeled as medium, and the yellow is for 13 500 patches which is labeled as large.
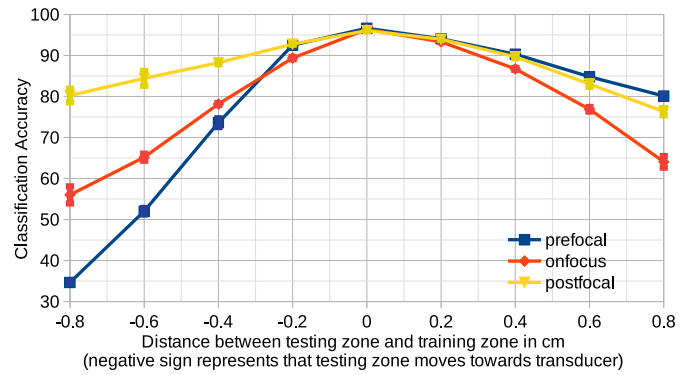

Fig. 7. Sweeping testing zone center for networks trained on different zones: classification accuracy versus distance between training and testing zones. Colors represent the training zone. The blue color is for the pre-focal zone, which is labeled as prefocal. The orange color is for the on-focus zone, which is labeled as onfocus. The yellow color is for the post-focal zone, which is labeled as postfocal.

2.6, and 2.8 cm, respectively. Overall, the y-axis represents classification accuracy and the x-axis represents the relative distance between the testing zone and the training zone. For instance, a value of -0.8 means that the testing zone is 0.8 cm closer to transducer than the training zone and $+0.8$ means that the testing zone is 0.8 cm farther away from transducer than the training zone. We repeat the experiments for different sizes of training sets. We used 675 image patches, 2700 image patches and 13 500 image patches in the training which correspond to 25 ultrasound images, 100 ultrasound images, and 500 ultrasound images, respectively. In the figure, colors indicate the size of the training set. Epoch numbers and learning rates in the training were chosen as 2000 and 5e-6 for 25 ultrasound images, 1500 and 1e-5 for 100 ultrasound images, 400 and 1e-5 for 500 ultrasound images.

Several observations can be made form Fig. 6. For the small and medium sets, when the testing zone moved closer to the transducer by 0.4 cm, classification accuracy dropped to 70%. However, when the testing zone moved away from the transducer by 0.4 cm, classification accuracy remained above 80%. Similarly, for the large set, when the testing zone moved closer to the transducer by 0.4 cm, classification accuracy dropped to below 80%. However, when the testing zone moved away from the transducer by 0.4 cm, classification accuracy remained well above 85%. Similar observations can be made at other spatial locations as well.

In Fig. 7, similar to Fig. 6, we plot classification accuracy as the y-axis and relative distance between testing zone and training zone as the x-axis. In this figure, we experiment with the pre-focal zone and the post-focal zone in addition to the on-focus zone. When we trained AlexNet using patches from the pre-focal zone, we tested the network with patches centered at 0.6 cm, 0.8 cm, 1 cm, 1.2 cm, 1.4 cm, 1.6 cm, 1.8 cm, 2 cm, and 2.2 cm. When we trained a CNN using patches from the post-focal zone, we tested the network with patches centered at 1.8 cm, 2 cm, 2.2 cm, 2.4 cm, 2.6 cm, 2.8 cm, 3 cm, 3.2 cm, and 3.4 cm. In this result, we used a fixed training set size, which was 13 500 image patches or 500 ultrasound images. In the figure, colors represent the
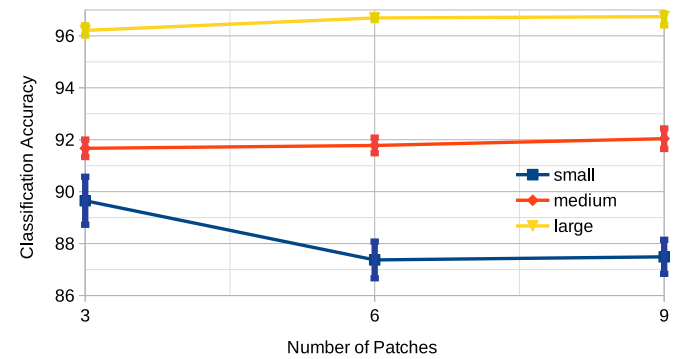

Fig. 8. Classification accuracy versus axial zone width for the on-focus zone. The colors indicate the size of the training set. The blue color is for 25 ultrasound images which is labeled as small, the red color is for 100 ultrasound images which is labeled as medium, and the yellow is for 500 ultrasound images which is labeled as large.

training zone. Epoch numbers and learning rates in the training were chosen as 400 and 1e-5 for all zones.

