A Data-Efficient Deep Learning Training Strategy for Biomedical Ultrasound Imaging: Zone Training

Ufuk Soylu and Michael L. Oelze

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 acquiring 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 deep learning training strategy, 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 required a factor of 2-5 less training data to achieve similar classification accuracies compared to a conventional training strategy.

Index Terms—Deep Learning, Tissue Classification, Biomedical Ultrasound Imaging

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 automatically rather than extracting features manually. Due to rapid increase in computational power and large data-sets, 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 (CNN) 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 remarkable results in the context of biomedical ultrasound imaging are classification, detection, segmentation and image reconstruction. Liu et al. [2] provide a comprehensive overview of the DL used in classification, detection and segmentation. A common classification application is classifying tumors into malignant or benign. Liu et al. [3] and Shi et al. [4] developed a supervised DL algorithm for tumor classification. As another example of classification, Nguyen et al. [5] demonstrated that CNNs are able to successfully classify liver tissues into fatty liver tissue and healthy liver tissue based on using raw RF backscattered data. Similarly, a common detection application is detecting tumors. Cao et al. [6] performed a comprehensive study for breast tumor detection by comparing state-of-the-art CNN based DL algorithms. Regarding segmentation, there are two types of segmentation: segmenting non-rigid organs such as left ventricle of the heart [7], [8] and segmenting rigid organs such as lymph nodes [9] or bones [10]. Moreover, Van Sloun et al. [11] provided a comprehensive overview of the DL used in ultrasound image reconstruction. As an example, even though delay-and-sum beamforming has become the standard due to its low computational cost, DL networks have been used as beamformers in multiple research studies to reduce any deterioration in image quality caused by constant speed of sound assumption or manual adjustment of channel weights in delay-and-sum beamforming [12]–[17]. Furthermore, DL algorithms have been employed in advanced ultrasound imaging applications such as super-resolution imaging of microvasculature structure via Ultrasound Localization Microscopy [18], [19].

Even though DL is promising for biomedical ultrasound imaging, there are certain roadblocks to wider adoption. A major roadblock is that acquiring large and diverse data-sets is expensive. Hence, data-efficient DL algorithms should be developed to overcome this limitation. Another roadblock is that there are huge variations in ultrasound images due to operator, patient or machine dependent factors. Therefore, improving robustness of DL algorithms against variations in ultrasound imaging 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]. 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 by using the data belonging to 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 performance in comparison to “Regular Training”.

In our experiments, we choose tissue classification as the primary application and test our proposed method to classify three distinct tissue-mimicking phantoms. Classifying tissues has recently evolved from model-based approaches such as quantitative ultrasound (QUS) techniques to model-free, DL-based techniques. Nguyen et al. [20] 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. In our study, we also use DL algorithms with RF ultrasound signal to utilize the phase and frequency-dependent information for the tissue classification problem.

Further details of our experiments can be found in Section II and III. Regarding our results, we present them in two categories. First, we investigate how to define zones optimally. Second, we investigate how much data is needed for “Zone Training” to achieve similar classification accuracy as “Regular Training”.

II. 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. 1 and their properties are summarized in Table I.

Phantom1, which mimicks human liver, has been described by Wear et al. [21]. Phantom1 had a measured attenuation coefficient slope of approximately 0.4 dB×cm⁻¹×MHz⁻¹. Its materials were produced based on the method of Madsen et al. [22] 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 to 90 μm-diameter glass beads (0.38 %), bovine milk concentrated 3 times by reverse osmosis (24.5 %), liquid Germall Plus preservative (International Specialty Products, Wayne, NJ) (1.88 %), and 18-MΩ-cm deionized water (66.3 %).

Phantom2 and Phantom3 are both low attenuation phantoms, whose properties have been described by Anderson et al. [23] and constructions have been described Madsen et al. [24]. 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×cm⁻¹×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.

