



HAL
open science

3D Complex-valued Convolutional Network for Fast Ultrasound Diverging Wave Imaging

Ahmed Bentaleb, Christophe Sintès, Chafiaa Hamitouche

► **To cite this version:**

Ahmed Bentaleb, Christophe Sintès, Chafiaa Hamitouche. 3D Complex-valued Convolutional Network for Fast Ultrasound Diverging Wave Imaging. RFIAP 2024, Jul 2024, Lille, France. hal-04617986

HAL Id: hal-04617986

<https://hal.science/hal-04617986>

Submitted on 20 Jun 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

3D Complex-valued Convolutional Network for Fast Ultrasound Diverging Wave Imaging

Ahmed Bentaleb

Christophe Sintès

Chafiaa Hamitouche

Département Image et Traitement de l'Information, Institut Mines-Télécom (IMT) Atlantique,
29200 Brest, France

ahmed.bentaleb@imt-atlantique.fr

Abstract

Diverging wave acquisition has become the standard technique for fast ultrasound imaging due to its high temporal resolution. However, this approach is limited by the number of transmitted waves. Recently, deep learning has taken place in a variety of studies to improve the quality of ultrasound imaging. Most of these approaches have been performed on Bmode images, radio frequency signals, and in-phase/quadrature posterior to the delay-and-sum beamformer. In this work, we employ a 3D complex convolutional neural network to reconstruct an enhanced ultrasound image from fewer diverging waves prior to the sum in delay-and-sum beamformer. The network was trained using simulated data and evaluated using in-vitro phantom data. We provide experimental evidence that our approach produces a high contrast and quality image from 3 steered acquisitions that is comparable to those obtained from diverging wave compounding.

Keywords

Complex convolutional neural networks, deep learning, image reconstruction, in-phase/quadrature signal, ultrasound imaging.

1 Introduction

Ultrasound (US) image reconstruction involves the reception of the reflected multichannel raw radio frequency (RF) signals. RF data is demodulated to produce in-phase/quadrature (IQ) data, followed by beamforming of the imaged target. Beamforming is the application of a time-of-flight correction process with spatial filtering to introduce selectivity into the signal to eliminate undesirable interference [1].

Diverging wave (DW) ability to image a target with a single transmission acquisition makes it achieve higher frame rate [2]. However images from a single DW emission produces poor quality. A standard technique to improve the quality consists of coherently compounding multiple acquisitions with different steering angles [3] [4] [5]. A trade-off be-

tween quality and frame rate needs to be made as the transmission of multiple waves increases the acquisition time, hence lower frame rate.

Recently, deep learning has played a key role in medical data processing due to the growth of computational power and various use cases. State-of-the-art results have been achieved in image classification [6], segmentation [7] [8], liver and breast lesion classification [9] [10]. This prompted US medical researchers to apply deep learning methods to their research. Deep neural networks (DNNs) were used to interpolate missing RF data [11], and the reconstruction of B-mode images [12]. A more promising approach is the use of convolutional neural networks (CNNs), which have shown improvement in image processing-related tasks; for instance CNN was used for compounded imaging [13], denoising and speckle reduction [14], also the use of a fully convolutional neural network (FCNN), which learned an MV beamformer transformation [15] and direct image segmentation from RF data [16]. Another use of deep learning was the learning of compounding for DW imaging [17] [18] where a real valued CNN was used to reconstruct a compounded image from fewer DWs.

Until recently, deep learning has predominantly focused on real-valued data, with limited exploration of complex-valued data. A recent study [19] compared the effectiveness of complex-value neural networks (CVNNs) versus real-value neural networks (RVNNs) to assess the utility of such data. Additionally, the application of CVNNs has been investigated in tasks involving the learning of complex representations of time-series and processing of seismic data [20].