Several observations can be made from Fig. 7 results. When the testing zone was closer to the transducer by 0.4 cm, classification accuracies were slightly lower than 90%, slightly lower than 80% and 75% for the post-focal zone, the on-focus zone and the pre-focal zone, respectively. However, when the testing zone moved away from the transducer by 0.4 cm, classification accuracies were around 90% for all zones. Second, we observed that the post-focal zone was the most robust zone against the shift in the testing. Classification accuracy for the post-focal zone remained approximately above 80% in all shifts.

In Fig. 8, we plot classification accuracy as the y-axis and axial zone width as the x-axis for the on-focus zone. In zone training, we extract three overlapping patches per ultrasound image as described in Section III-C and shown in Fig. 3. In this result, we make an exception to experiment with zone width, which is defined in terms of number of patches. We now extract 3, 6, and 9 overlapping patches from each ultrasound image for the on-focus zone and these numbers form the x-axis. Specifically, extracting three patches coincides with the original on-focus zone definition, while
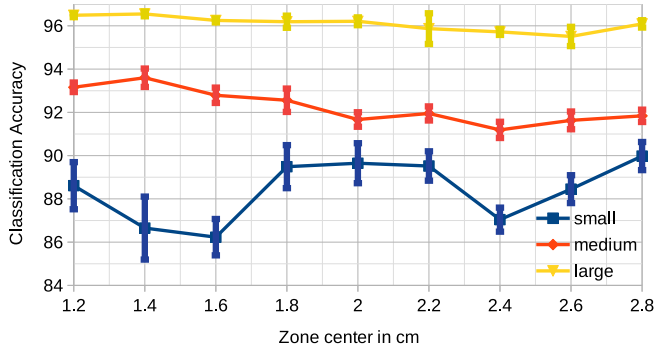
Fig. 9. Classification accuracy versus zone center for different dataset sizes. Colors represent training set sizes where the blue color is for 675 patches which is labeled as small, the red color is for 2700 patches which is labeled as medium, and the yellow is for 13 500 patches which is labeled as large.

extracting six patches coincides with merging pre-focal and on-focus zones; and extracting nine patches coincides with regular training. Additionally, we used three different sizes for the training set. We used 25 ultrasound images, which corresponds to 675, 1350, 2025 training image patches when we extract 3, 6, and 9 patches from each ultrasound image, respectively. Similarly, we used 100 ultrasound images, which corresponds to 2700, 5400, 8100 training image patches and we used 500 ultrasound images, which corresponds to 13 500, 27 000, 40 500 training image patches. As a side note, for this graph, we used the same training and testing zones, unlike the previous two graphs, and colors in the graph represent training set sizes. Moreover, epoch numbers and learning rates were were chosen in accordance with the previous figures. From the Fig. 8, one can observe that for the small dataset size increasing the number of patches, i.e., broadening the zone size, resulted in poorer classification.

In Fig. 9, we plot classification accuracy as the y-axis and zone center as the x-axis. For this graph, we tested and trained networks from the same zone while sweeping the zone center axially. We trained and tested our networks for zones centered at 1.2, 1.4, 1.6, 1.8, 2.0, 2.2, 2.4, 2.6, and 2.8 cm. We repeat the experiments for different sizes of training sets. We used 675 image patches (25 ultrasound images), 2700 image patches (100 ultrasound images), and 13 500 image patches (500 ultrasound images) in the training. Epoch numbers and learning rates were chosen in accordance with the previous figures.