![Tissue-mimicking Phantoms](image)

**TABLE I**

| Phantom | Phantom1 | Phantom2 | Phantom3 |
|---------|----------|----------|----------|
| Sphere diameter (μm) | 75-90 | 41 ± 2 | 50 ± 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 |

B. Ultrasound Imaging Device and Its Settings

Ultrasound gel was placed on the surfaces of the phantoms and then the phantoms were scanned with an L9-4/38 transducer using 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 total 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. Further imaging parameters can be found in Table II.

| Imaging Parameters          | Values         |
|-----------------------------|----------------|
| Center Frequency of Pulse   | 9 MHz          |
| Pulse Shape                 | Cosine signal  |
| Pulse Duration              | 1 Period       |
| Focus                       | Single at 2 cm |
| Output Power                | -5 dB          |
| Imaging Depth               | 4 cm           |

**C. Data-set**

After acquiring 1007 ultrasound images per phantom, we extracted square image patches whose sizes were 256 pixels × 256 pixels, to be used in training and test sets. The size of an ultrasound image frame was 2080 pixels × 256 pixels. There were 2080 samples along the axial direction that corresponds to 4 cm depth. Even though the L9-4/38 transducer has 128 channels, SonixOne system interpolates to 256 channels that correspond to 256 lateral pixels.

From one ultrasound image, we could extract 15 image patches when we used the complete field of view as in "Regular Training". While extracting image patches for "Regular Training", we didn’t use the first 400 pixels in the ultrasound image and obtained individual patches by jumping 100 pixels along the axial depth. The process of obtaining image patches for "Regular Training" is depicted in Fig. 2.

We now describe how to obtain image patches for "Zone Training". First, we developed definitions for the zones based on the diffraction pattern for a single focused transducer. In this work, we broke the set into five zones: a pre-focal zone centered at 1 cm, a pre-focal zone centered at 1.5 cm, an on focus zone centered at 2 cm, a post focal zone centered at 2.5 cm and a post focal zone centered at 3 cm. For each zone, we extracted three overlapping patches per ultrasound image: one patch at the zone center, one patch at 100 pixels above the zone center and one patch at 100 pixels below the zone center. The process of obtaining image patches for the on focus zone representing "Zone Training" is also depicted in Fig. 2.

Example image patches for each phantom and some zones can be found in Fig. 3. Additionally, the data-set of ultrasound images is also publicly available at https://osf.io/7ztg3/ (DOI 10.17605/OSF.IO/7ZTG3).

**D. Network Structure**

In this work, we used a CNN architecture named AlexNet [25]. CNNs have several advantages among other DL structures for the tasks related to 2D images. CNNs are similar to the human visual system, which makes them effective at learning and extracting abstractions of 2D images. They also have significantly fewer parameters and so they can be trained more efficiently [26]. Generally, CNNs have two parts: feature extractors that consist of convolution layers, max-pooling layers and non-linear activation functions, and a classifier that consists of fully connected layers and non-linear activation functions.

In 2012, Krizhevsky et al. [25] proposed a wider and deeper network, AlexNet, compared to LeNet and won the ImageNet challenge for visual object recognition called the ImageNet.
Large Scale Visual Recognition Challenge (ILSVRC). Alom et al. [26] described that moment as the historical moment where DL began. AlexNet was a significant development in the field of machine learning and computer vision for visual recognition and classification tasks and it is the point in history where interest from academia and industry in DL increased significantly [26]. The Zone Training approach can be implemented with any network structure. To able to solve our tissue classification problem, we make minor modifications to the original AlexNet structure. First, we modify the first convolution layer where it takes single input channel rather than threechannels. Second, it outputs three probabilities that correspond to three phantoms rather than 1000 classes in the original network. Further details related to the architecture can be found in Table III.

| Layer Name | Output Size | AlexNet |
|------------|-------------|---------|
| conv1      | 62 x 62 x 96 | 11 x 11, stride4, relu |
| maxpool1   | 30 x 30 x 96 | 3 x 3, stride2 |
| conv2      | 30 x 30 x 256 | 5 x 5, pad2, relu |
| maxpool2   | 14 x 14 x 256 | 3 x 3, stride2 |
| conv3      | 14 x 14 x 384 | 3 x 3, pad1, relu |
| conv4      | 14 x 14 x 384 | 3 x 3, pad1, relu |
| conv5      | 14 x 14 x 256 | 3 x 3, pad1, relu |
| maxpool3   | 6 x 6 x 256 | 3 x 3, stride2 |
| fully connected1 | 4096 | 9216 x 4096 connections, relu |
| fully connected2 | 4096 | 4096 x 4096 connections, relu |
| fully connected3 | 4096 | 4096 x 3 connections |