In the field of US imaging, CNNs based on complex IQ data has been introduced for speckle reduction [14], multi-line acquisition and transmission improvement [21], where the complex data was trained as a separated real and imaginary parts in a two-branch network structure as in color channel information, such approach does not take in consideration the nature of complex data and its calculation. As demonstrated in [19], a complex-valued model provides

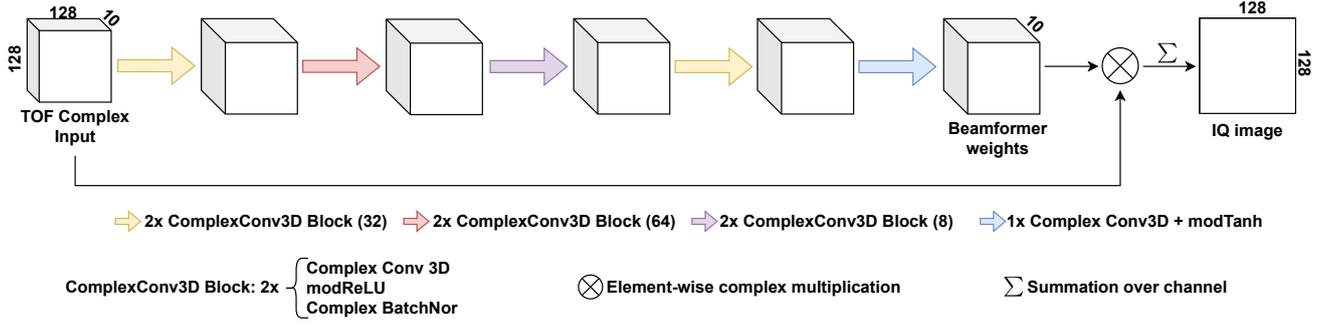


Figure 1: Proposed complex-valued architecture. The network consists of five (5) complex convolution 3D blocks, where each block includes doubled complex convolution 3d, modReLU activation and complex batch normalisation. Finally single 3D complex convolution with modTanh activation to produce the network weights. The weights are applied to the input data and the result is summed over channel data to produce complex IQ image.

a more constrained system than a model based on real-valued parameters. A recent study [22] demonstrated the superior representational capacity of complex-valued neural networks in acoustic applications such as speech spectrum prediction and music transcription.

A different approach was considered by [23] [24], where a CVNN was defined for DW imaging; their work consisted of building a complex equivalent of the work in [18]. The network consisted of a 2D convolution between the IQ data produces from delay-and-sum (DAS) beamformer as real and imaginary parts with complex weights represented as real and imaginary parts.

In this context, we propose a 3D CVNN architecture for fast DW imaging using only 3 DW acquisitions. Our network employees data after time-of-flight (TOF) correction that produces 3D complex channel signals rather than the 2D IQ signals posterior to DAS. Our network uses 3D complex convolution to incorporate the phase information from each channel to enhance the learning to produce high quality US image.

The present study is structured as follows: Section 2 deals with the proposed architecture and training strategy. Section 3 presents the experimental setup for the data acquisition, network training, and performance metrics. Section 4 presents the result of the work and a comparison with the approach presented in [24] which is named CID-Net. Finally, conclusion and remarks are given in section 5.

2 Methods

2.1 Delay and Sum

Assume an M element linear array aligned on the x -axis facing the positive direction of the z -axis, and a pixel grid of R rows and C columns with p_{rc} being one pixel from the grid such that $r \in [1, R]$ and $c \in [1, C]$. Denoting $IQ_m(t)$ the demodulated RF echo signal received at an element m , and τ_{rcm} the round trip time of the pixel p_{rc} to the element m of the transmitted wave.

DAS beamformer performs a time of flight correction by

combining the IQ signals at a delay τ_{rcm} to generate complex-valued pixel data that is out of phase producing a frame of size $R \times C \times M$. In order to correct the pixel data, a correction weights w_{DAS} are applied to the pixel data and summed over the transducer's elements as follows:

$$I_{DAS} = \sum_{m=1}^M w_{DAS} \cdot IQ_m(\tau_{rcm}). \quad (1)$$

The correction weights are defined as:

$$w_{DAS} = \exp(2\pi j f_c \tau_{rcm}), \quad (2)$$

where $j = \sqrt{-1}$ is the imaginary unit and f_c is the transducer center frequency. V. Perrot *et al.* [25] implemented a DAS beamformer as a part of their toolbox that handles ultrasound imaging which will be used for data processing in our work.