B. Training Set Size Versus Classification Accuracy

This section compares the performance of zone training against that of regular training and depth-aware training under various data conditions from low data size regimes to larger data size regimes to investigate if zone training is more successful when there is a low amount of data. Tables III–VIII are confusion matrices that list the classification accuracies for different training and testing strategies using training set sizes of 10, 25, 50, 100, 200, and 500 ultrasound images. Rows represent training strategies: The first row, denoted as pre-focal, is for training with patches from the pre-focal zone. The second row, denoted as on-focus, is for training with

TABLE III
CLASSIFICATION ACCURACIES WITH TEN ULTRASOUND IMAGES

	Pre	On	Post
Pre-Focal	81.01±3.20	75.18±3.67	38.99±1.91
On Focus	53.00±4.25	86.77±2.81	74.56±3.33
Post Focal	42.78±3.90	74.00±4.82	86.96±1.88
Regular	78.62±2.53	83.52±2.00	85.14±2.13
Depth-Aware	80.74±2.43	83.96±3.17	87.27±2.45

TABLE IV
CLASSIFICATION ACCURACIES WITH 25 ULTRASOUND IMAGES

	Pre	On	Post
Pre-Focal	86.65±2.91	77.70±4.14	42.73±9.93
On Focus	54.19±2.86	89.65±1.85	73.07±3.11
Post Focal	43.11±3.61	73.17±4.40	88.45±1.29
Regular	81.64±2.80	87.49±1.31	87.35±1.73
Depth-Aware	82.79±1.52	87.89±0.91	88.81±1.23

TABLE V
CLASSIFICATION ACCURACIES WITH 50 ULTRASOUND IMAGES

	Pre	On	Post
Pre-Focal	91.66±0.93	83.91±2.41	56.60±8.91
On Focus	53.42±3.12	90.42±0.96	73.98±3.51
Post Focal	44.75±3.10	77.87±2.98	89.63±1.12
Regular	89.66±1.61	89.23±1.85	89.01±1.06
Depth-Aware	90.15±1.90	90.32±1.49	89.91±1.95

TABLE VI
CLASSIFICATION ACCURACIES WITH 100 ULTRASOUND IMAGES

	Pre	On	Post
Pre-Focal	93.60±0.85	84.09±1.51	62.19±8.20
On Focus	57.60±3.94	91.67±0.67	74.40±2.06
Post Focal	48.62±3.13	78.53±4.54	91.63±0.82
Regular	93.36±1.22	92.04±0.79	92.08±0.57
Depth-Aware	93.58±1.17	92.61±0.99	92.77±1.06

patches from the on-focus zone. The third row, denoted as post-focal, is for training with patches from the post-focal zone. The fourth row, denoted as regular, is for training with regular training strategy. The last row, denoted as depth-aware, is for training with Depth-Aware Training strategy. Columns represent testing strategies: testing with patches from the pre-focal zone, testing with patches from the on-focus zone, testing with patches from the post-focal zone and testing with complete field of view, respectively, from first to last column. Epoch numbers and learning rates were chosen in accordance with the previous figures, which were 2500 and 5e-6 for ten ultrasound images, 2000 and 5e-6 for 25 ultrasound images, 2000 and 1e-5 for 50 ultrasound images, 1500 and 1e-5 for 100 ultrasound images, 1000 and 1e-5 for 200 ultrasound images, 400 and 1e-5 for 500 ultrasound images.

The tables verify that zone training had better classification accuracy than regular training and Depth-Aware Training in the low data regime. When we used 10, 25, or 50 ultrasound images in training, zone training performed 1%–5% better than regular training and 1%–4% better than Depth-Aware Training. Additionally, depth-aware training performed approximately 1% better than Regular Training for all training set sizes.

TABLE VII
CLASSIFICATION ACCURACIES WITH 200 ULTRASOUND IMAGES

	Pre	On	Post
Pre-Focal	94.69±0.41	84.43±0.56	63.77±3.93
On Focus	61.45±4.47	94.19±0.35	76.06±2.28
Post Focal	57.95±5.66	83.79±1.98	93.41±0.53
Regular	94.79±0.35	94.50±0.41	94.00±0.35
Depth-Aware	95.17±0.27	95.00±0.29	95.08±0.36

TABLE VIII
CLASSIFICATION ACCURACIES WITH 500 ULTRASOUND IMAGES

	Pre	On	Post
Pre-Focal	96.55±0.22	84.53±0.88	66.53±5.62
On Focus	62.25±4.92	96.21±0.36	75.87±3.05
Post Focal	58.97±8.04	82.34±2.74	95.51±0.89
Regular	97.09±0.77	96.74±0.67	96.75±0.32
Depth-Aware	97.40±0.10	97.32±0.16	97.15±0.16

Lastly, the performance of the zones varied as the size of the training set was reduced. The classification accuracy dropped around 17% when we used ten ultrasound images in training in comparison to 500 ultrasound images for the pre-focal pattern. For the on-focus pattern, the same percentage drop was approximately 12 points and for the post-focal pattern, the same percentage drop was around ten points.