E. Training

DL training was done by using two machines each with a single GPU. One machine had TITAN RTX and the other machine had RTX A500. All implementations were done with the PyTorch library [27]. The batch number was chosen as 64 through out all experiments. In the training, dropout layers [28] with 0.5 probabilities were added to improve the regularization and deal with overfitting, before fully connected1 and fully connected2 layers as in the original implementation of AlexNet. Horizontal flip with 0.5 probability was implemented as a data augmentation step in the training process. Additionally, the models were trained by using cross entropy loss with uniform class weights, defined in (1), which includes built-in softmax function in PyTorch implementation [27],

\[ l(x, y) = \frac{1}{N} \sum_{n=1}^{N} -log \frac{\exp(x_{n,y_a})}{\sum_{c=1}^{C} \exp(x_{n,c})} \]  

(1)

where C is the number of classes, N is the batch size, \( x_{n,c} \) is the logit, which corresponds to class c and \( x_{n,y_a} \) is the logit with correct class \( y_a \).

Furthermore, the learning rate was fixed at 1e-5 and we used the Adam algorithm [29] as the optimizer in all experiments. Initial weights for AlexNet were chosen based on the original paper. The weights of convolutional layers were initialized by using a Gaussian distribution whose mean is 0 and standard deviation is 0.01. While the biases for conv1 and conv3 were initialized with 0, the biases for conv2, conv4, and conv5 are initialized with 1. Weights and biases for fully connected layers were set by PyTorch default settings.

In the experiments, we fixed all the hyper-parameters such as network structure, learning rate, etc. Therefore, we didn’t use any validation set to search for hyper-parameters. In the results, we directly report classification accuracies for the test set. We repeated each experiment 10 times and report mean and standard deviation of classification accuracies for each experiment. Regarding epoch number, in general, we picked the epoch number that gave the best classification accuracy for a given experiment. In Section III we report epoch number, mean classification accuracy and standard deviation for each experiment.

The process of forming training and test sets starts with randomly selecting the desired number of ultrasound images per phantom. Then, we extract patches as described in Section II-C. Subsequently, we randomly split the patches into training and testing sets. The ratio between training set size and test set size was four.

As we described in Section II we propose to divide the complete field of view into multiple zones. Then, we train separate neural networks per each zone by using the patches belonging to the corresponding zone. We name this strategy as “Zone Training” as opposed to ”Regular Training”, which uses all patches coming from the complete field of view to train a single neural network. Briefly, the main hypothesis 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. In Section III we present our results related to “Zone Training” and compare it to ”Regular Training”.

III. RESULTS

Results are organized in two subsections. In the first subsection, we present results, which help us to determine if our zone definitions are optimal by experimenting with axial zone widths, axial zone locations and by sweeping testing zone centers around training zone centers. Our purpose in this subsection is to determine an optimal way to divide the field of view into multiple zones, which is required for ”Zone Training”. In the second subsection, we investigate the relationship between training set size and classification accuracy for “Zone Training” and ”Regular Training”. Our purpose in this subsection is to demonstrate that ”Zone Training” requires less data, which is the main observation of this paper.

A. How To Determine Zones Optimally

We now present four results that are helpful in determining optimal 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 by 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 centered at 1.5 cm and the post focal zone centered 2.5 cm 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, the pre-focal zone at 1.5 cm and the post focal zone at 2.5 cm when we increase the number of overlapping patches used in patch extraction. The third result, together with first and second results, investigate diffraction pattern behaviour at the pre-focal, on focus and post focal zones. In the fourth result, we experiment with axial zone location and we plot classification accuracy at different zone centers.

In Fig. 4, classification accuracy is plotted as the testing zone center is swept by 1 cm towards and away from the transducer around the training zone center. We train a neural network (AlexNet) by using patches from the on focus zone centered at 2 cm depth, and then we test the neural network with patches from zones centered at 1 cm, 1.25 cm, 1.5 cm, 1.75 cm, 2 cm, 2.25 cm, 2.5 cm, 2.75 cm and 3 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, negative one means that the testing zone is 1 cm closer to transducer than the training zone and positive one means that the testing zone is 1 cm farther away from transducer than the training zone. We repeat the experiments for different sizes of training sets. We use 900 image patches, 1800 image patches and 7248 image patches in the training which correspond to 300 ultrasound images, 600 ultrasound images and 2416 ultrasound images (complete data-set), respectively. In the figure, colors indicate the size of the training set. Epoch numbers in the training are chosen as 150 for 300 ultrasound images, 120 for 600 ultrasound images and 110 for the complete data-set. Epoch numbers are chosen by observing classification accuracy in the test set and identifying convergence. Additionally, we repeat each experiment 10 times and plot average classification accuracy on the graph. We report standard deviations of 10 experiments as vertical error bars.

In Fig. 5, similar to Fig. 4, 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 centered at 1.5 cm and the post focal zone centered at 2.5 cm in addition to the on focus zone. When we train AlexNet by using patches from the pre-focal zone at 1.5 cm, we test the network with patches centered at 0.5 cm, 0.75 cm, 1 cm, 1.25 cm, 1.5 cm, 1.75 cm, 2 cm, 2.25 cm and 2.5 cm. When we train AlexNet by using patches from post focal zone at 2.5 cm, we test the network with patches centered at 1.5 cm, 1.75 cm, 2 cm, 2.25 cm, 2.5 cm, 2.75 cm, 3 cm, 3.25 cm and 3.5 cm. In this result, we use fixed training set size, which is 7248 image patches or 2416 ultrasound images. In the figure, colors represent the training zone. Epoch number was chosen as 110 in accordance with Figure 4. Moreover, the plot was formed by averaging 10 experiments where error bars represent standard deviations.

In Fig. 6, we plot classification accuracy as the y axis and axial zone width as the x axis. In Zone Training, we extract three overlapping patches per ultrasound image as described in Section III-B. In this result, we make an exception to experiment with zone width, which is defined in terms of number of patches. In the original zone definitions, we extracted three patches per zone from one ultrasound image. We now extract 3, 4, 5 and 6 patches per zone from each ultrasound image to be used in Zone Training and these numbers form the x axis. Training set size was fixed at 600 ultrasound images throughout the experiment, which corresponds to 1800, 2400, 3000, 3600 training image patches when we extract 3, 4, 5 and 6 patches from each ultrasound image, respectively. 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/testing zones. Additionally, the plot was formed by averaging 10 experiments where error bars represent standard deviations. Moreover, the epoch number was chosen as 120 in accordance with previous results.

In Fig. 7, we again plot classification accuracy as the y axis and zone center as the x axis. For this graph, we test and train networks from the same zone while sweeping the zone center axially. We train and test our networks for the pre-focal zone centered at 1 cm, the pre-focal zone centered at 1.5 cm, the on
focus zone, the post focal zone centered at 2.5 cm and the post focal zone centered at 3 cm. Epoch numbers were chosen as 150 for the small set, 120 for the medium set and 110 for the large set in accordance with Fig. 4. Each of the experiments were repeated 10 times and average classification accuracies are plotted in the graph, standard deviations are represented as the vertical error bars.

Fig. 6. Axial Zone Width: Classification accuracy vs axial zone width. The colors represent training/testing zones. The yellow color is for the post focal zone centered at 2.5 cm, which is labeled as postfocus. The blue color is for the on focus zone centered at 2 cm, which is labeled as onfocus. The orange color is for the pre-focal zone centered at 1.5 cm, which is labeled as prefocus one half.

Fig. 7. Determining Zone Center: Classification accuracy vs zone center for different data-set sizes. Colors represent training set sizes where the blue color is for 900 patches, the orange color is for 1800 patches and the yellow color is for 7248 patches.

B. Training Set Size vs Classification Accuracy

We now compare the proposed "Zone Training" with "Regular Training" to demonstrate that the proposed method requires less data to achieve similar accuracy. In Fig. 8, we plot classification accuracy as the y axis and training set size in terms of number of ultrasound images as the x axis. We repeat the experiments for Regular Training and five zones. Training set sizes were chosen as 100, 200, 300, 400 and 500 ultrasound images, and patch extraction for Regular Training and Zone Training was done as described in Section II-C. Average classification accuracies are plotted in the graph by repeating each experiment 10 times and standard deviations are indicated by the vertical error bars. Moreover, epoch numbers are chosen for each zone type and training set size when the classification accuracy reaches its peak and they vary between 160-110.

Fig. 8. Classification accuracy vs training set size in terms of patches. The blue color represents Regular Training; the orange color represents the on focus zone, the yellow color represents the pre-focal zone centered at 1.5 cm, green color represents the pre-focal zone centered at 1 cm, the post focal zone centered at 2.5 cm and the post focal zone centered at 3 cm.

Furthermore, Table IV, V and VI are confusion matrices that list the classification accuracies for different training and testing strategies by using training set sizes of 100, 200 and 300 ultrasound images. Rows represent training strategies: The first row, denoted as pre, is for training with patches from the pre-focal zone centered at 1.5 cm. The second row, denoted as on, is for training with patches from the on focus zone centered at 2 cm. The third row, denoted as post, is for training with patches from the post-focal zone centered at 2.5 cm. The last row, denoted as regular, is for training with Regular Training strategy. Columns represent testing strategies: testing with patches from the pre-focal zone centered at 1.5 cm, testing with patches from the on focus zone centered at 2 cm, testing with patches from the post-focal zone centered at 2.5 cm and testing with complete field of view, respectively from first to last column. All experiments were repeated 10 times and epoch numbers were fixed at 150.

### TABLE IV
|                  | Pre    | On    | Post   | Regular |
|------------------|--------|-------|--------|---------|
| Pre              | 92.9±4.1 | 75.2±4.9 | 86.1±3.7 | 64.9±3.0 |
| On               | 48.9±3.4 | 95.3±3.7 | 94.8±5.1 | 64.3±1.6 |
| Post             | 33.1±0.6 | 76.3±4.8 | 98.3±0.8 | 68.7±1.2 |
| Regular          | 61.0±3.2 | 91.2±5.9 | 96.5±1.3 | 88.9±2.0 |

### TABLE V
|                  | Pre    | On    | Post   | Regular |
|------------------|--------|-------|--------|---------|
| Pre              | 96.5±2.3 | 72.7±2.8 | 81.3±5.8 | 65.4±1.4 |
| On               | 49.4±2.6 | 96.5±1.6 | 95.4±2.4 | 60.9±1.6 |
| Post             | 33.4±0.3 | 80.6±3.7 | 99.2±0.9 | 66.6±1.5 |
| Regular          | 52.5±4.1 | 95.4±2.1 | 98.3±1.4 | 96.2±0.8 |
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. In Section III, we presented results to investigate if our zone definitions were optimal and to assess whether Zone Training reduced the needed training set size without losing classification accuracy.

In Fig. 4, we quantified how much classification accuracy decreased when the testing zone moved away from the training zone for training with patches from the on focus zone. We observed that as the testing zone moved towards the transducer, classification accuracy dropped faster and it was valid for small, medium and large training set sizes. When the testing zone moved closer to the transducer by 0.25 cm, classification accuracy dropped to 80 percent. However, when the testing zone moved away from the transducer by 0.25 cm, classification accuracy remained around 100 percent. Similarly, when the testing zone moved closer to the transducer by 0.5 cm, classification accuracy dropped to around 40 percent. However, when the testing zone moved away from the transducer by 0.5 cm, classification accuracy only dropped to 90 percent. The observation indicates that the pre-focal diffraction pattern is more complicated and it changes faster than the post focal diffraction pattern, which changes smoothly. This observation will repeat in Fig. 5 as well. Additionally, another observation is that classification accuracy dropped faster for the larger training set. For example, at 0 cm, classification accuracy for the large training set was the highest, while classification accuracy for the small training set was the lowest. However, at 1 cm, it was the opposite. This indicates that when we used a smaller training set, even though it performs worse for the case where the training zone and testing zone were matched, it could generalize better for other zones.

In Fig. 5, we quantified how much 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 all zones, which further verified our previous observation stating that the pre-focal pattern changed quickly in comparison to the post focal pattern. For example, when the testing zone was closer to the transducer by 0.5 cm, classification accuracies were 80 percent, 50 percent and 40 percent 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 0.5 cm, classification accuracies were around 100 percent, 80 percent and 70 percent for the post focal zone, the on focus zone and the pre-focal zone, respectively. Second, we observed that the post focal zone was the most robust zone against the shift in the testing zone, while the pre-focal zone was the least robust zone against the shift in the testing zone. Classification accuracy for the post focal zone remained constant and around 100 percent from -0.25 cm, (i.e., when the testing zone moved closer to the transducer by 0.25 cm), to 0.5 cm, (i.e., when the testing zone moved away from the transducer by 0.5 cm). On the other hand, classification accuracy for the pre-focal zone dropped to below 60 percent at -0.25 cm and below 80 percent at 0.25 cm shifts. Another minor observation was that for the pre-focal training, classification accuracy deteriorated slowly when the distance between the testing zone and training zone reached to around 0.25-0.5 cm. This is the point where the testing zone started to move from the pre-focal area to the on focus zone.

In Fig. 6, we investigated the relationship between classification accuracy and zone width in terms of overlapping patches for the pre-focal, the on focus and the post focal zones. We observed that classification accuracy remained constant when we used 3, 4, 5 and 6 overlapping patches in the post focal zone definition. Additionally, we observed that classification accuracy remained relatively constant when we used 3 and 4 overlapping patches but degraded when we used 5 and 6 overlapping patches in the on focus zone definition. Moreover, classification accuracy degraded immediately when we increased the number of overlapping patches in the pre-focal zone definition. These observations also verified our observations from Figs. 4 and 5. Overall, by considering Figs. 4, 5 and 6 together, the results suggest splitting the pre-focal zone into multiple zones due to its rapidly changing nature and have multiple narrow zones in the pre-focal zone. However, there is more flexibility with the post focal zone due to its smoothly changing nature. We can use wider post focal zone definitions. Potentially, the complete area after the focus can be defined as the post focal zone without losing any significant classification power.

In Fig. 7, we determined the best zone location axially in terms of classification accuracy. We observed that setting the zone center at the post focal zone around 2.5 cm and 3 cm was optimal. In other words, for using a single zone to classify the phantoms, it would be ideal to locate the zone center at the post focal zone. However, this scenario is only meaningful when the phantoms are uniform and we don’t loose 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 by 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. 7 is still useful for determining which expert network should have higher effect in the voting schema. The general observation from this figure is that while the post focal zones are the most reliable zones, the pre-focal zones are the least reliable zones.

Figure 8 provides important results where we compare the proposed “Zone Training” with “Regular Training” in terms of training set size. First, we observed that Regular Training had the lowest classification accuracy for all training set sizes.

| TABLE VI | Classification Accuracies with 300 ultrasound images |
|----------|------------------------------------------------------|
|          | Pre | On  | Post | Regular |
| Pre      | 97.3±3.3 | 75.2±2.0 | 79.7±5.1 | 64.8±1.9 |
| On       | 46.4±2.0 | 97.9±1.4 | 93.9±2.4 | 39.2±1.9 |
| Post     | 33.5±0.3 | 74.2±2.8 | 99.5±0.5 | 68.1±0.9 |
| Regular  | 49.4±3.1 | 95.0±2.4 | 97.7±2.2 | 96.9±1.3 |
of 100, 200, 300, 400 and 500 ultrasound images. On the other hand, the post focal zones had the highest accuracy, which was very close to 100 percent for all training set sizes. Among all zones, the pre-focal at 1.5 cm had the lowest classification accuracy, and therefore, it required more training data. The most remarkable observation from this figure was that Regular Training required at least 500 ultrasound images to achieve similar performance with the post focal training by using only 100 ultrasound images. In addition, Regular Training required at least 200 images to achieve similar performance with the on focus training and the pre-focal training by using only 100 ultrasound images. These observations verify that in order for “Regular Training” to achieve similar accuracies as “Zone Training”, it requires more training data.

Lastly, in Table VII we present confusion tables to quantify classification accuracies. While these tables verify our observations from Fig. 8, we make two additional observations. First, in Zone Training, we obtained neural networks that produced higher classification accuracies for only their zones in comparison to Regular Training. As an example, the post focal training provided classification accuracies around 98-99 percent for the post focal zone, which was higher than accuracies provided by Regular Training. However, training in the post focal zone provided significantly lower accuracies when using test data from the on focus zone and the pre-focal zone. Second, Regular Training suffers from having to learn the pre-focal zone centered at 1.5 cm and, therefore, produced significantly lower accuracies around 50 percent. This observation by itself could be useful in certain applications of DL powered ultrasound imaging and tissue classification.

In this work, our proposed method is applied to only classification. However, it is applicable to all other DL applications such as detection, segmentation and image formation. In future work, we will investigate zone training for those applications as well. Additionally, “Zone Training” is not related to the neural network structure; therefore it is applicable to all types of neural network structures.

V. CONCLUSION

We have presented a data-efficient DL training strategy for biomedical ultrasound imaging and classification, which we named Zone Training. Even though, in this work, Zone Training was investigated for a certain DL application and a specific DL network type, it is potentially applicable and beneficial to all DL applications and all network types. Therefore, Zone Training should be considered as an approach for DL powered ultrasound imaging and sample classification because of its ability to achieve high classification accuracy with less training data.

REFERENCES

[1] Z. Akkus, J. Cai, A. Boonrod, A. Zielinski, A. D. Weston, K. A. Philbrick, and B. J. Erickson, “A survey of deep-learning applications in ultrasound: Artificial intelligence–powered ultrasound for improving clinical workflow,” Journal of the American College of Radiology, vol. 16, no. 9, pp. 1318–1328, 2019.

[2] S. Liu, Y. Wang, X. Yang, B. Lei, L. Liu, S. X. Li, D. Ni, and T. Wang, “Deep learning in medical ultrasound analysis: A review,” Engineering, vol. 5, no. 2, pp. 261–275, 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 International Workshop on Machine Learning in Medical Imaging. 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, 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 in Medicine & Biology, vol. 47, no. 3, pp. 556–568, 2021.

[6] Z. Cao, L. Duan, G. Yang, T. Yue, Q. Chen, H. Fu, and Y. Xu, “Breast tumor detection in ultrasound images using deep learning,” in International Workshop on Patch-based Techniques in Medical Imaging. Springer, 2017, pp. 121–128.

[7] 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 Transactions on Image Processing, vol. 21, no. 3, pp. 968–982, 2011.

[8] G. Carneiro and J. C. Nascimento, “Combining multiple dynamic models and deep learning architectures for tracking the left ventricle endocardium in ultrasound data,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 11, pp. 2592–2607, 2013.

[9] Y. Zhang, M. T. 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 2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE, 2016, pp. 443–448.

[10] N. Yang, S. Leenstra, and T. van Walsum, “Ultrasound aided vertebral level localization for lumbar surgery,” IEEE Trans. Med. Im., vol. 36, no. 10, pp. 2138–2147, 2017.

[11] R. J. Van Sloun, R. Cohen, and Y. C. Eldar, “Deep learning in ultrasound imaging,” Proceedings of the IEEE, vol. 108, no. 1, pp. 11–29, 2019.

[12] S. Vedula, O. Senouf, G. Zurakhov, A. Bronstein, O. Michailovich, and M. Zibulevsky, “Learning beamforming in ultrasound imaging,” arXiv preprint arXiv:1812.08042, 2018.

[13] D. Perdios, A. Besson, M. Arditi, and I.-P. Thiran, “A deep learning approach to ultrasound image recovery,” in 2017 IEEE International Ultrasonics Symposium (IUS). IEEE, 2017, pp. 1–4.

[14] W. Simson, R. Göbl, M. Paschali, M. Krönke, S. Scheidhauer, W. Weber, and N. Navab, “End-to-end learning-based ultrasound reconstruction,” arXiv preprint arXiv:1904.04696, 2019.

[15] S. Khan, J. Huh, and J. C. Ye, “Universal deep beamformer for variable frequency ultrasound,” IEEE Ultras. Im. and Comput. Conf. (IEEE UIC), 2019.

[16] A. C. Luchies and B. C. Byram, “Deep neural networks for ultrasound beamforming,” IEEE transactions on medical imaging, vol. 37, no. 9, pp. 2010–2021, 2018.

[17] D. Hyun, L. L. Brickson, K. T. Looby, and J. J. Dahl, “Beamforming and speckle reduction using neural networks,” IEEE transactions on ultrasonics, ferroelectrics, and frequency control, vol. 66, no. 5, pp. 896–910, 2019.

[18] X. Liu, T. Zhou, M. Lu, Y. Yang, Q. He, and J. Luo, “Deep learning for ultrasound localization microscopy,” IEEE Trans. Med. Im., vol. 39, no. 10, pp. 3064–3078, 2020.

[19] R. J. Van Sloun, O. Solomon, M. Bruce, Z. K. Zhaing, H. Wijkstra, Y. C. Eldar, and M. Mischi, “Super-resolution ultrasound localization microscopy through deep learning,” IEEE Trans. Med. Im., 2020.

[20] T. N. Nguyen, A. S. Podkowa, A. Y. Tam, E. C. Arnold, R. J. Miller, T. H. Park, M. N. Do, and M. L. Oelze, “Characterizing fatty liver in vivo in rabbits, using quantitative ultrasound,” Ultrasound in Medicine & Biology, vol. 45, no. 8, pp. 2049–2062, 2019.

[21] K. A. Wear, T. A. Stiles, G. R. Frank, E. L. Madsen, F. Cheng, E. J. Feleppa, C. S. Hall, B. S. Kim, P. Lee, W. D. O’Brien Jr et al., “Interlaboratory comparison of ultrasound backscatter coefficient measurements from 2 to 9 mhz,” Journal of Ultrasound in Medicine, vol. 24, no. 9, pp. 1235–1250, 2005.

[22] E. L. Madsen, G. R. Frank, and F. Dong, “Liquid or solid ultrasoundically tissue-mimicking materials with very low scatter,” Ultrasound in Medicine & Biology, vol. 24, no. 4, pp. 535–542, 1998.

[23] J. J. Anderson et al., “Interlaboratory comparison of backscatter coefficient estimates for tissue-mimicking phantoms,” Ultrasonic Imaging, vol. 32, no. 1, pp. 48–64, 2010.

[24] E. L. Madsen, J. A. Zamborski, R. A. Banjavie, and R. E. Jutila, “Tissue mimicking materials for ultrasound phantoms,” Medical Physics, vol. 5, no. 5, pp. 391–394, 1978.
[25] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Advances in Neural Information Processing Systems*, vol. 25, pp. 1097–1105, 2012.

[26] M. Z. Alom, T. M. Taha, C. Yakopcic, S. Westberg, P. Sidike, M. S. Nasrin, B. C. Van Esen, A. A. S. Awwal, and V. K. Asari, “The history began from alexnet: A comprehensive survey on deep learning approaches,” *arXiv preprint arXiv:1803.01164*, 2018.

[27] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga et al., “Pytorch: An imperative style, high-performance deep learning library,” *Advances in Neural Information Processing Systems*, vol. 32, pp. 8026–8037, 2019.

[28] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: A simple way to prevent neural networks from overfitting,” *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[29] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization.” *arXiv preprint arXiv:1412.6980*, 2014.