2.2 Complex Convolution

In this work, complex-valued convolution layer is utilized based on the definition of complex-valued convolution [22]. Assuming a complex-valued input X and a kernel W , the complex-valued convolution Z is defined as:

$$\begin{bmatrix} \Re(Z) \\ \Im(Z) \end{bmatrix} = \begin{bmatrix} \Re(W) & -\Im(W) \\ \Im(W) & \Re(W) \end{bmatrix} * \begin{bmatrix} \Re(X) \\ \Im(X) \end{bmatrix}, \quad (3)$$

where $\Re(\cdot)$ and $\Im(\cdot)$ denote the real and imaginary part of the complex value, respectively. $(*)$ denotes convolutional operation.

2.3 Activation Function

Different activation were investigated in the context of complex-valued networks, however, our aim is to preserved the phase information and only influence that value with the complex convolution operation. Arjovsky *et al.* proposed modReLU [26], which is a variant of ReLU activation that is defined as:

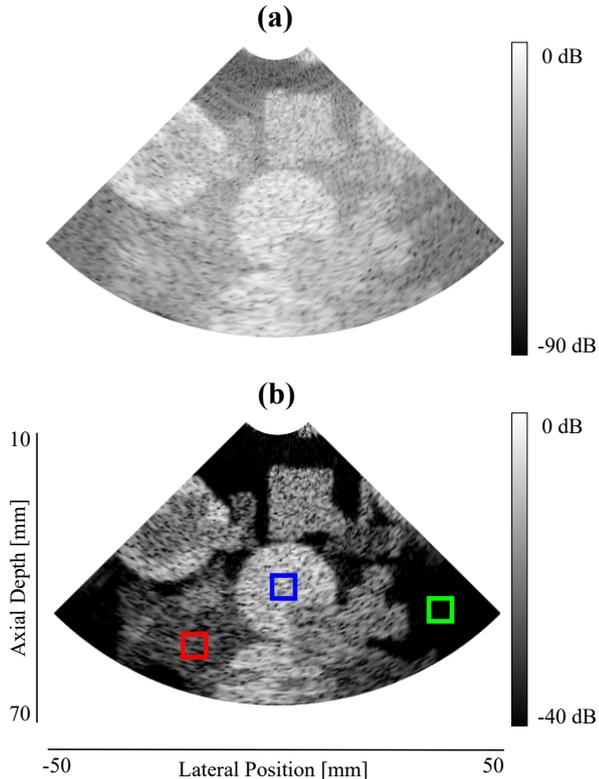


Figure 2: Sample from training set. Red area indicates background region, green area indicates anechoic region, and blue area indicates hyperechoic region. (a) B-mode image of standard compounding 3 DWs as training input with 90 dB, (b) B-mode image of standard compounding 20 DWs as target with 40 dB.

$$\text{modReLU}(Z) = \text{ReLU}(|Z| + b)e^{j\theta_Z}, \quad (4)$$

where θ_Z is the phase of Z , $b \in \mathbf{R}$ is a bias parameter of the non-linearity, and $|\cdot|$ is the magnitude of the complex number. By applying such activation the phase θ_Z is preserved in contrary to applying a separate ReLUs on both the real and imaginary parts. Following the definition of modeReLU, we created modTanh as a final activation as follow:

$$\text{modTanh}(Z) = \tanh(|Z| + b)e^{j\theta_Z}. \quad (5)$$

2.4 Loss Function

Network learning is not complete without applying back-propagation to update the complex kernels. Back-propagation implies differentiable loss function. In a complex-valued network a complex differentiable function is needed (i.e., holomorphic functions that satisfy the Cauchy-Riemann conditions) [20]. However, this restriction is not necessary, as it was shown that a real differentiable loss function with respect to the real and imaginary parts is compatible with back-propagation[19, 22].

We suggest to use a complex magnitude mean-squared-error (MSE) between the desired output image Y and network prediction \hat{Y} :

$$\mathcal{L}_{MSE} = \frac{1}{N} \sum (\hat{Y} - Y)^2. \quad (6)$$

2.5 Network Architecture

The main building components of the network are: 3D complex convolution, modReLU, and complex batch normalization [27] with a final modTanh activation. The network consist of a series of blocks that learns the optimal weights to produce a high quality and contrast. Overall, the network is defined as:

$$I_{new} = \sum_{m=1}^M w_{new} \cdot w_{DAS} \cdot IQ_m(\tau_{rcm}), \quad (7)$$

where w_{new} is the learned weights of the network applied to DAS and summed over channel axis.

The proposed network (3D-CVCNN) presented in Figure 1; consists of 5 complex convolution blocks. One block is constructed using the previously mentioned components. The network uses proper padding to preserve the spatial features with the same size. modReLU preserves the phase throughout the network to reconstruct w_{new} form the target data. w_{new} is multiplied to the input and the result is summed over channel axis to produce complex IQ image.

Table 1: Ultrasound Imaging Configuration.

Parameters	Value
Number of elements	80
Center frequency	2.8 MHz
Sampling frequency	11.2 MHz
Pitch	240 μm
Kerf	40 μm
Fractional bandwidth	75%
Speed of sound	1540 m/s

3 Experiment

3.1 Data Acquisition

A phased array configuration derived from VERMON P2-8 transducer (Table 1) was used with SIMUS [28, 29] to simulate training dataset by steered diverging wave acquisitions. Each acquisition was performed by emitting 20 DWs with tilt angles between $\pm 45^\circ$. The simulated RF data were demodulated using a phaser at the center frequency, and a fifth-order Butterworth low-pass filter was used for bandwidth reduction. Time-of-flight correction was performed with a modified version of the DAS beamformer [25] to

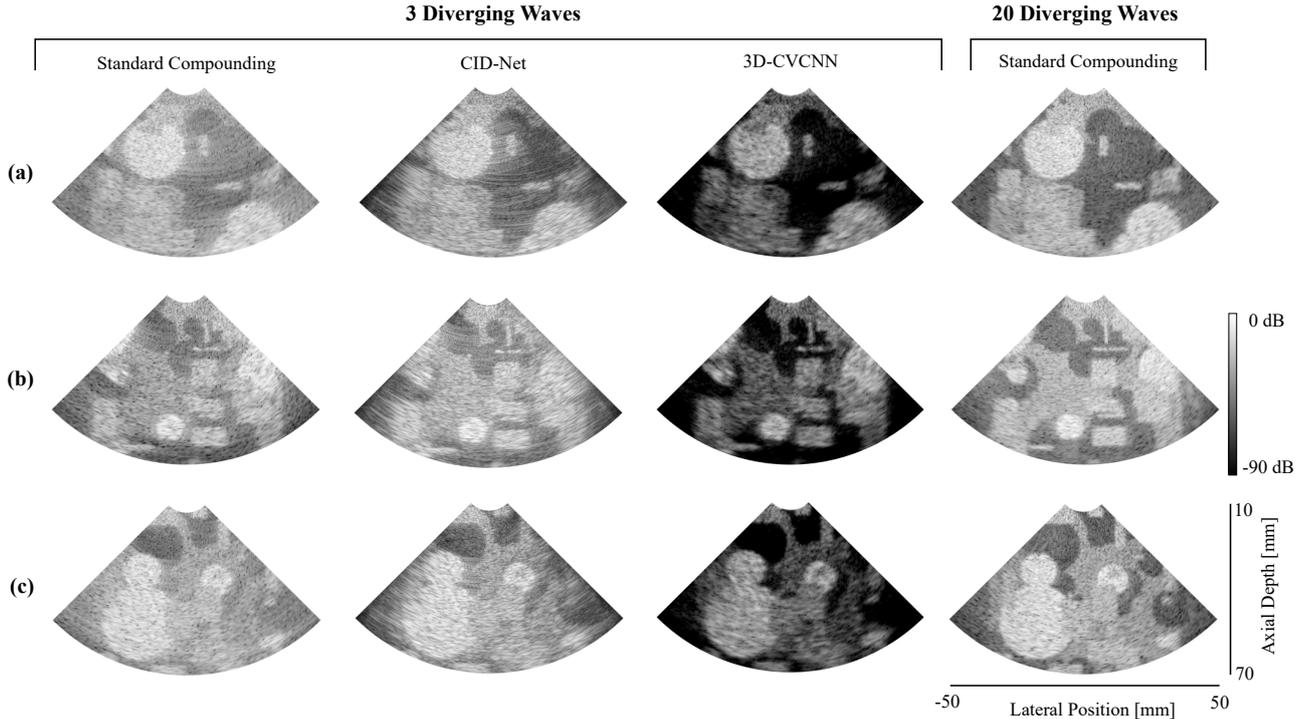


Figure 3: B-mode samples from simulated test set comparing different methods. Left to right: standard compounding of 3 DWs; CID-Net (3 DWs); 3D-CVCNN (3 DWs); and standard compounding of 20 DWs.

Method	Hyperechoic region		Anechoic region	
	CR	CNR	CR	CNR
Standard compounding (3 DWs)	0.20	2.31	0.18	2.00
CID-Net (3 DWs)	0.20	2.58	0.23	3.01
3D-CVCNN (3 DWs)	0.39	3.86	0.21	2.53
Standard compounding (20 DWs)	0.20	2.16	0.22	2.34

Table 2: CR, CNR of different methods on the test set.

create the 3D complex input data that contains IQ signals for each element of the transducer (channel), using a mean sound velocity of 1540 m/s. The IQ signals were formed on a 128×128 grid, which corresponds to a physical dimensions of 70 mm in length, and a sectorial angle of 90° .

From each acquisition, target data was obtained by coherently compounding all 20 TOF signal and summed over channel axis to produce a 128×128 IQ image, while a subset of 3 DWs TOF data that correspond to steered angles $[-30^\circ, 0^\circ, +30^\circ]$ were summed to produce the complex training TOF data $128 \times 128 \times 80$. To be able to fit the training data and prevent the loss of information (phase), we averaged the input data over the channel axis to a size of $128 \times 128 \times 10$. A total of 5000 training pairs (TOF, IQ) were simulated.

In addition, we acquired in-vitro data using CIRS phan-

tom model 040GS, with VERMON P2-8 transducer on a Vantage 256; Verasonics scanner. The acquisition was performed with 31 DWs with title angles between $\pm 45^\circ$ and processed with the same procedure as the simulated data. The in-vitro data were not included in training, and only used for evaluation.

3.2 Network Training

The network was implemented in Python using Keras API [30], and Keras Complex API [27]. A selection of 3500 pairs (TOF, IQ) were used as training, 1000 for validation, and the remaining 500 for testing. Due to the size of training data and GPU RAM limitations, a data loader was used to feed the training data to the network by batch. The network was trained using Adam optimizer [31] with a batch size of 1 and an initial learning rate of 1×10^{-4} . Early stopping mechanism was implemented to prevent over-fitting,

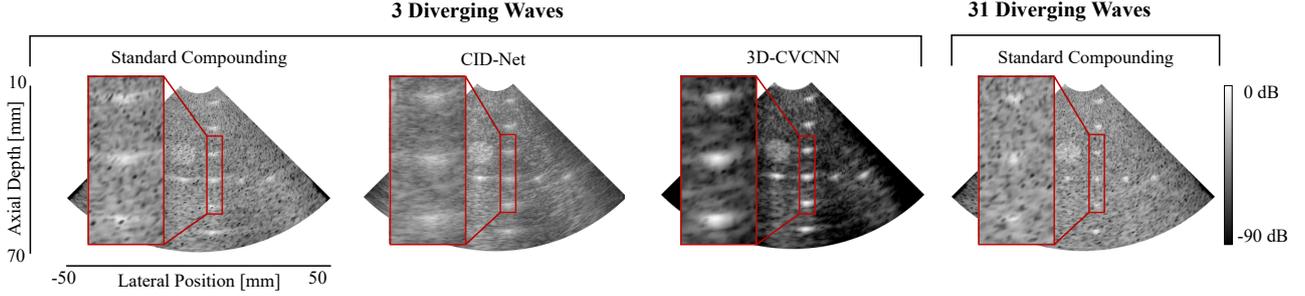


Figure 4: B-mode samples from in-vitro phantom (CIRS; model 040GSE) comparing different methods. Left to right: standard compounding of 3 DWs; CID-Net (3 DWs); 3D-CVCNN (3 DWs); and standard compounding of 31 DWs.

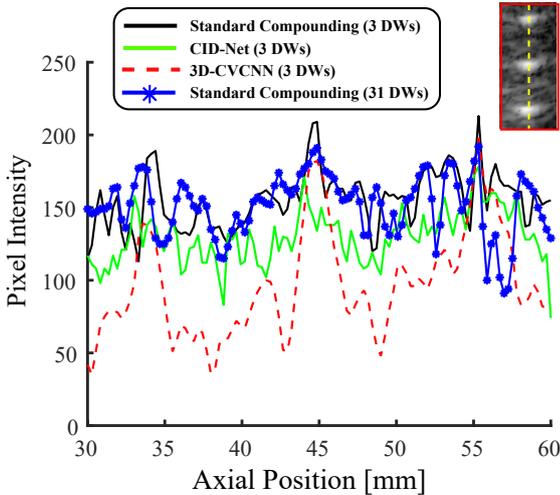


Figure 5: Axial profiles of the point reflectors represented in Figure 4 (a).

with learning rate scheduler. The learning rate was halved if there was no reduction in validation loss for 10 epochs, and training was terminated if there were 20 epochs without a reduction in validation loss with a minimum change of 0.0001 that qualify as an improvement. The training was conducted on a NVIDIA GeForce RTX 3080 GPU.

3.3 Evaluation Metrics

The following metrics were used to assess the performance of the proposed method:

Contrast: Contrast (CR) is defined as the ratio of the mean value in the region of interest to the mean value in the region of the background [32].

$$CR = \frac{|\mu_R - \mu_B|}{\sigma_B}. \quad (8)$$

Contrast to Noise Ratio: Contrast to Noise Ratio (CNR) measures the signal intensity ratio between the region of interest and the background [33].

$$CNR = \frac{|\mu_R - \mu_B|}{\sqrt{\sigma_R^2 + \sigma_B^2}}, \quad (9)$$

where μ_R, μ_B (σ_R, σ_B) are the mean (standard deviation) of the region of interest and the background, respectively.

4 Results

4.1 Learning Convergence

After 56 epochs the training was ended according to the implemented early stopping mechanism and relative to the size of the model that has a total of 4.9×10^3 trainable parameters. Training took approximately 15 hours and the weights obtained from the last training session were used for further analysis.

4.2 Image Quality

Figure 3 depicts B-mode sample from test set comparing different methods. The methods included standard compounding, CID-Net, and 3D-CVCNN with 3 DWs, in addition to standard compounding with 20 DWs. We observe that 3D-CVCNN improved the overall image quality and contrast by suppressing background noise. This improvement is related to the learning strategy where we fed the network with high contrast data that is reconstructed from coherently compounding 20 DW acquisition.

The previously defined metrics in section 3.3 were computed over the test set. The results are presented in Table 2. We notice that 3D-CVCNN improves the contrast and that is reflected into the values of CR and CNR compared to standard compounding and CID-Net, However CID-Net provides higher CNR in anechoic regions.

4.3 Image resolution

Figure 4 depicts results from the in-vitro phantom acquisition comparing different methods. As 3D-CVCNN improves the contrast it lacks the sharpness of standard compounding of 31 DWs. Another observation is presented in Figure 5 showing the axial profile of the point reflectors (in red rectangle). We notice that 3D-CVCNN provides a separation between the point reflectors and background compared to standard compound and CID-Net with 3 DWs, However as the input data is from 3 DWs, the blurring in point reflectors is to be expected and that is to be improved during acquisition.

5 Conclusion

In this paper, we presented a 3D complex-valued convolutional network for fast DW imaging. The proposed method was trained for the reconstruction of high quality imaging and frame rate only using 3 DWs. Through experimental evaluation, we demonstrated the effectiveness of this method, yielding an image quality comparable to that obtained from coherent compound of more than 20 DWs. For future work, we aim to explore methods to improve the pixel resolution and sharpness of the reconstructed IQ image.

References

- [1] H Trees. *Optimum array processing (detection, estimation, and modulation theory, part iv)* wiley. 2002.
- [2] Hideyuki Hasegawa and Hiroshi Kanai. “High-frame-rate echocardiography using diverging transmit beams and parallel receive beamforming”. In: *Journal of medical ultrasonics* 38 (2011), pp. 129–140.
- [3] Gabriel Montaldo et al. “Coherent plane-wave compounding for very high frame rate ultrasonography and transient elastography”. In: *IEEE transactions on ultrasonics, ferroelectrics, and frequency control* 56.3 (2009), pp. 489–506.
- [4] Mickael Tanter and Mathias Fink. “Ultrafast imaging in biomedical ultrasound”. In: *IEEE transactions on ultrasonics, ferroelectrics, and frequency control* 61.1 (2014), pp. 102–119.
- [5] Jonathan Porée et al. “High-frame-rate echocardiography using coherent compounding with Doppler-based motion-compensation”. In: *IEEE transactions on medical imaging* 35.7 (2016), pp. 1647–1657.
- [6] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. “ImageNet Classification with Deep Convolutional Neural Networks”. In: *Advances in Neural Information Processing Systems*. Ed. by F. Pereira et al. Vol. 25. Curran Associates, Inc., 2012.
- [7] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. “U-net: Convolutional networks for biomedical image segmentation”. In: *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III* 18. Springer. 2015, pp. 234–241.
- [8] Xiaomeng Li et al. “H-DenseUNet: Hybrid Densely Connected UNet for Liver and Tumor Segmentation From CT Volumes”. In: *IEEE Transactions on Medical Imaging* 37.12 (2018), pp. 2663–2674. DOI: 10.1109/TMI.2018.2845918.
- [9] Kaizhi Wu, Xi Chen, and Mingyue Ding. “Deep learning based classification of focal liver lesions with contrast-enhanced ultrasound”. In: *Optik - International Journal for Light and Electron Optics* 125 (Aug. 2014). DOI: 10.1016/j.ijleo.2014.01.114.
- [10] Jie-Zhi Cheng et al. “Computer-aided diagnosis with deep learning architecture: applications to breast lesions in US images and pulmonary nodules in CT scans”. In: *Scientific reports* 6.1 (2016), pp. 1–13.
- [11] Yeo Hun Yoon and Jong Chul Ye. “Deep Learning for Accelerated Ultrasound Imaging”. In: *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2018, pp. 6673–6676. DOI: 10.1109/ICASSP.2018.8462304.
- [12] Ben Luijten et al. “Deep Learning for Fast Adaptive Beamforming”. In: *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2019, pp. 1333–1337. DOI: 10.1109/ICASSP.2019.8683478.
- [13] Maxime Gasse et al. “High-Quality Plane Wave Compounding Using Convolutional Neural Networks”. In: *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* 64.10 (2017), pp. 1637–1639. DOI: 10.1109/TUFFC.2017.2736890.
- [14] Dongwoon Hyun et al. “Beamforming and Speckle Reduction Using Neural Networks”. In: *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* 66.5 (2019), pp. 898–910. DOI: 10.1109/TUFFC.2019.2903795.
- [15] Walter Simson et al. “Deep Learning Beamforming for Sub-Sampled Ultrasound Data”. In: *2018 IEEE International Ultrasonics Symposium (IUS)*. 2018, pp. 1–4. DOI: 10.1109/ULTSYM.2018.8579818.
- [16] Arun Asokan Nair et al. “A Deep Learning Based Alternative to Beamforming Ultrasound Images”. In: *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2018, pp. 3359–3363. DOI: 10.1109/ICASSP.2018.8461575.
- [17] Jingfeng Lu et al. “Fast diverging wave imaging using deep-learning-based compounding”. In: *2019 IEEE International Ultrasonics Symposium (IUS)*. IEEE. 2019, pp. 2341–2344.
- [18] Jingfeng Lu et al. “Reconstruction for diverging-wave imaging using deep convolutional neural networks”. In: *IEEE transactions on ultrasonics, ferroelectrics, and frequency control* 67.12 (2020), pp. 2481–2492.