V. DISCUSSION

We proposed a DL training strategy, named zone training, where we split the complete field of view into zones such as the pre-focal, the on-focus and the post-focal zones. Then, we trained separate networks for each zone. We investigated zone training thoroughly by experimenting with zone definitions and their behavior under different training set sizes.

The figures provide several important observations. From Fig. 6, we observed that as the testing zone moved toward the transducer, classification accuracy dropped faster and it was valid for small, medium, and large training set sizes. The observation indicates that the pre-focal diffraction pattern was more complicated and it changed faster than the post-focal diffraction pattern.

In Fig. 7, we quantified how classification accuracy decreased when the testing zone moved away from the training zone for training with the on-focus zone, the pre-focal zone, and the post-focal zone. First, we observed that when the testing zone was closer to the transducer, classification accuracies dropped faster for the pre-focal and on-focus zones. For the post-focal zone, classification accuracies were relatively symmetric around the zone center. This further verified our previous observation stating that the pre-focal pattern was more complicated and changed quickly in comparison to the post-focal pattern. Another observation was that for the pre-focal training, classification accuracy deteriorated slowly when the testing zone moved away from the transducer in comparison to the testing zone moving toward the transducer, which further illustrates the complicated behavior of the pre-focal pattern.

In Fig. 8, we investigated the relationship between classification accuracy and zone width in terms of overlapping patches for the on-focus zone. First, we observed that as we increased the number of patches, classification accuracy remained relatively constant for the medium size training set, while classification accuracy slightly increased for the large size training set. However, classification accuracy dropped as we increased the number of patches for the small size training set. Specifically, classification accuracy dropped to around 87% from 90% as we increased the number of patches from 3 to 6 and it stayed relatively constant when we increased the number of patches to 9. These observations indicate that zone training was more robust when the training set size was smaller. However, regular training can be preferable when the training set size was larger.

In Fig. 9, we determined the best zone location axially in terms of classification accuracy. We observed that the relationship between classification accuracy and the zone location depended on the training set size. For the large training set size, the classification accuracy stayed relatively constant around 96% in all axial locations. For the medium training set size, the classification accuracy degraded from approximately 94% to 92% when the zone location moved from 1.2 to 2.8 cm. That indicates the pre-focal zones are the most desirable zones for the medium size training set size. However, for the small training set size, the zones around the on-focus zone are the most desirable. Fig. 9 is useful to determine the most optimal zone center to characterize tissue samples for different training conditions. However, using a single zone is only meaningful when the phantoms are uniform and we do not lose any information by discarding other zones in our decision process. If there is some spatial information to be taken advantage of in our classification decision or we want to increase classification accuracies using all information that we have, then we need to separate the complete field of view into multiple zones and train multiple expert networks to be used in a voting schema. In that case, Fig. 9 is still useful for determining which expert network should have higher effect in a voting schema.

In Tables III–VIII, we presented confusion matrices to quantify classification accuracies with respect to different training set sizes, which are 10, 25, 20, 100, 200, and 500 ultrasound images. First, zone training had better classification accuracy than regular training when the training data were scarce. However, when the training data size was larger, regular training performed better than zone training. For example, when we used 200 or 500 ultrasound images in training, regular training performed around 1% better than zone training for all zones. However, when we used 10, 25 or 50 ultrasound images in training, zone training performed better than regular training. When we used 100 ultrasound images, regular training and zone training performed similarly. Second, depth-aware training was always better than regular training. Third, zone training was slightly better than depth-aware training when the training data were scarce. When we used 10, 25 or 50 ultrasound images in training, zone training performed better than depth-aware training for the pre-focal and on-focus zones. However, they performed similarly for the post-focal zone. Moreover, the post-focal pattern was more robust against decreasing training set size in comparison to the on-focus pattern; and the on-focus pattern was more robust in

comparison to the pre-focal pattern for all training strategies (*Zone, Regular, and Depth-Aware Training*). Finally, the results from the tables indicate that the training set should comprise data specifically from areas that are task relevant. For example, if a clinical imaging session investigates data in the pre-focal zone, but the training data came from the focal or post-focal zone, the classification performance may degrade significantly.

