Ultrasound Scatterer Density Classification Using Convolutional Neural Networks and Patch Statistics

Ali K. Z. Tehrani, Mina Amiri, Ivan M. Rosado-Mendez,
Timothy J. Hall, and Hassan Rivaz

***Abstract*—Quantitative ultrasound (QUS) can reveal crucial information on tissue properties such as scatterer density. If the scatterer density per resolution cell is above or below 10, the tissue is considered as fully developed speckle (FDS) or under-developed speckle (UDS), respectively. Conventionally, the scatterer density has been classified using estimated statistical parameters of the amplitude of backscattered echoes. However, if the patch size is small, the estimation is not accurate. These parameters are also highly dependent on imaging settings. In this paper, we adapt convolutional neural network (CNN) architectures for QUS, and train them using simulation data. We further improve the network's performance by utilizing patch statistics as additional input channels. Inspired by deep supervision and multi-task learning, we propose a second method to exploit patch statistics. We evaluate the networks using simulation data and experimental phantoms. We also compare our proposed methods with different classic and deep learning models and demonstrate their superior performance in the classification of tissues with different scatterer density values. The results also show that we are able to classify scatterer density in different imaging parameters with no need for a reference phantom. This work demonstrates the potential of CNNs in classifying scatterer density in ultrasound images.**

***Index Terms*—Quantitative Ultrasound, Scatterer density, Convolutional Neural Network, Patch statistics.**

I. Introduction

Ultrasound imaging is increasingly attracting the attention of researchers and clinicians due to being a real-time and non-ionizing imaging modality, and being less expensive and more portable compared to other medical imaging techniques. However, several types of artifacts make interpretation of ultrasound images difficult. Cells, collagen, microcalcifications, and other microstuctural components are often smaller than the wavelength of the ultrasound wave, and scatter the wave and create the granular appearance in B-mode images called speckles. The scattered signal from scatterers provides useful information about characteristics of the scatterers, which are highly related to the tissue properties. Quantitative ultrasound (QUS) measures the tissue characteristics by analysing the ultrasound signal [1]–[8]. It aims to provide quantitative estimations of tissue characteristics which cannot be otherwise obtained from the B-mode image. It has been employed in many different applications such as liver fibrosis and steatosis assessment [9], [10], bone quality measurement [11], breast tumor classification [12], [13] and cardiac tissue characterization [14]. Improving QUS techniques can eventually broaden the applications of this safe and cost-effective method in diagnosis and treatment of a large number of disorders.

QUS methods can be classified into two broad categories: spectral-based and envelope-based methods [15]. Parameters such as the backscatter coefficient and attenuation coefficient can be estimated by spectral-based methods, usually with a requirement of a reference phantom to remove system-dependent effects [5], [7], [16], [17]. In envelope-based methods, different characteristics of the tissue are usually estimated by analysing and modelling the envelope of the ultrasound RadioFrequency (RF) data by fitting a probability density function. The sample size, wave frequency, and the attenuation can affect the accuracy of the distribution modelling, and therefore its parameter estimations [18], [19].

Among different QUS parameters, the scatterer density has attracted a great attention. If there are many scatterers (more than 10 in a resolution cell (an ellipsoidal volume defined by - 6 dB point of the beam profile [20])), the envelope data is considered as a fully developed speckle (FDS), and when the number of scatterers is low, it is considered as under-developed speckle (UDS). Classifying scatterer density into FDS and UDS is very critical since for estimation of the QUS parameters, different assumptions must be taken for UDS regions. Disregarding the density of scatterers results in unreliable estimates of other QUS parameters [21]. Reliable classification of UDS and FDS can pave the way for differentiating tissues with many small scatterers from those with few strong scatterers, and potentially use them as disease biomarkers. In addition, many downstream ultrasound applications usually work better under FDS conditions, such as sensorless 3D ultrasound [22] and elastography [23]. Furthermore, the presence of UDS can also affect the accuracy and precision of other biomarkers that are currently being explored in different clinical settings, such as the evaluation of fat infiltration in the liver. Detecting this condition during data acquisition can help define ways to improve the accuracy of the biomarker by implementing spatial or angular compounding strategies to compensate for the limited samples of the scattering process under UDS [24].

The statistics of echo-envelope data, extracted by either model-based or model-free parameters, provide information about tissue properties. Model-based parameters try to fit a distribution to the envelope data. If envelope data is FDS, the RF data can be modelled by the Gaussian distribution;

A. K. Z. Tehrani, M. Amiri and H. Rivaz are with the Department of Electrical and Computer Engineering, Concordia University, Canada. Ivan M. Rosado-Mendez is with Universidad Nacional Autonoma de Mexico, Mexico. Timothy J. Hall is with the Department of Medical Physics, University of Wisconsin, United States. e-mail: A_Kafaei@encs.concordia.ca, Amirim@encs.concordia.ca, irosado@fisica.unam.mx, tjhall@wisc.edu and hrivaz@ece.concordia.ca

Manuscript received; revised

therefore, envelope values follow the Rayleigh distribution [1], [25], [26]. However, the Rayleigh distribution fails to model the envelope statistics of UDS regions and other distributions such as K-distribution [14], Homodyned K-distribution [27], [28] and Nakagami distribution [12] can be utilized. Among these, the Homodyned K-distribution is the most comprehensive but the most complex one that does not have a closed-form solution. The Nakagami distribution provides a good estimate of the envelope signal with low-complexity and is widely used in QUS studies.

The Nakagami distribution, applied to ultrasound data for the first time by Shankar *et al.* [12], can be used to describe different probability density functions, and to characterize various scatterer patterns in tissues. It has been shown to be useful in discriminating different scatterer and tissue types. The Nakagami image can depict tissue properties that are not visible in ultrasound B-mode images, and has been employed in several studies for tissue characterization [9], [29]–[31].

Model-free parameters such as the envelope signal to noise ratio (SNR), skewness (S) and entropy [26], [32] are statistical parameters that change with different scatterer distributions. Entropy-based parametric imaging is a QUS imaging technique, which uses a small sliding window throughout the image to measure the entropy (the overall level of variations) of the backscattered RF signal. It has been shown to be effective in differentiating tissues with different scattering properties, and can provide higher accuracy in a smaller patch size compared to Nakagami imaging [32].

Deep Learning (DL) techniques have been utilised in many fields of medical image processing. They have also proved useful in different ultrasound applications such as segmentation [33], [34], super resolution imaging [35]–[37] and elastography [38]–[40]. A few studies have also attempted to tackle the challenge of extracting quantitative measures from ultrasound images using DL techniques. Byra *et al.* [41] used Nakagami images to train a convolutional neural network for the task of breast lesion classification. Wang *et al.* [42] have proposed a 3D convolutional network for breast cancer detection. However, the appearance and even statistics of ultrasound images can change with changes in imaging parameters such as time gain compensation and focal points. Such changes are well studied in DL and are referred to as domain shift [43]. If not accounted for, domain shift renders DL estimates grossly inaccurate. In fact, this is one of the reasons that DL is less explored in QUS compared to other ultrasound applications.