- [19] Akira Hirose and Shotaro Yoshida. “Generalization Characteristics of Complex-Valued Feedforward Neural Networks in Relation to Signal Coherence”. In: *IEEE Transactions on Neural Networks and Learning Systems* 23.4 (2012), pp. 541–551. DOI: 10.1109/TNNLS.2012.2183613.
- [20] Andy M Sarroff, Victor Shepardson, and Michael A Casey. “Learning representations using complex-valued nets”. In: *arXiv preprint arXiv:1511.06351* (2015).
- [21] Ortal Senouf et al. “High frame-rate cardiac ultrasound imaging with deep learning”. In: *Medical Image Computing and Computer Assisted Intervention—MICCAI 2018: 21st International Conference, Granada, Spain, September 16–20, 2018, Proceedings, Part I*. Springer, 2018, pp. 126–134.
- [22] Chiheb Trabelsi et al. “Deep complex networks”. In: *arXiv preprint arXiv:1705.09792* (2017).
- [23] Jingfeng Lu et al. “Complex convolutional neural networks for fast diverging wave imaging”. In: *2020 IEEE International Ultrasonics Symposium (IUS)*. IEEE, 2020, pp. 1–3.
- [24] Jingfeng Lu et al. “Complex convolutional neural networks for ultrafast ultrasound imaging reconstruction from in-phase/quadrature signal”. In: *IEEE transactions on ultrasonics, ferroelectrics, and frequency control* 69.2 (2021), pp. 592–603.
- [25] Vincent Perrot et al. “So you think you can DAS? A viewpoint on delay-and-sum beamforming”. In: *Ultrasonics* 111 (2021), p. 106309.
- [26] Martin Arjovsky, Amar Shah, and Yoshua Bengio. “Unitary evolution recurrent neural networks”. In: *International conference on machine learning*. PMLR, 2016, pp. 1120–1128.
- [27] Jesper Sören Dramsch and Contributors. *Complex-Valued Neural Networks in Keras with Tensorflow*. 2019. DOI: 10.6084/m9.figshare.9783773. URL: https://figshare.com/articles/Complex_Valued_Neural_Networks_in_Keras_with_Tensorflow/9783773/1.
- [28] Amanda Cigier, François Varray, and Damien Garcia. “SIMUS: An open-source simulator for medical ultrasound imaging. Part II: Comparison with four simulators”. In: *Computer Methods and Programs in Biomedicine* 220 (2022), p. 106774.
- [29] Damien Garcia. “SIMUS: an open-source simulator for medical ultrasound imaging. Part I: theory & examples”. In: *Computer Methods and Programs in Biomedicine* 218 (2022), p. 106726.
- [30] François Chollet et al. *Keras*. <https://keras.io>. 2015.
- [31] Diederik P Kingma and Jimmy Ba. “Adam: A method for stochastic optimization”. In: *arXiv preprint arXiv:1412.6980* (2014).
- [32] Mengling Xu et al. “Spatio-temporally smoothed coherence factor for ultrasound imaging [correspondence]”. In: *IEEE transactions on ultrasonics, ferroelectrics, and frequency control* 61.1 (2014), pp. 182–190.
- [33] Ole Marius Hoel Rindal et al. “Understanding contrast improvements from capon beamforming”. In: *2014 IEEE International Ultrasonics Symposium*. 2014, pp. 1694–1697. DOI: 10.1109/ULTSYM.2014.0420.