In this work, our proposed method was applied to tissue classification. However, it can be applicable to other DL applications such as detection, segmentation, and image formation. Therefore, zone training can be tested for different applications. Additionally, it would be interesting to test zone training with different types of neural network structures even though zone training is not directly related to the neural network structure. Moreover, the optimal number of zones should be investigated in greater detail. The optimal number of zones can change from problem to problem depending on the imaging substrates, imaging system, problem complexity, and imaging settings. It is also important to consider varying zone widths for different focal zones, as this can have a significant impact on the optimal number of zones, e.g., breaking the pre-focal zone into multiple, smaller zones might improve the accuracy in the pre-focal region. Furthermore, different patch sizes including pixel-wise classification, which is known as image segmentation, can be investigated within the context of zone training. Ultimately, we would like to identify the most data-efficient DL algorithm in the context of tissue characterization and zone training can be a useful tool in low data regimes for ultrasound imaging.

VI. CONCLUSION

We have presented a data-efficient DL strategy for tissue classification with ultrasound imaging, which we named zone training. Zone training has the ability to maintain high classification accuracy, while reducing training set size. Therefore, it should be considered as a robust approach for DL-powered ultrasound imaging in the context of tissue characterization.

REFERENCES

[1] Z. Akkus et al., "A survey of deep-learning applications in ultrasound: Artificial intelligence–powered ultrasound for improving clinical workflow," *J. Amer. College Radiol.*, vol. 16, no. 9, pp. 1318–1328, 2019.

[2] S. Liu et al., "Deep learning in medical ultrasound analysis: A review," *Engineering*, vol. 5, no. 2, pp. 261–275, Apr. 2019.

[3] X. Liu, J. Shi, and Q. Zhang, "Tumor classification by deep polynomial network and multiple kernel learning on small ultrasound image dataset," in *Proc. Int. Workshop Mach. Learn. Med. Imag.*, Cham, Switzerland: Springer, 2015, pp. 313–320.

[4] J. Shi, S. Zhou, X. Liu, Q. Zhang, M. Lu, and T. Wang, "Stacked deep polynomial network based representation learning for tumor classification with small ultrasound image dataset," *Neurocomputing*, vol. 194, pp. 87–94, Jun. 2016.

[5] T. N. Nguyen, A. S. Podkowa, T. H. Park, R. J. Miller, M. N. Do, and M. L. Oelze, "Use of a convolutional neural network and quantitative ultrasound for diagnosis of fatty liver," *Ultrasound Med. Biol.*, vol. 47, no. 3, pp. 556–568, Mar. 2021.

[6] A. Han et al., "Noninvasive diagnosis of nonalcoholic fatty liver disease and quantification of liver fat with radiofrequency ultrasound data using one-dimensional convolutional neural networks," *Radiology*, vol. 295, no. 2, pp. 342–350, May 2020.

[7] M. Byra et al., "Liver fat assessment in multiview sonography using transfer learning with convolutional neural networks," *J. Ultrasound Med.*, vol. 41, no. 1, pp. 175–184, Jan. 2022.

[8] Z. Cao et al., "Breast tumor detection in ultrasound images using deep learning," in *Proc. Int. Workshop Patch-Based Techn. Med. Imag.*, Cham, Switzerland: Springer, 2017, pp. 121–128.

[9] G. Carneiro, J. C. Nascimento, and A. Freitas, "The segmentation of the left ventricle of the heart from ultrasound data using deep learning architectures and derivative-based search methods," *IEEE Trans. Image Process.*, vol. 21, no. 3, pp. 968–982, Mar. 2012.

[10] G. Carneiro and J. C. Nascimento, "Combining multiple dynamic models and deep learning architectures for tracking the left ventricle endocardium in ultrasound data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2592–2607, Nov. 2013.

[11] Y. Zhang, M. T. C. Ying, L. Yang, A. T. Ahuja, and D. Z. Chen, "Coarse-to-fine stacked fully convolutional nets for lymph node segmentation in ultrasound images," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Dec. 2016, pp. 443–448.