In a recent work, we designed a CNN to classify FDS and UDS [44]. The CNN was fed with envelope data and the RF data spectrum from small patches of RF data, and was compared with a Multi-Layer Perceptron (MLP) classifier, which used SNR and skewness as inputs. We used patches to analyse a small area of the image and therefore, to provide a high resolution. The effect of patch size was also investigated (with patches sized 5 to 10 $\times$ wavelength). The results showed that in small patch sizes, the CNN outperformed the MLP classifier, whereas for larger patch sizes, where the statistics of the patch could be reliably estimated, the MLP classifier outperformed the CNN.

In another recent work, we segmented simulated images with three different scatterer densities using a U-Net [45]. We found that the network was able to segment precisely when the intensity difference between the inclusion and the background was high and thus the network could associate the intensity to the scatterer density.

In [46], the mean scatterer intensity, which is another QUS parameter, was estimated for the whole image. They assumed that all regions have FDS, which is a limiting factor in real ultrasound images. In this study, we aim to classify FDS from UDS regions using CNNs in small patches (Note that the patch size is different for simulation and experimental phantom data) where classical statistical parameters commonly used in QUS studies cannot be estimated accurately. Our ultimate goal is to reveal tissue scatterer information similar to that of Table I using ultrasound envelope patches. The intermediate goal of this paper is to quantitatively evaluate the scatterer density under arbitrary conditions (i.e. different imaging settings). We use the ultrasound envelope data as the input to the network, since statistics such as SNR and Nakagami parameters are histogram-based, meaning that they ignore image texture. We hypothesize that the texture of the ultrasound envelope image contains crucial information which can be useful to determine the density of scatterers.

We use a large amount of simulated data to train the networks, and test the networks on simulated and experimental phantom data. We show that the CNN networks are more robust to the domain shift [43] compared to statistics used in conventional QUS methods. We modify well-known classification networks such as MobileNet V2 [47], Inception [48], DenseNet [49] and ResNext [50] for the task at hand, and train them using simulation data. The aforementioned networks are tested on unseen phantom data which are being imaged by a different imaging setting. In the next step, we combine statistics with CNNs by two different methods: fusion strategy and deep supervision. Our contributions can be summarized as follows:

- Different CNN architectures are utilized to classify scatterer density using envelope data.
- A novel training strategy and input channel are proposed to avoid over-fitting on domain information which enabled us to classify ultrasound patches without any reference phantom.
- The networks are further improved by exploiting patch statistics.
- The three different classifiers of support vector machine (SVM), random forest and MLP are used to classify based on patch statistics.
- Experimental phantom data is employed to validate our work in different imaging settings.

II. Methods

In this section, we first describe different datasets we analysed. We then present the scatterer density classification methods developed in this work, which include both classical (SVM, random forest and MLP) and DL methods (CNN and CNN with patch statistics as additional inputs), and provide intuitions for using different inputs.

A. Data

We employed two different datasets to investigate the performance of our proposed methods as outlined below.

1) Simulation data: We simulated 200 phantoms of size 30 mm$\times$ 30 mm$\times$ 1 mm using the Field II pro toolbox [51], with the center frequency of 6.67 MHz. The sampling frequency was 100 MHz and the RF signals were then downsampled to 50 MHz. Other imaging parameters can be found in Supplementary Materials.

We randomly distributed point scatteres in the phantoms. In 100 FDS phantoms, we included 16 scatterers per resolution cell. In the remaining 100 UDS phantoms, we included 2 scatterers per resolution cell. The resolution cell size was determined by calculating the correlation between the data and a moving window in different regions [52]. The size was 0.15 mm^2 at the focal point (The out of plane resolution cell size is not computed). We randomly cropped 5000 patches of size 256$\times$32 (4.04 mm $\times$ 5 mm which is 17 and 21 $\times$ wavelength in axial and lateral directions, respectively) from different depths as the training set and 1000 patches as the validation set. For the test set, we simulated 20 more phantoms with a random scatterer density value of 2 or 16 $\pm$ 10% in order to make the test data more challenging. We randomly selected 500 patches from these phantoms as the test set to evaluate the methods. This dataset will be publicly available online at data.sonography.ai.

2) Experimental phantom: Three different phantoms were used to validate our method. The phantoms were of size 15cm$\times$ 5cm$\times$ 15cm, built from homogeneous mixture of agarose gel media and glass beads as scattering agents. The glass bead diameter range and bead concentration in the phantoms are reported in Table I. For more information on construction details, the speed of sound and attenuation coefficient of these phantoms refer to [53]. The phantoms were imaged using an 18L6 transducer operating at 10 MHz frequency using an Acuson S2000 scanner (Siemens Medical Solutions, Malvern, PA) and we used envelope of RF data which was acquired using Axius Direct Ultrasound Research Interface [54]. There are 456 A lines, separated by 0.1242 mm and the depth is 40mm. The sampling frequency is 40MHz and the exact operation frequency was 8.89MHz. However, because of attenuation, the center frequency of the spectrum was lower. We computed the resolution cell size using correlation method at different depths and it varied between 0.284 mm^3 (at the top where resolution was poor) and 0.036 mm^3 (at the focal point where resolution was the highest). This high variation of the resolution cell size can have an adverse effect on the classification, especially when this variation is not observed by the network during training. The numbers of scatterers per resolution cell for different depth are given in Table I.

We used the experimental phantoms as the test data to evaluate the performance of different models optimized or trained on the simulation data. Phantom A (high concentration) belongs to the FDS class and Phantoms B (medium concentration) and C (low concentration) belong to the UDS class. The B-mode images of these phantoms are illustrated in Supplementary Materials.

B. Classical Statistical Parameters

Several parameters have been proposed in literature for estimating the scatterer density in ultrasound images. SNR and skewness are among the most important parameters proposed to classify different scatterer densities:

$$R = SNR = \frac{\overline{A^v}}{\sqrt{\overline{A^{2v}} - (\overline{A^v})^2}}, \qquad S = skewness = \frac{\overline{(A^v - \overline{A^v})^3}}{(\overline{A^{2v}} - (\overline{A^v})^2)^{1.5}} \tag{1}$$

where A is the envelope of RF data, v is the signal power and $\overline{(...)}$ denotes mean operation. While in [1], v smaller than 1 was suggested due to having higher dynamic range and lower estimation error, Prager *et al.* proposed 1.8 as the optimal value [55] in terms of the estimation error. We analysed both recommended values of 0.5 and 1.8, and obtained significantly better results on the validation set using the 0.5 value (Area Under Curve (AUC) of 0.894 vs. 0.876 when employing the MLP, and 0.802 vs. 0.794 when employing the SVM classifier). We, therefore, set v to 0.5 in this study.

When the patch size is big enough, the estimation error of R and S, and therefore the classification error based on these parameters is low. But for small-size patches, the classification becomes difficult [1], [44]. This is especially important in clinical applications where tissues are rarely homogeneous and a large patch may include different scattering properties [26], [27].

