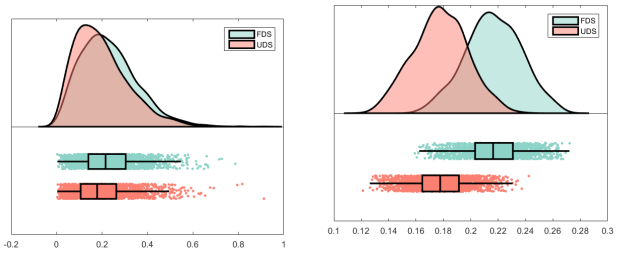
Supplimentary Material for Ultrasound Scatterer Density Classification Using Convolutional Neural Networks and Patch Statistics

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



(a) Envelope Amplitudes (b) Mean of Envelope Amplitudes

Fig. S1: The distribution of the patch statistics for FDS and UDS in simulated training data.

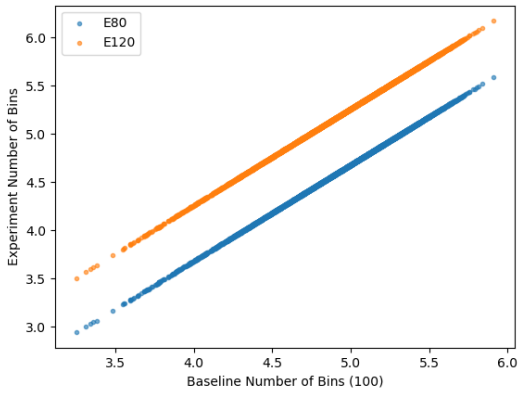


Fig. S2: Entropy with 80 (blue) and 120 (orange) bins versus 100 bins.

TABLE SI: Correlation between the statistical parameters.

	R	S	Entropy	m	T
R	1	-0.783	0.621	0.998	-0.883
S	-0.783	1	-0.778	-0.808	0.711
Entropy	0.621	-0.778	1	0.651	-0.575
m	0.998	-0808	0.651	1	-0.884
T	-0.883	0.711	-0.575	-0.884	1

A. Distribution of Envelope

Similar to Fig. 1 of the paper, we provided the distribution of all envelope amplitudes and the average of each patch in Fig. S1.

B. Correlation Analysis of QUS Features

We used Spearman Rho metric to analyse the correlation between the statistical features. The correlations between the features are given in Table SI. It is evident from the Table that m is highly correlated with R and T. Therefore, we removed m from our feature space.

C. Entropy Number of Bins

We used 100 bins to compute the histogram and subsequently entropy which changes with number of bins. Here, we computed the entropy of training data patches using 80, 100 and 120 bins which are shown in Fig. S2. The entropy with 80 bins (blue) and 120 bins (orange) are plotted versus 120 bins. We then used linear curve fitting tools to find the slopes and biases of the fitted lines. The slopes of the fitted lines for 80 and 120 bins versus 100 bins were 1.00 and 0.998, respectively. The bias were -0.31 (80 bins) and 0.25 (120 bins). According to these results, only the bias of computed entropy changes with different number of bins.

D. End to End Training

In the manuscript, we mentioned that training the network end to end results in low generalization and sensitivity to the initial point. To show this, we trained a CNN fused with MLP end to end 5 times with different initial points. The AUCs of simulation test data and experimental phantoms are given in Table SIV. The AUCs of experimental phantom CNN with fusion vary between 0.548-0.899. By training separately we avoided such large variations. It worth mentioning that the training schedule was the same as the networks reported in the manuscript.

This work is supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) RGPIN-2020-04612.

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

TABLE SII: The AUC and 95% confidence interval when different features are used to train the Random forest classifier

Features	Simulation AUC	Phantom AUC
R, S	0.886 (0.858- 0.915)	0.887 (0.869- 0.905)
R, S, entropy	0.883 (0.854- 0.911)	0.885 (0.866- 0.905)
R, S, T	0.890 (0.863- 0.916)	0.893 (0.876- 0.910)
R, S, entropy, T	**0.894** (0.868- 0.920)	**0.895** (0.880- 0.913)

E. Results Using Different Combinations of Features

The simulation and experimental phantom results of the MLP classifier using different combinations of features are given in Table SII. According to the Table, the best combination is to include all four features (R, S, $entropy$ and T) which confirms that the features we used were not redundant and including all of them were beneficial.

F. Simulation Imaging Parameter

We simulated 200 phantoms with the imaging setting given in Table SIII.

TABLE SIII: Imaging parameters used in simulation dataset.

Parameter	Value
Center Frequency	6.6 MHz
Sampling Frequency	100 MHz
Total Number of Elements	192
Active Number of Elements	64
Element Height	10 mm
Kerf	0.1 mm
Phantom Width	30 mm
Phantom Height	30 mm
Phantom Start Height	30 mm
Focal Point	45 mm
Number of Lines	128
Decimation Scale in Axial	2
Interpolation in Lateral	3
Axial Resolution Size	0.35 mm
Lateral Resolution Size	0.48 mm

G. Evaluation of results

The AUCs versus methods (CNN only, CNN+fusion and CNN+DS) and different networks are shown in the Fig. S4. F1 score of simulation and phantom results are given in Table SV. The ROC curve of experimental phantoms for different settings of DenseNet121 is depicted in Fig. S5.

H. B-mode images of Experimental Phantoms

The B-mode images of the experimental phantoms are illustrated in Fig. S6.

I. Training Results

Simulation and experimental test results are presented in the paper. Training results of different variants of DenseNet121 are given in Table SVI. The summary of the training is given in Table SVII.

J. Learning Curves

The training and validation learning curves of the networks used in this paper are depicted in Fig. S3. The evaluated loss for both training and validation is binary cross entropy. The fluctuation in the training loss is due to using cyclic learning rate. The number of epochs for the networks is not fixed and early stopping is used to terminate the training. It can be observed that DenseNet121 has the lowest training and validation loss among the compared methods and also it requires the least number of epochs to converge.

K. Visualizing More Results

The results of some other networks which are not presented in the manuscript are given in Fig. S7.

TABLE SIV: AUCs of training the models end to end 5 times having random initialization.

	1st		2nd		3rd		4th		5th	
	Sim	Exp	Sim	Exp	Sim	Exp	Sim	Exp	Sim	Exp
CNN with fusion	0.847	0.851	0.846	0.890	0.859	**0.548**	0.825	0.899	0.922	0.814

TABLE SV: F1 score of simulation and phantom results

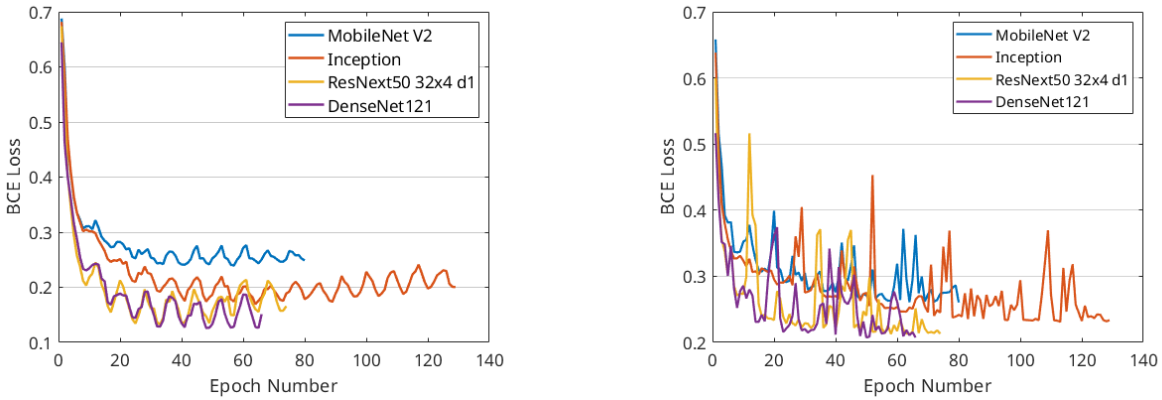
Model	Fusion	DS	F1 (simulation)	F1 (phantom)
SVM	✗	✗	0.808	0.370
Random Forest	✗	✗	0.811	0.528
MLP	✗	✗	0.818	0.393
MobileNet V2	✗	✗	0.866	0.714
MobileNet V2	✓	✗	0.830	0.561
MobileNet V2	✗	✓	0.866	0.687
Inception	✗	✗	0.898	0.794
Inception	✓	✗	0.909	0.762
Inception	✗	✓	0.885	0.756
ResNext50_32x4d1	✗	✗	**0.923**	0.722
ResNext50_32x4d1	✓	✗	0.835	0.592
ResNext50_32x4d1	✗	✓	0.914	0.699
DenseNet121	✗	✗	0.890	0.770
DenseNet121	✓	✗	0.880	**0.818**
DenseNet121	✗	✓	0.886	0.816

TABLE SVI: Training Results

Model	Fusion	DS	AUC	Sensitivity	Precision	Accuracy	Youden's Index
DenseNet121	✗	✗	0.992 (0.991-0.994)	0.939	0.962	0.956	0.921 (0.35)
DenseNet121	✓	✗	0.972 (0.968-0.975)	0.960	0.826	0.892	0.820 (0.67)
DenseNet121	✗	✓	0.990 (0.988-0.992)	0.908	0.965	0.944	0.911 (0.33)

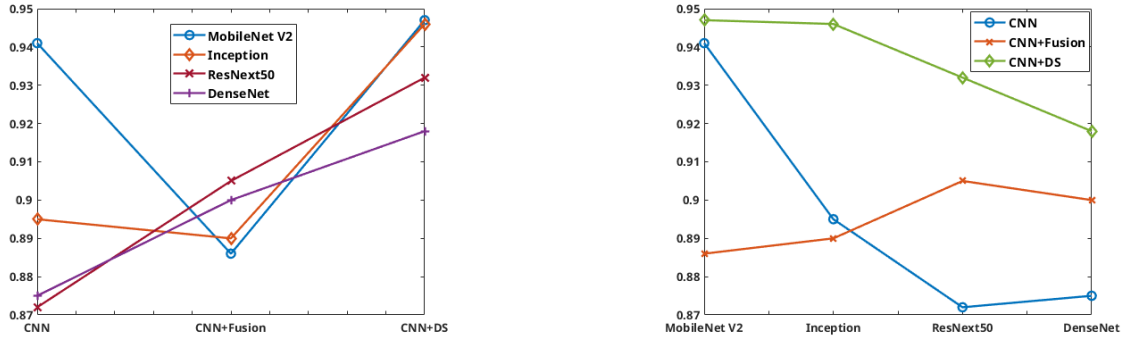
TABLE SVII: Summary of the training hyper parameters

Parameter	Value
Number of simulation training data	5000
Number of simulation validation data	1000
Number of simulation test data	500
Number of patches for each expermental phantom	318
Patch statistics used	SNR, skewness, entropy and T
CNN models evaluated	MobileNet V2, Inception, ResNext32, DenseNet121
Methods of combining patch statistics	Fusion using MLP and Multi-task learning (deep supervision)
Stopping criteria	Early stopping
Learning rate	Cyclic learning rate (1e-8 - 1e-5)
Input Channels	Envelope and Envelope multiplied by log compressed Envelope



(a) Training (b) Validation

Fig. S3: The learning curves of the networks without using the statistical features.



(a) AUCs vs methods. (b) AUCs vs networks

Fig. S4: Experimental phantoms AUCs versus methods of using patch statistics, CNN, CNN+Fusion and CNN+DS (a). Experimental phantoms AUCs versus different network architectures compared in the paper (b).

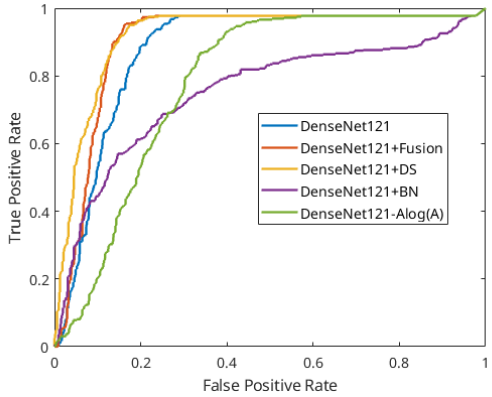
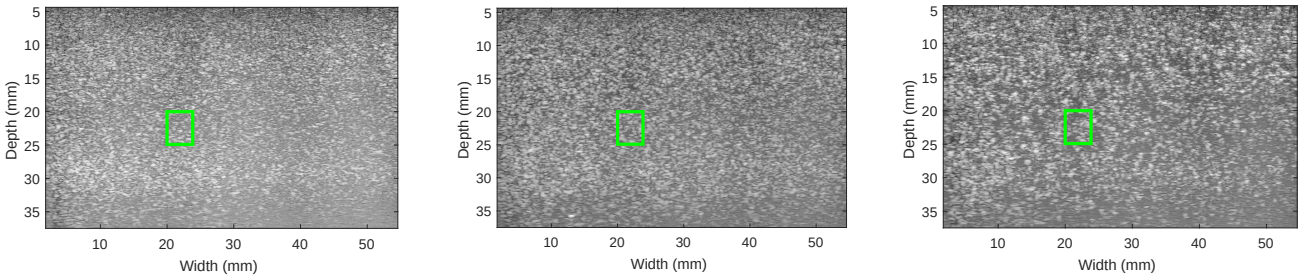


Fig. S5: ROC curve of different settings of DenseNet.



(a) Phantom A (High) (b) Phantom B (Medium) (c) Phantom C (Low)

Fig. S6: B-mode images of the experimental phantoms. The patch size is specified by the highlighted windows.

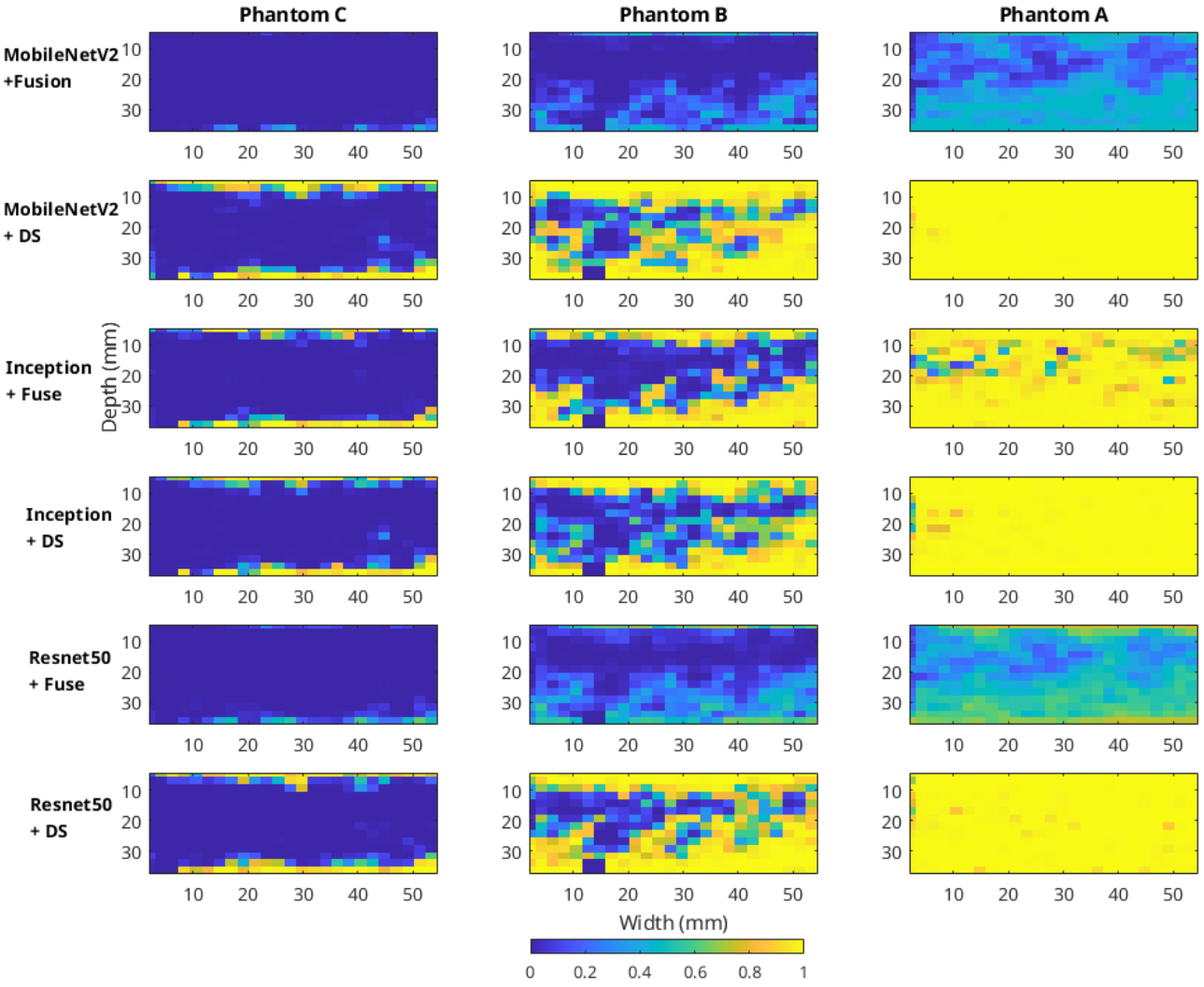


Fig. S7: The results of MLP, MobileNet V2, Inception and ResNext50 models with fusion or deep supervision (DS) on the experimental phantoms. The color code represents the predicted output of the networks, from 0 (UDS) to 1 (FDS). Correct classes are 0 (UDS) for phantoms C and B, and 1 for phantom A.