[12] N. Baka, S. Leenstra, and T. V. Walsum, "Ultrasound aided vertebral level localization for lumbar surgery," *IEEE Trans. Med. Imag.*, vol. 36, no. 10, pp. 2138–2147, Oct. 2017.

[13] R. J. G. Van Sloun, R. Cohen, and Y. C. Eldar, "Deep learning in ultrasound imaging," *Proc. IEEE*, vol. 108, no. 1, pp. 11–29, Aug. 2020.

[14] A. C. Luchies and B. C. Byram, "Deep neural networks for ultrasound beamforming," *IEEE Trans. Med. Imag.*, vol. 37, no. 9, pp. 2010–2021, Sep. 2018.

[15] D. Xiao, W. M. K. Pitman, B. Y. S. Yiu, A. J. Y. Chee, and A. C. H. Yu, "Minimizing image quality loss after channel count reduction for plane wave ultrasound via deep learning inference," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 69, no. 10, pp. 2849–2861, Oct. 2022.

[16] J. Zhang, Q. He, Y. Xiao, H. Zheng, C. Wang, and J. Luo, "Ultrasound image reconstruction from plane wave radio-frequency data by self-supervised deep neural network," *Med. Image Anal.*, vol. 70, May 2021, Art. no. 102018.

[17] J. Lu, F. Millioz, D. Garcia, S. Salles, D. Ye, and D. Friboulet, "Complex convolutional neural networks for ultrafast ultrasound imaging reconstruction from in-phase/quadrature signal," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 69, no. 2, pp. 592–603, Feb. 2022.

[18] Y. Chen, J. Liu, X. Luo, and J. Luo, "ApodNet: Learning for high frame rate synthetic transmit aperture ultrasound imaging," *IEEE Trans. Med. Imag.*, vol. 40, no. 11, pp. 3190–3204, Nov. 2021.

[19] D. Perdios, M. Vonlanthen, F. Martinez, M. Arditi, and J.-P. Thiran, "CNN-based image reconstruction method for ultrafast ultrasound imaging," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 69, no. 4, pp. 1154–1168, Apr. 2022.

[20] M. Gasse, F. Millioz, E. Roux, D. Garcia, H. Liebgott, and D. Friboulet, "High-quality plane wave compounding using convolutional neural networks," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 64, no. 10, pp. 1637–1639, Oct. 2017.

[21] D. Hyun, L. L. Brickson, K. T. Looby, and J. J. Dahl, "Beamforming and speckle reduction using neural networks," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 66, no. 5, pp. 898–910, May 2019.

[22] A. K. Z. Tehrani, M. Sharifzadeh, E. Boctor, and H. Rivaz, "Bi-directional semi-supervised training of convolutional neural networks for ultrasound elastography displacement estimation," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 69, no. 4, pp. 1181–1190, Apr. 2022.

[23] X. Wei et al., "Unsupervised convolutional neural network for motion estimation in ultrasound elastography," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 69, no. 7, pp. 2236–2247, Jul. 2022.

[24] A. K. Z. Tehrani, M. Ashikuzzaman, and H. Rivaz, "Lateral strain imaging using self-supervised and physically inspired constraints in unsupervised regularized elastography," *IEEE Trans. Med. Imag.*, early access, Dec. 19, 2022, doi: 10.1109/TMI.2022.3230635.

[25] E. Evain, K. Faraz, T. Grenier, D. Garcia, M. D. Craene, and O. Bernard, "A pilot study on convolutional neural networks for motion estimation from ultrasound images," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 67, no. 12, pp. 2565–2573, Dec. 2020.

[26] R. Delaunay, Y. Hu, and T. Vercauteren, "An unsupervised learning approach to ultrasound strain elastography with spatio-temporal consistency," *Phys. Med. Biol.*, vol. 66, no. 17, Sep. 2021, Art. no. 175031.

[27] X. Liu, T. Zhou, M. Lu, Y. Yang, Q. He, and J. Luo, "Deep learning for ultrasound localization microscopy," *IEEE Trans. Med. Imag.*, vol. 39, no. 10, pp. 3064–3078, Oct. 2020.

[28] R. J. G. van Sloun et al., "Super-resolution ultrasound localization microscopy through deep learning," *IEEE Trans. Med. Imag.*, vol. 40, no. 3, pp. 829–839, Mar. 2021.