Entropy has been employed for scatterer density classification [32]:

$$E = \sum_{n=1}^{N} p(i) log[p(i)] \tag{2}$$

where E denotes entropy, and statistical histogram of the envelope data square is represented by p, and N is the number of bins for calculating the histogram, which is arbitrarily set to 100 in this study (changing the number of bins might results in different outcomes and it is investigated in Supplementary Materials). Entropy increases as the density of scatterers increases (moving from UDS to FDS). The entropy measure is shown to be effective when using a small window for QUS analysis [32].

Another parameter that has been shown useful in estimation of scatterer density is the Nakagami model parameters m (a maximum likelihood estimator of the shape parameter) and T (a generalized likelihood ratio test statistic) [25]:

$$m = \frac{(\overline{A^2})^2}{var[A^2]}, \qquad T = 2K(log\frac{m^m}{\Gamma(m)} + (m-1)[\overline{log(I)} - log(\overline{I}) - 1]) \tag{3}$$

where A is the envelope data and Γ represents the Gamma function. I is a vector representing K independent and identically distributed samples of the intensity from a specific patch. Different values of m explain different properties; when the

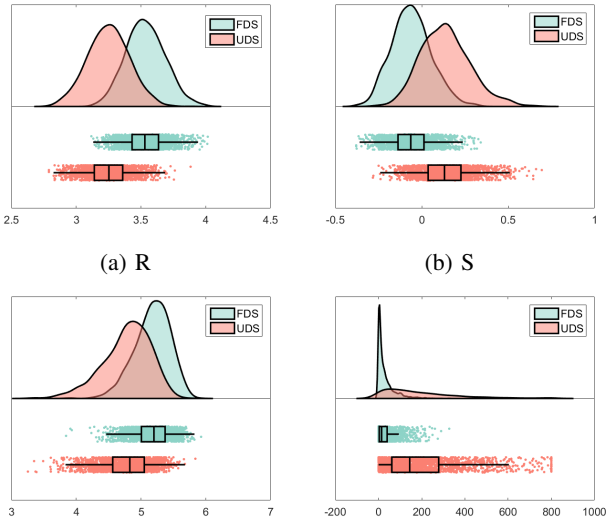
TABLE I: Characteristics of the experimental phantoms and their scatterer concentration per resolution cell using 18L6 transducer (the range shows the minimum and maximum values derived from different depths).

Phantom	Diameter of Random Scatterers (μm)	Scatterers Concentration per mm^3	Scatterers Concentration per resolution cell
A (High)	5-40	236	8.50-67
B (Medium)	75-90	9	0.32-2.55
C (Low)	126-151	3	0.11-0.85

m parameter approaches 1, the distribution approaches the Rayleigh distribution. The m parameter above and below 1 represent post- and pre-Rayleigh distributions, respectively, which are forms of a more general family of distributions, called Rician [1], [12], [20], [56].

There is a strong correlation between features m and T. The features m and R are also highly correlated. We therefore, remove the feature m from the feature list to eliminate the redundancy between different features. We consider aforementioned parameters as a set of features to classify FDS and UDS patches. The Supplementary Materials contain detailed correlation analysis of these statistical parameters.

Fig. 1 shows the distribution of different features, extracted from the simulation training data for UDS and FDS classes. The patch size is small so that for all features, a considerable overlap exists between the distributions of the two classes, which makes the classification highly erroneous using only a single feature. As opposed to our previous work [44] where only parameters R and S were used for classification, we use R, S, entropy and T together to obtain higher performance in classification. As shown in Fig. 1, the dynamic ranges of the features are not similar, hence we employ normalization (they are normalized to be in range 0-1) across each feature in the training data. The test and validation data were also normalized using the same coefficient obtained from the training data.



(a) R (b) S

(c) Entropy (d) T

Fig. 1: The distribution of the patch statistics for FDS and UDS in simulated training data. The patch size is small enough such that FDS and UDS classes overlap.

C. Machine Learning Methods

In order to classify FDS and UDS classes, we developed classical machine learning techniques in addition to DL methods. In this section, we describe the details of these classic techniques.

1) Support Vector Machine (SVM): We used SVM as a classical machine learning algorithm to classify FDS versus UDS. We analysed different SVMs with linear and non-linear (Radial Basic Function (RBF) and polynomial) kernels. An SVM with an RBF kernel led to the best results on the validation set, and was selected throughout this manuscript. We did a search to find optimum value of the C parameter of SVM, and subsequently, $C = 2.65$ was chosen.

2) Random forest classifier: Random forest is a learning method based on the decision tree algorithm and ensemble of different trees' outputs, and is among the top classification algorithms. By changing different parameters of a random forest model, we found the best performing model on the validation set, and used this model to classify different patches of simulation and experimental phantom data.

D. Deep Learning Methods

1) Multi-Layer Perceptron (MLP): We proposed an MLP structure to classify FDS and UDS groups. To find the best network architecture for classifying scatterer density using the aforementioned features, we investigated the performance of different MLP architectures on the validation data. We obtained the best results with a 3-layer network. Further increase in the number of layers did not improve the results and lead to overfitting, a common problem with MLPs. We also analysed different numbers of neurons in each hidden layer. We incrementally increased the number of neurons in two hidden layers. Including 128 neurons in the first hidden layer, and 32 neurons in the second hidden layer led to the best result. However, it is important to note that the results reached a plateau and did not change substantially by changing the number of neurons. We employed Dropout [57] in the second layer. The activation functions were Leaky Relu for the first two layers and Sigmoid for the last layer. The loss function was binary cross entropy and the network was optimized using the Adam optimizer.

2) Convolutional Neural Network (CNN): We used different state-of-the-art pretrained CNN networks to classify scatterer density. ResNext [50], MobileNet V2 [47], Inception [48] and DenseNet [49] were employed. In order to use

these networks, we replaced the last fully connected layer with a fully connected layer with one output having Sigmoid activation function. We used the pre-trained weights of the networks for initialization to facilitate the training and avoid over-fitting.

3) CNN with patch statistics as additional inputs: To further enhance the network, we proposed to utilize the patch statistics (R, S, entropy and T) as additional inputs. We tested different settings to determine the optimal way to fuse the information of patch statistics to the CNN. Fig. 2 shows the outcome (A denotes envelope). The CNN part is the same as the CNN network described in Section II-D2 and the patch statistic classifier part is similar to the MLP explained in Section II-D1. These parameters are fed to an MLP to generate a feature map, which is concatenated to the feature map obtained from the CNN. The resulted feature map is then used for a final classification.

Our first intuition was to train the whole network end to end. However, the CNN and the MLP have vastly different numbers of parameters and this resulted in a low generalization and a large sensitivity to the initial seeds. To mitigate the imbalanced number of parameters, we proposed training each part separately. We then trained the fusing part while the CNN and MLP weights were kept fixed. Please refer to Supplementary Materials for more information.

4) Deeply supervised CNN (Multi-task learning): Another way of using patch statistics is that using them as additional outputs. Multi-task learning [58] and deep supervision [59] have shown to improve the generality and performance of CNNs. We added R, S and m as additional outputs to force the networks learn in a way to have more generalization abilities. Values of R, S and m are highly correlated with the scatterer density, which led us to train the network to learn these features as additional outputs. The loss function for the networks with additional outputs was defined as a weighted summation of scatterer classification loss and parameter estimation part which can be written as:

$$\begin{aligned} loss = \frac{1}{N}\sum_{i=1}^{N}(y_i log(\widetilde{y_i}) + (1-y_i) log(1-\widetilde{y_i}) \\ +\beta \times ((\widetilde{R}-R)^2 + (\widetilde{S}-S)^2 + (\widetilde{m}-m)^2)) \end{aligned} \tag{4}$$

where y is the classification labels, $\widetilde{y}$ is the predicted scatterer density, R, $\widetilde{R}$, S, $\widetilde{S}$, m and $\widetilde{m}$ are ground truth and estimated values of SNR, Skewness and Nakagami parameter, respectively. N is the number of data in the corresponding mini-batch and β is the weight associated to the axillary loss. The classification performance is not very sensitive to β, which is set empirically to 0.2.

E. Training Schedule

To augment the data, random Gaussian noise, elastic deformation and random flipping in lateral direction were employed. The networks were trained with the Adam optimizer and the binary cross entropy was used as the loss function. Due to the fact that there were different networks with different inputs and to have a good generalization, we adopted a variant of early stopping which could be considered as a form of implicit regularization [60]. For early stopping, the validation AUC was selected as the stopping criteria; when the best validation AUC was reached during the training and remained the best after 20 epochs, we stopped the training. The cyclic learning rate was also used in order to avoid bad local minima [61].

1) Reduction of domain specific information: Batch normalization has been used widely in the classification networks. It has been found to facilitate the training and remove covariance shift [62]. During training, batch normalization layers compute the mean and standard deviation of the layer. The output of the layer is normalized by computed mean and standard deviation. During the test time, the computed mean and variance of the training data are used for normalization. The networks we employed contain several batch normalization layers. Domain information are mostly kept in the estimated mean and standard deviation of batch normalization layers [63]. In order to avoid learning domain specific information, we proposed not to update the mean and standard deviation of batch normalization layers during the training. We used pre-trained values of the mean and standard deviation for each batch normalization layers. The effect of batch normalization is studied in Section III-C.

F. Input Channels

In [28], log compression of envelope along with the envelope have been used ($log(A)$ and $A^2 \times log(A^2)$) for estimating statistics using the Homodyned K-distribution. Inspired by their work, we used $A \times log(A)$ as a novel input to the proposed CNNs. We therefore used the amplitude A and $A \times log(A)$. We also tested other inputs including $log(A)$ and $\sqrt{A}$ but did not observe any improvements. For brevity, the results are not included in the manuscript. The effect of including $A \times log(A)$) is studied in Section III-C.

G. Evaluation Metrics

To evaluate the classification performance, we used AUC of the Receiver Operating characteristic Curve (ROC), accuracy, sensitivity (recall), precision and also Youden's Index [64]. We estimated the 95% confidence interval of the metrics by employing boot strapping (i.e. sampling the data with replacement, for 1000 times). Youden's Index is a measure of both sensitivity and specificity:

$$\begin{aligned} J &= \frac{TP}{TP+FN} + \frac{TN}{TN+FP} - 1 \\ &= Sensitivity + Specificity - 1 \end{aligned} \tag{5}$$

where TP, FN, TN and FP denote true positive, false negative, true negative and false positive, respectively. It should be noted that accuracy, sensitivity and precision are reported for the threshold of 0.5. We calculated the Youden's Index for different threshold values and the highest values are compared between different models. The value of threshold which results in highest Youden's index can be considered as the optimal threshold value to have highest average of sensitivity and specificity. F1 score is also reported in the Supplementary Material.

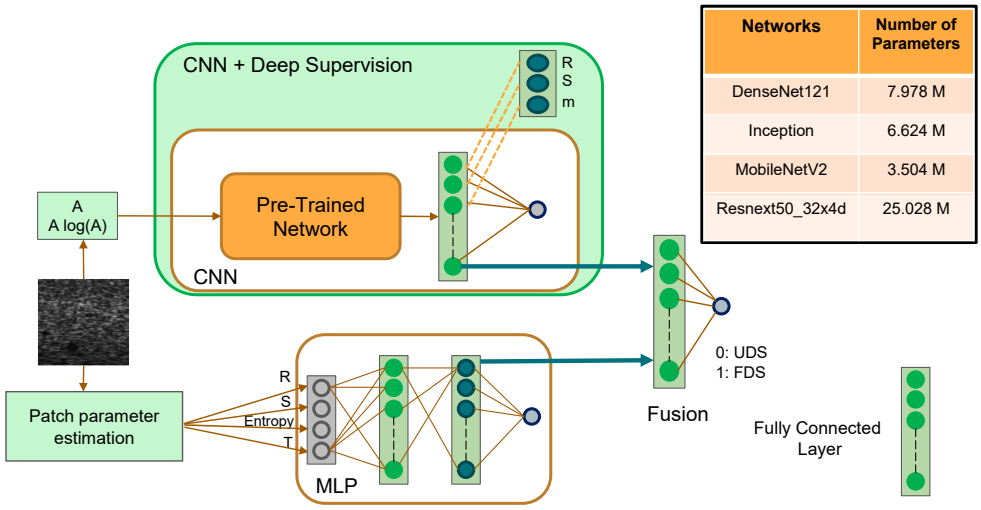


Fig. 2: Proposed architectures for different networks. MLP, CNNs, CNNs with deep supervision and CNNs with fusion with MLP. The pre-trained networks with their number of parameters in Million are specified.

TABLE II: Simulation Results

Model	Fusion	DS	AUC	Sensitivity	Precision	Accuracy	Youden's Index
SVM	✗	✗	0.892 (0.867-0.920)	0.816	0.800	0.802	0.620 (0.51)
Random Forest	✗	✗	0.894 (0.868-0.919)	0.816	0.806	0.806	0.620 (0.04)
MLP	✗	✗	0.890 (0.860-0.917)	0.850	0.786	0.806	0.610 (0.52)
MobileNet V2	✗	✗	0.949 (0.929-0.964)	0.867	0.870	0.866	0.7529 (0.31)
MobileNet V2	✓	✗	0.905 (0.887-0.927)	0.831	0.830	0.828	0.656 (0.50)
MobileNet V2	✗	✓	0.950 (0.930-0.965)	0.847	0.885	0.866	0.755 (0.38)
Inception	✗	✗	0.969 (0.952-0.981)	**0.988**	0.823	0.886	0.825 (0.90)
Inception	✓	✗	0.970 (0.953-0.981)	0.906	0.913	0.908	0.837 (0.61)
Inception	✗	✓	0.945 (0.923-0.960)	0.914	0.857	0.878	0.766 (0.45)
ResNext50_32x4d1	✗	✗	**0.975** (0.957-0.984)	0.926	0.920	**0.920**	0.848 (0.47)
ResNext50_32x4d1	✓	✗	0.918 (0.893-0.939)	0.851	0.819	0.828	0.677 (0.52)
ResNext50_32x4d1	✗	✓	0.973 (0.956-0.984)	**0.988**	0.850	0.906	**0.855** (0.73)
DenseNet121	✗	✗	0.964 (0.947-0.976)	0.863	0.917	0.889	0.798 (0.16)
DenseNet121	✓	✗	0.947 (0.923-0.961)	0.925	0.840	0.872	0.759 (0.59)
DenseNet121	✗	✓	0.967 (0.952-0.978)	0.851	**0.923**	0.888	0.801 (0.45)

III. Results

In this section, we provide the results of the proposed models for classification of FDS and UDS classes when analyzing different datasets. We tested three classifiers without including any CNN (i.e. a SVM, a random forest and an MLP model). Different CNNs were also analyzed, by exploiting the patch statistics using fusion (Fusion) and also deep supervision (DS). All DL models were trained according to the training schedule explained in Section II-E. The weights of the top-performing networks will be publicly available online after acceptance of this paper at code.sonography.ai.

A. Simulation Results

All proposed models were evaluated on the simulation data. Envelope and envelope multiplied by log compressed envelope are the input channels of all CNNs. Networks alone, fused with patch statistics (section II-D3) and with axillary outputs (section II-D4) are evaluated. The results are shown in Table II. The values inside the parenthesis in the AUC column represent the confidence interval and the value inside parenthesis in Youden's index represents the threshold which results in the highest Youden's index. As seen in Table II, CNN-based models provide better results compared to the MLP and SVM and random forest models which use only patch statistics. ResNext50_32x4d1 has the highest AUC (0.975) and accuracy (0.920). Whereas DenseNet121 with deep supervision has the highest precision (0.923). Both sensitivity and precision of CNN models are high meaning that the networks perform very well on identifying both positive and negative classes. The most of the optimum threshold of the Youden's Index was

close to 0.5 which is the threshold value used for classification.

B. Experimental Phantom Results

The results of classifying small patches from phantom A vs. phantoms B and C are provided in Table III. The patch size in terms of number of pixels is the same as the simulation data but it differs in terms of size in mm (4.92 mm$\times$ 4.28 mm). Due to the presence of domain shift, machine learning methods which use only patch statistics (SVM, random forest and MLP) have good precision but poor sensitivity. While, CNN models have high sensitivity and moderate precision (as opposed to simulation results where both sensitivity and precision were high). we expected that by adding patch statistics information, the sensitivity is decreased and precision improves. The effect of domain shift can be seen in the optimal threshold value of Youden's index. While in simulation the optimal values are close to 0.5, in experimental results most of them are far from 0.5.

Comparing the networks, MobileNet V2 has the highest sensitivity but the precision is low. Inception has the highest average of sensitivity and precision among the CNNs that do not exploit patch statistics. DenseNet121 achieves the highest precision (0.688) compared to other CNNs (without supervision or fusion). Exploiting patch statistics by fusion method increases the precision of the most of the networks (except Inception) and decreases the sensitivity in some models. By adding the patch statistics using deep supervision, the precision which is low in CNN models (0.562-0.688) was improved (0.628-0.809) with a slight decrease of sensitivity in some cases. Networks exploiting the patch statistics by deep supervision have higher AUCs compared to CNNs and CNNs with fusion. DenseNet + deep supervision achieves the highest accuracy among all the evaluated models. Please refer to Supplementary Materials for graphs of AUCs versus different networks.

C. Ablation Experiment

We conducted ablation experiments [65] (note that we are not referring to the thermal ablation treatment) on one of the networks (DenseNet121) to validate the input choice (Section II-F) and training strategy (Section II-E). The results are listed in Table IV. DenseNet121+BN denotes training DenseNet121 with updating batch normalization layers (training all layers). DenseNet121-Alog(A) represents that the network input is only envelope and the proposed input (Alog(A)) has been removed. Although precision of DenseNet121+BN is slightly better than the proposed method (0.755 Vs 0.688), the sensitivity is very poor (0.223 Vs 0.875). Accuracy is also lower (0.717 Vs 0.826); therefore, it can be concluded that freezing batch normalization statistics is beneficial. Regarding the selection of the input channels, it can be observed that adding envelope multiplied by log compressed envelope deteriorates the simulation results. However, it substantially improves the experimental phantom results which indicates that adding Alog(A) can reduce over-fitting on the simulation data.

D. Visualizing Experimental Phantoms and Ablation Experiment Results

Fig. 3 depicts some examples of the studied images using different models. We split each image into overlapping patches (50% overlap), and feed all patches to the networks. As seen in Fig. 3, CNNs perform very well in classification of patches from phantoms A and C but they perform modestly for patches from phantom B which belongs to the class 0 (UDS). Whereas, MLP produces low probability for all three phantoms. The main reason is that the statistics are also system dependent. It can be observed that by using proposed training strategy, the CNNs have less system dependency and produce better results compared to MLP, which uses only patch statistics. Among the CNNs, the model with deep supervision (DenseNet121+DS) has the best results on the challenging phantom B with introducing slight error on the phantom A. DenseNet121+BN which mean and standard deviation of batch normalization layers are trained as well as other layers, produces poor results on phantom A that shows the benefits of using the proposed training strategy.

We should also mention that the networks estimate higher probability of FDS in the top and bottom regions of the phantoms. The main reason is that in these regions, the resolution cell size is larger than that at the focal region, leading to a larger number of scatterers per resolution cell. Please refer to the Supplementary Materials for more results including ROC curve of different settings of DenseNet.

IV. Discussion

The density of scatterers in different parts of a tissue is an important property of that tissue which may discriminate normal and abnormal regions. Ultrasound images can be utilized to estimate this property non-invasively. This will eventually guide invasive procedures such as biopsy, leading to less expensive and safer diagnosis methods for different types of diseases.

In this work, we employed DL techniques to classify the scatterer density in ultrasound images. Based on our results, different numbers of scatterers result in different texture patterns in the ultrasound image. We used CNNs to detect texture patterns and employ histogram-based features (SNR, skewness, entropy and T) to improve the performance. Regarding the choice of features, we only used histogram-based features since CNNs can detect texture-based features but they need a large receptive field (as large as the entire patch) to be able to detect histogram-based features. Further investigations may reveal effectiveness of other features for CNNs.

The network trained on simulation data was able to classify the experimental phantom data, despite the fact that the number of scatterers and the imaging properties are completely different in these two datasets.

Comparing the two methods of exploiting patch statistics, deep supervised networks have higher AUCs than the fusion method. In addition to this, deep supervision does not need calculation of the patch statistics separately which reduce computation overhead. The networks we employed have a large receptive field which enables them to estimate statistic

TABLE III: Experimental phantom results.

Model	Fusion	DS	AUC	Sensitivity	Precision	Accuracy	Youden's Index
SVM	✗	✗	0.646 (0.623-0.669)	0.235	**0.873**	0.733	0.350 (0.03)
Random Forest	✗	✗	0.895 (0.880-0.913)	0.389	0.821	0.768	0.710 (0.46)
MLP	✗	✗	0.887 (0.879-0.895)	0.303	0.558	0.744	0.716 (0.20)
MobileNet V2	✗	✗	0.941 (0.935-0.948)	**0.978**	0.562	0.738	**0.830** (0.98)
MobileNet V2	✓	✗	0.886 (0.878-0.894)	0.446	0.754	0.767	0.707 (0.11)
MobileNet V2	✗	✓	**0.947** (0.940-0.953)	**0.978**	0.530	0.703	**0.830** (0.99)
Inception	✗	✗	0.895 (0.886-0.901)	0.945	0.685	0.835	0.726 (0.51)
Inception	✓	✗	0.890 (0.883-0.897)	0.950	0.636	0.802	0.684 (0.54)
Inception	✗	✓	0.946 (0.939-0.952)	0.975	0.617	0.790	0.825 (0.96)
ResNext50_32x4d1	✗	✗	0.872 (0.864-0.880)	0.967	0.576	0.752	0.720 (0.94)
ResNext50_32x4d1	✓	✗	0.905 (0.898-0.913)	0.467	0.809	0.786	0.746 (0.17)
ResNext50_32x4d1	✗	✓	0.932 (0.924-0.938)	0.977	0.544	0.720	0.783 (0.98)
DenseNet121	✗	✗	0.875 (0.866-0.883)	0.875	0.688	0.770	0.701 (0.15)
DenseNet121	✓	✗	0.900 (0.893-0.908)	0.872	0.770	0.869	0.789 (0.27)
DenseNet121	✗	✓	0.918 (0.911-0.925)	0.871	0.768	**0.870**	0.776 (0.21)

TABLE IV: Ablation experiment on DenseNet121. DenseNet121+BN : Batch normalization layers statistics are updated during training. DenseNet121-Alog(A) : Only envelope (A) is used as input and Alog(A) is not employed. DenseNet121: Proposed training strategy which batch normalization layers statistics are kept fixed during the training and Alog(A) is used alongside the envelope as an input channel.

	Simulation					Experimental Phantom				
Model	**AUC**	**Sensitivity**	**Precision**	**Accuracy**	**Youden's Index**	**AUC**	**Sensitivity**	**Precision**	**Accuracy**	**Youden's Index**
DenseNet121+BN	0.955 (0.935-0.969)	0.894	0.884	0.886	0.782 (0.23)	0.751 (0.738-0.762)	0.223	**0.755**	0.717	0.421 (0.02)
DenseNet121-Alog(A)	**0.966** (0.950-0.977)	**0.922**	0.900	**0.908**	**0.816** (0.57)	0.781 (0.771-0.790)	0.445	0.568	0.702	0.529 (0.02)
DenseNet121	0.964 (0.947-0.976)	0.863	**0.917**	0.890	0.798 (0.16)	0.875 (0.866-0.883)	**0.875**	0.688	**0.826**	**0.701** (0.15)

information related to the entire patch. Networks with a small receptive field are not able to estimate these statistics as well as deeper ones.

In a fixed imaging setting, a larger number of scatterers results in a brighter ultrasound image. However, by changing the imaging machine settings, the image intensity can vary. Even though the average density of ultrasound images contains information about the scatterers concentration, it is not a reliable feature for classifying the number of scatterers, as it can be easily altered by changing the imaging setting. We eliminated the effect of the average intensity by normalizing each individual patch such that the intensity of all studied patches was in the range [0,1]. It should be noted that the normalization method is not robust for regions having saturation. Those regions should be excluded to be able to use the proposed method correctly. We also reduced the domain related information by avoiding updating statistic coefficients of batch normalization layers during the training. Using reference phantoms to reduce the system specific effects can be an area of future works.

The effective number of scatterers per resolution cell varies by depth. Generally, at the focal point, the resolution cell is the smallest. Therefore, there are fewer scatterers per resolution cell at the focal point compared to other regions.

We included the data recorded from three phantoms in this study. The density of the scatterers is not the only parameter which differs between these phantoms. The size of the included scatterers is also different (Table I). However, considering the operating frequency, the size of the scatterers is still smaller than the wavelength, and does not substantially affect the results. Phantoms with different scatterer sizes and densities warrants further investigations.

V. Conclusion

In this manuscript, we proposed different CNN models to classify small patches of ultrasound images as FDS or UDS. We proposed to use both envelope and envelope multiplied by log compressed envelope as two separate input channels to the proposed CNNs. We also proposed to freeze batch normalization layers during the training to avoid learning domain specific information. We further benefited from patch statistics using fusion and axillary outputs (deep supervision). We did not use any data from experimental phantom during the training phase, which can degrade the results due to the domain shift. Nevertheless, we found that CNN models result in high sensitivity but moderate precision for experimental phantoms, due to this domain shift. The precision is improved by using statistic information as additional inputs/outputs with a slight decrease in sensitivity. On a broad picture, these results show the ability of our model to learn scatterer density from simulation data alone without the need to perform transfer learning on experimental data.

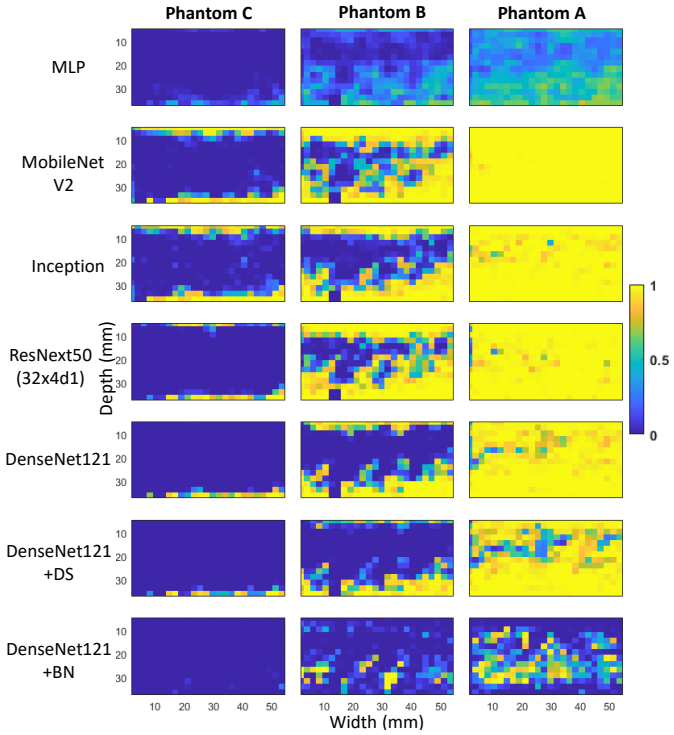


Fig. 3: The results of MLP, MobileNet V2, Inception, ResNext50, DenseNet121, DenseNet121+DS (with deep supervision) and DenseNet121+BN (with updating batch norm coefficients) models on the experimental phantoms. The color code represents the predicted output of the networks, from 0 (UDS) to 1 (FDS). Correct classes are 0 (UDS) for phantoms C and B, and 1 for phantom A.

ACKNOWLEDGMENT

We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC) RGPIN-2020-04612.

REFERENCES

[1] V. Dutt and J. F. Greenleaf, "Speckle analysis using signal to noise ratios based on fractional order moments," *Ultrasonic Imaging*, vol. 17, pp. 251–268, 1995.

[2] K. Nam, J. A. Zagzebski, and T. J. Hall, "Simultaneous backscatter and attenuation estimation using a least squares method with constraints," *Ultrasound in Medicine and Biology*, vol. 37, no. 12, pp. 2096–2104, 2011.

[3] M. F. Insana, R. F. Wagner, D. G. Brown, and T. J. Hall, "Describing small-scale structure in random media using pulse-echo ultrasound," *The Journal of the Acoustical Society of America*, vol. 87, no. 1, pp. 179–192, 1990.

[4] J. Mamou and M. L. Oelze, *Quantitative ultrasound in soft tissues*. Springer, 2013.

[5] Z. Vajihi, I. M. Rosado-Mendez, T. J. Hall, and H. Rivaz, "Low variance estimation of backscatter quantitative ultrasound parameters using dynamic programming," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 65, no. 11, pp. 2042–2053, 2018.

[6] J. Rouyer, T. Cueva, T. Yamamoto, A. Portal, and R. J. Lavarello, "In vivo estimation of attenuation and backscatter coefficients from human thyroids," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 63, no. 9, pp. 1253–1261, 2016.

[7] F. L. Lizzi, S. K. Alam, S. Mikaelian, P. Lee, and E. J. Feleppa, "On the statistics of ultrasonic spectral parameters," *Ultrasound in medicine & biology*, vol. 32, no. 11, pp. 1671–1685, 2006.

[8] N. J. Bureau, F. Destrempes, S. Acid, E. Lungu, T. Moser, J. Michaud, and G. Cloutier, "Diagnostic accuracy of echo envelope statistical modeling compared to b-mode and power doppler ultrasound imaging in patients with clinically diagnosed lateral epicondylosis of the elbow," *Journal of Ultrasound in Medicine*, vol. 38, no. 10, pp. 2631–2641, 2019.

[9] M.-C. Ho *et al.*, "Using ultrasound nakagami imaging to assess liver fibrosis in rats," *Ultrasonics*, vol. 52, no. 2, pp. 215–222, 2012.

[10] A. M. Pirmoazen, A. Khurana, A. El Kaffas, and A. Kamaya, "Quantitative ultrasound approaches for diagnosis and monitoring hepatic steatosis in nonalcoholic fatty liver disease," *Theranostics*, vol. 10, no. 9, p. 4277, 2020.

[11] K.-Y. Chin and S. Ima-Nirwana, "Calcaneal quantitative ultrasound as a determinant of bone health status: what properties of bone does it reflect?," *International journal of medical sciences*, vol. 10, no. 12, p. 1778, 2013.

[12] P. M. Shankar *et al.*, "Classification of ultrasonic b-mode images of breast masses using nakagami distribution," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 48, no. 2, pp. 569–580, 2001.

[13] A. Larrue and J. A. Noble, "Modeling of errors in nakagami imaging: Illustration on breast mass characterization," *Ultrasound in Medicine and Biology*, vol. 40, no. 5, pp. 917–930, 2014.

[14] L. Clifford, P. Fitzgerald, and D. James, "Non-rayleigh first-order statistics of ultrasonic backscatter from normal myocardium," *Ultrasound in Medicine and Biology*, vol. 19, no. 6, pp. 487–495, 1993.

[15] M. L. Oelze and J. Mamou, "Review of quantitative ultrasound: Envelope statistics and backscatter coefficient imaging and contributions to diagnostic ultrasound," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 63, no. 2, pp. 336–351, 2016.

[16] L. X. Yao, J. A. Zagzebski, and E. L. Madsen, "Backscatter coefficient measurements using a reference phantom to extract depth-dependent instrumentation factors," *Ultrasonic imaging*, vol. 12, no. 1, pp. 58–70, 1990.

[17] F. Deeba *et al.*, "A spatially weighted regularization method for attenuation coefficient estimation," in *2019 IEEE International Ultrasonics Symposium (IUS)*, pp. 2023–2026, IEEE, 2019.

[18] G. Cloutier, M. Daronat, D. Savéry, D. Garcia, L.-G. Durand, and F. S. Foster, "Non-gaussian statistics and temporal variations of the ultrasound signal backscattered by blood at frequencies between 10 and 58 mhz," *The Journal of the Acoustical Society of America*, vol. 116, no. 1, pp. 566–577, 2004.

[19] P.-H. Tsui, Z. Zhou, Y.-H. Lin, C.-M. Hung, S.-J. Chung, and Y.-L. Wan, "Effect of ultrasound frequency on the nakagami statistics of human liver tissues," *PloS one*, vol. 12, no. 8, 2017.

[20] R. F. Wagner, "Statistics of speckle in ultrasound b-scans," *IEEE Trans. Sonics & Ultrason.*, vol. 30, no. 3, pp. 156–163, 1983.

[21] E. Jakeman, "Speckle statistics with a small number of scatterers," *Optical Engineering*, vol. 23, no. 4, p. 234453, 1984.

[22] A. H. Gee, R. J. Housden, P. Hassenpflug, G. M. Treece, and R. W. Prager, "Sensorless freehand 3d ultrasound in real tissue: speckle decorrelation without fully developed speckle," *Medical image analysis*, vol. 10, no. 2, pp. 137–149, 2006.

[23] J. Ophir, I. Cespedes, H. Ponnekanti, Y. Yazdi, and X. Li, "Elastography: a quantitative method for imaging the elasticity of biological tissues," *Ultrasonic imaging*, vol. 13, no. 2, pp. 111–134, 1991.

[24] N. Rubert and T. Varghese, "Scatterer number density considerations in reference phantom-based attenuation estimation," *Ultrasound in medicine & biology*, vol. 40, no. 7, pp. 1680–1696, 2014.

[25] I. M. Rosado-Mendez, L. C. Drehfal, J. A. Zagzebski, and T. J. Hall, "Analysis of coherent and diffuse scattering using a reference phantom," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 63, no. 9, pp. 1306–1320, 2016.

[26] H. Rivaz, E. M. Boctor, and G. Fichtinger, "Ultrasound speckle detection using low order moments," in *2006 IEEE Ultrasonics Symposium*, pp. 2092–2095, Oct 2006.

[27] R. W. Prager, A. H. Gee, G. M. Treece, and L. H. Berman, "Analysis of speckle in ultrasound images using fractional order statistics and the homodyned k-distribution," *Ultrasonics*, vol. 40, no. 1-8, pp. 133–137, 2002.

[28] F. Destrempes, J. Porée, and G. Cloutier, "Estimation method of the homodyned k-distribution based on the mean intensity and two log-moments," *SIAM journal on imaging sciences*, vol. 6, no. 3, pp. 1499–1530, 2013.

[29] P.-H. Tsui, C.-K. Yeh, C.-C. Chang, and Y.-Y. Liao, "Classification of breast masses by ultrasonic nakagami imaging: a feasibility study," *Physics in Medicine & Biology*, vol. 53, no. 21, p. 6027, 2008.

[30] P.-H. Tsui, C.-K. Yeh, C.-C. Chang, and W.-S. Chen, "Performance evaluation of ultrasonic nakagami image in tissue characterization," *Ultrasonic imaging*, vol. 30, no. 2, pp. 78–94, 2008.

[31] P.-H. Tsui, C.-C. Huang, L. Sun, S. H. Dailey, and K. K. Shung, "Characterization of lamina propria and vocal muscle in human vocal fold tissue by ultrasound nakagami imaging," *Medical physics*, vol. 38, no. 4, pp. 2019–2026, 2011.

[32] P.-H. Tsui *et al.*, "Small-window parametric imaging based on information entropy for ultrasound tissue characterization," *Scientific Reports*, vol. 7, no. 41004, 2017.

[33] M. Amiri, R. Brooks, B. Behboodi, and H. Rivaz, "Two-stage ultrasound image segmentation using u-net and test time augmentation," *International journal of computer assisted radiology and surgery*, 2020.

[34] K. M. Meiburger, U. R. Acharya, and F. Molinari, "Automated localization and segmentation techniques for b-mode ultrasound images: A review," *Computers in biology and medicine*, vol. 92, pp. 210–235, 2018.

[35] R. J. Van Sloun, O. Solomon, M. Bruce, Z. Z. Khaing, Y. C. Eldar, and M. Mischi, "Deep learning for super-resolution vascular ultrasound imaging," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1055–1059, IEEE, 2019.

[36] K. G. Brown, D. Ghosh, and K. Hoyt, "Deep learning of spatiotemporal filtering for fast super-resolution ultrasound imaging," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 2020.

[37] S. Goudarzi, A. Asif, and H. Rivaz, "Multi-focus ultrasound imaging using generative adversarial networks," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pp. 1118–1121, 2019.

[38] Z. Gao *et al.*, "Learning the implicit strain reconstruction in ultrasound elastography using privileged information," *Medical image analysis*, vol. 58, p. 101534, 2019.

[39] E. Evain, K. Faraz, T. Grenier, D. Garcia, M. De Craene, and O. Bernard, "A pilot study on convolutional neural networks for motion estimation from ultrasound images," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 2020.

[40] A. K. Z. Tehrani and H. Rivaz, "Displacement estimation in ultrasound elastography using pyramidal convolutional neural network," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, pp. 1–1, 2020.

[41] M. Byra, H. Piotrzkowska-Wróblewska, K. Dobruch-Sobczak, and A. Nowicki, "Combining nakagami imaging and convolutional neural network for breast lesion classification," in *2017 IEEE International Ultrasonics Symposium (IUS)*, pp. 1–4, 2017.

[42] Y. Wang *et al.*, "Deeply-supervised networks with threshold loss for cancer detection in automated breast ultrasound," *IEEE Transactions on Medical Imaging*, vol. 39, no. 4, pp. 866–876, 2020.

[43] K. Stacke, G. Eilertsen, J. Unger, and C. Lundström, "A closer look at domain shift for deep learning in histopathology," *arXiv preprint arXiv:1909.11575*, 2019.

[44] A. K. Tehrani, M. Amiri, I. M. Rosado-Mendez, T. J. Hall, and H. Rivaz, "A pilot study on scatterer density classification of ultrasound images using deep neural networks," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 2059–2062, IEEE, 2020.

[45] M. Amiri, A. K. Tehrani, and H. Rivaz, "Segmentation of ultrasound images based on scatterer density using u-net," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 2063–2066, IEEE, 2020.

[46] L. Zhang, V. Vishnevskiy, and O. Goksel, "Deep network for scatterer distribution estimation for ultrasound image simulation," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 2020.

[47] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4510–4520, 2018.

[48] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818–2826, 2016.

[49] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700–4708, 2017.

[50] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1492–1500, 2017.

[51] J. A. Jensen, "Field: A program for simulating ultrasound systems," in *10th Nordicbaltic Conference On Biomedical Imaging, VOL. 4, Supplement 1, PART 1: 351–353*, Citeseer, 1996.

[52] H. Rivaz, R. Zellars, G. Hager, G. Fichtinger, and E. Boctor, "9c-1 beam steering approach for speckle characterization and out-of-plane motion estimation in real tissue," in *2007 IEEE Ultrasonics Symposium Proceedings*, pp. 781–784, IEEE, 2007.

[53] I. M. Rosado-Mendez, *Advanced spectral analysis methods for quantification of coherent ultrasound scattering: Applications in the breast*. PhD thesis, The University of Wisconsin-Madison, 2014.

[54] S. S. Brunke, M. F. Insana, J. J. Dahl, C. Hansen, M. Ashfaq, and H. Ermert, "An ultrasound research interface for a clinical system.," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 54, no. 1, pp. 198–210, 2007.

[55] R. Prager, A. Gee, G. Treece, and L. Berman, "Speckle detection in ultrasound images using first order statistics," *University of Cambridge, Department of Engineering*, 2001.

[56] T. Tuthill, R. Sperry, and K. Parker, "Deviations from rayleigh statistics in ultrasonic speckle," *Ultrasonic imaging*, vol. 10, no. 2, pp. 81–89, 1988.

[57] 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.

[58] Y. Zhang and Q. Yang, "An overview of multi-task learning," *National Science Review*, vol. 5, no. 1, pp. 30–43, 2018.

[59] L. Wang, C.-Y. Lee, Z. Tu, and S. Lazebnik, "Training deeper convolutional networks with deep supervision," *arXiv preprint arXiv:1505.02496*, 2015.

[60] C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals, "Understanding deep learning requires rethinking generalization," *arXiv preprint arXiv:1611.03530*, 2016.

[61] L. N. Smith, "Cyclical learning rates for training neural networks," in *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 464–472, IEEE, 2017.

[62] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.

[63] Y. Li, N. Wang, J. Shi, X. Hou, and J. Liu, "Adaptive batch normalization for practical domain adaptation," *Pattern Recognition*, vol. 80, pp. 109–117, 2018.

[64] W. J. Youden, "Index for rating diagnostic tests," *Cancer*, vol. 3, no. 1, pp. 32–35, 1950.

[65] L. Du, "How much deep learning does neural style transfer really need? an ablation study," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 3150–3159, 2020.