[29] T. N. Nguyen et al., "Characterizing fatty liver in vivo in rabbits, using quantitative ultrasound," *Ultrasound Med. Biol.*, vol. 45, no. 8, pp. 2049–2062, Aug. 2019.

[30] H. Su, V. Jampani, D. Sun, O. Gallo, E. Learned-Miller, and J. Kautz, "Pixel-adaptive convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 11166–11175.

[31] M.-H. Guo et al., "Attention mechanisms in computer vision: A survey," *Comput. Vis. Media*, vol. 8, no. 3, pp. 331–368, 2022.

[32] A. Adadi, "A survey on data-efficient algorithms in big data era," *J. Big Data*, vol. 8, no. 1, pp. 1–54, Jan. 2021.

[33] B. Shneiderman, "Human-centered AI," *Issues Sci. Technol.*, vol. 37, no. 2, pp. 56–61, 2021.

[34] K. A. Wear et al., "Interlaboratory comparison of ultrasonic backscatter coefficient measurements from 2 to 9 MHz," *J. Ultrasound Med.*, vol. 24, no. 9, pp. 1235–1250, 2005.

[35] E. L. Madsen, G. R. Frank, and F. Dong, "Liquid or solid ultrasonically tissue-mimicking materials with very low scatter," *Ultrasound Med. Biol.*, vol. 24, no. 4, pp. 535–542, 1998.

[36] J. J. Anderson et al., "Interlaboratory comparison of backscatter coefficient estimates for tissue-mimicking phantoms," *Ultrason. Imag.*, vol. 32, no. 1, pp. 48–64, Jan. 2010.

[37] E. L. Madsen, J. A. Zagzebski, R. A. Banjavie, and R. E. Jutila, "Tissue mimicking materials for ultrasound phantoms," *Med. Phys.*, vol. 5, no. 5, pp. 391–394, Sep. 1978.

[38] A. Paszke et al., "PyTorch: An imperative style, high-performance deep learning library," in *Proc. 33rd Conf. Neural Inf. Process. Syst. (NeurIPS)*, Vancouver, BC, Canada, vol. 32, May 2019, pp. 8026–8037.

[39] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

[40] M. Z. Alom et al., "The history began from AlexNet: A comprehensive survey on deep learning approaches," 2018, *arXiv:1803.01164*.

[41] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, vol. 25. Stateline, NV, USA, Dec. 2012, pp. 1097–1105.

Ufuk Soylu (Graduate Student Member, IEEE) received the B.S. degree in electrical engineering from Middle East Technical University, Ankara, Turkey, in 2018, and the M.S. degree in electrical engineering from the University of Illinois at Urbana–Champaign, Urbana, IL, USA, in 2020, where he is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering.

His current research interests are deep learning-based biomedical ultrasound imaging.

Michael L. Oelze (Senior Member, IEEE) was born in Hamilton, New Zealand, in 1971. He received the B.S. degree in physics and mathematics from Harding University, Searcy, AR, USA, in 1994, and the Ph.D. degree in physics from The University of Mississippi, Oxford, MS, USA, in 2000.

From 2000 to 2002, he was a Postdoctoral Researcher with the Bioacoustics Research Laboratory, Department of Electrical and Computer Engineering (ECE), University of Illinois at Urbana–Champaign (UIUC), Urbana, IL, USA. From 2002 to 2004, he was an NIH Fellow conducting research in quantitative ultrasound techniques for biomedical ultrasound applications in cancer detection. In 2005, he joined the Faculty of ECE, UIUC, where he is currently a Professor and the Associate Head for Graduate Affairs in ECE, a Health Innovator Professor with the Carle Illinois College of Medicine, and the Frederick G. and Elizabeth H. Nearing Scholar in the Grainger College of Engineering. His research interests include biomedical ultrasound, quantitative ultrasound imaging for improving cancer diagnostics and monitoring therapy response, ultrasound tomography, ultrasound-based therapy, beamforming, coded excitation, and communications using ultrasound.

Dr. Oelze is currently a fellow of the AIUM and a member of ASA. He is a member of the Technical Program Committee of the IEEE Ultrasonics Symposium. He currently serves as an Associate Editor for the IEEE TRANSACTIONS ON ULTRASONICS, FERROELECTRICS, AND FREQUENCY CONTROL, *Ultrasonic Imaging*, and the IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING.