Minimizing Image Quality Loss After Channel Count Reduction for Plane Wave Ultrasound via Deep Learning Inference

Di Xiao, Graduate Student Member, IEEE, William M. K. Pitman, Graduate Student Member, IEEE, Billy Y. S. Yu, Member, IEEE, Adrian J. Y. Chee, Member, IEEE, and Alfred C. H. Yu, Senior Member, IEEE

Abstract—High-frame-rate ultrasound imaging uses unfocused transmissions to insonify an entire imaging view for each transmit event, thereby enabling frame rates over 1000 frames per second (fps). At these high frame rates, it is naturally challenging to realize real-time transfer of channel-domain raw data from the transducer to the system back end. Our work seeks to halve the total data transfer rate by uniformly decimating the receive channel count by 50% and, in turn, doubling the array pitch. We show that despite the reduced channel count and the inevitable use of a sparse array aperture, the resulting beamformed image quality can be maintained by designing a custom convolutional encoder–decoder neural network to infer the radio frequency (RF) data of the nullified channels. This deep learning framework was trained with in vivo human carotid data (5-MHz plane wave imaging, 128 channels, 31 steering angles over a 30° span, and 62 799 frames in total). After training, the network was tested on an in vitro point target scenario that was dissimilar to the training data, in addition to in vivo carotid validation datasets. In the point target phantom image beamformed from inferred channel data, spatial aliasing artifacts attributed to array pitch doubling were found to be reduced by up to 10 dB. For carotid imaging, our proposed approach yielded a lumen-to-tissue contrast that was on average within 3 dB compared to the full-aperture image, whereas without channel data inferencing, the carotid lumen was obscured. When implemented on an RTX-2080 GPU, the inference time to apply the trained network was 4 ms, which favors real-time imaging. Overall, our technique shows that with the help of deep learning, channel data transfer rates can be effectively halved with limited impact on the resulting image quality.

Index Terms—Convolutional neural network (CNN), high-frame-rate ultrasound, plane wave imaging, sparse arrays.

Manuscript received 20 April 2022; accepted 3 July 2022. Date of publication 21 July 2022; date of current version 27 September 2022. This work was supported in part by the Natural Sciences and Engineering Council of Canada under Grant CREATE-528202-2019 and Grant RGPIN-2022-04157, in part by the Canadian Space Agency under Grant 21FAWA TA08, and in part by the E. W. R. Steacie Memorial Fellowship under Grant SMFSU-556263-2021. (Corresponding author: Alfred C. H. Yu.)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Clinical Research Ethics Committee of the University of Waterloo under Approval No. 31694.

The authors are with the Schlegel Research Institute for Aging and the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: alfred.yu@uwaterloo.ca).

Digital Object Identifier 10.1109/TUFFC.2022.3192854

I. INTRODUCTION

ULTRASOUND imaging is well regarded as a noninvasive medical imaging modality. For current clinical scanners with array scanning capabilities, they typically form each image frame via a beamline-based pulse-echo sensing paradigm that involves sending one focused pulse firing along each of a set of beamlines in the imaging view. From each pulse-echo sensing event, one axial line of the image frame is generated by performing delay-and-sum (DAS) beamforming based on the channel-domain radio frequency (RF) data samples acquired from all array elements [1]. Real-time frame rates can be achieved with this conventional ultrasound imaging paradigm. In contrast, in the past decade, unfocused pulsing of planar or diverging wavefronts, which can transiently insonify the entire imaging view, has demonstrated efficacy in achieving high frame rates of over 1000 frames per second (fps) [2], [3]. For this high-frame-rate imaging paradigm, one image frame can be generated from the channel RF data array of the corresponding unfocused pulse-echo firing. The beamformed images obtained from unfocused pulse-echo firings with different steering angles can also be coherently compounded to improve the lateral image resolution [4], [5]. Accordingly, researchers have successfully imaged hemodynamics in vivo on a time-resolved basis with submillisecond temporal resolution [6]–[8]. Others have also developed new applications, such as shear wave elastography and functional ultrasound imaging [9], [10].

While conceptually powerful, high-frame-rate ultrasound imaging is not without its practical implementation challenges. Its two primary hurdles are: 1) high computational capacity is required to generate images in real time [11] and 2) high streaming bandwidth is needed in transferring the channel RF data from the array front end to the computing back end [2]. The first hurdle can possibly be overcome by applying software beamforming methods that make use of graphics processing unit (GPU) computing devices to parallelize the DAS beamforming process over different pixel positions [4], [12]. Nevertheless, the second hurdle is more challenging to address. Specifically, upward of 10 GB of channel RF data can be transmitted every second, and few scanners are designed to accommodate this level of data streaming traffic [2]. In light of
this bottleneck, it is of practical interest to reduce the amount of channel RF data that needs to be streamed to the computing back end.

Several approaches have attempted to circumvent the data streaming bottleneck in high-frame-rate ultrasound imaging. One emerging direction in recent years is compressed sensing (CS), which combines sub-Nyquist sampling with assumptions of sparsity to achieve close to 90% data reduction [13]–[16]. However, these methods often require an iterative solution to optimization problems (i.e., there is no closed-form solution), so it is inherently challenging to implement them in real time even though attempts to accelerate CS have been made [17], [18]. Another algorithmic solution is to perform sparse convolutional beamforming that works by using a sum coarray algorithm to improve image quality while nonuniformly reducing the total number of array elements. Nevertheless, this solution has yet to be demonstrated in real time, and its feasibility in high-frame-rate imaging via unfocused transmissions has not been shown [19].

Other approaches to reduce the data streaming traffic in ultrasound imaging involve hardware modifications, either by reducing the total number of array channels or by executing preprocessing within the front-end hardware. The use of sparse arrays is one example of such an approach. Although their use can directly reduce the data streaming traffic, sparse arrays typically have higher sidelobe levels in their field profile, and this issue is known to compromise image quality [20]–[23]. Alternatively, the incorporation of hardware microbeamformers within the transducer front end, which execute part of the DAS beamforming process, also allows for fewer data samples to be streamed to the system back end [24], [25]. Yet, this approach inherently makes the system less reconfigurable, so it is ill-suited for the implementation of new channel-domain imaging methods [2]. Perhaps, a less restrictive data transfer reduction solution is to add a lossless compression–decompression algorithm to the system’s data streaming operations [26]. Nonetheless, the compression ratio that can be achieved tends to vary depending on the characteristics of the acquired channel RF samples, so it is uncertain as to whether the size of the compressed data streaming traffic can always be accommodated by the system.

In this article, we present a new real-time framework for data transfer reduction when performing high-frame-rate ultrasound imaging via a software beamforming approach. Our proposed framework works by uniformly halving the channel count on receive and then reconstructing the other half of the channels using a data-driven, deep-learning approach. We hypothesize that the total number of receive channels for robust image formation can be reduced by exploiting redundancies between the pulse echoes of adjacent channels. This redundancy stems from the prevailing concept in pulse-echo sensing that different transducer channels on the same aperture would tend to receive, for a given scatterer, a similar echo signature that is merely shifted in time. Accordingly, we posit that a convolutional neural network (CNN) architecture can be trained to capture the redundant structure within the channel RF data array. The trained network can then be applied to reconstruct full-channel RF datasets from truncated ones acquired with fewer array elements used on reception. In doing so, data streaming traffic can be effectively halved with minimal loss in image quality and with bearable computational cost that does not compromise real-time feasibility, as will be shown later in this article.

II. THEORY

A. Background Considerations

In principle, it is readily possible to reduce raw data transfer in ultrasound imaging by uniformly decimating every other channel on the array aperture during pulse-echo reception. Nevertheless, in doing so, the array pitch for beamforming would be concomitantly doubled. Using a large pitch relative to the acoustic wavelength (exceeding the spatial Nyquist limit of \( \lambda/2 \)) results in spatial aliasing that, in turn, introduces imaging artifacts along the lateral dimension and reduces the contrast of echolucent regions [27]. Fortunately, this drawback is in theory addressable because, according to modern signal processing, even the Nyquist sampling limit can be overcome with additional knowledge of the structure of the signal [28], [29]. As explained in the Appendix, ultrasound RF data are indeed highly structured across adjacent channels. Accordingly, our proposed framework has been devised to implicitly take advantage of this structural redundancy via a data-driven, deep-learning approach to recover decimated channel-domain RF data and, in turn, generate ultrasound images with comparable quality as those formed using RF data acquired from the fully populated array.

B. Overview of the Proposed Framework

A convolutional encoder–decoder network is designed to use learned convolutional kernels to extract features in the RF data and to perform data recovery accordingly. This network forms the basis of our proposed pre-beamformed data formation framework as shown in Fig. 1, together with the overall imaging system. After a transmit event, the framework (dashed box) first acquires from half of the receive channels (odd-indexed) of the ultrasound probe, then infers the unacquired channels (even-indexed) using the neural network, and finally interleaves the acquired and inferred channels to reform a full-channel set of data. The reconstructed full dataset is then passed to a beamforming module to generate an image frame.

The encoder–decoder network is chosen because this architecture is suitable for the RF data inference task based on image processing literature. In modern image processing, encoder–decoders are used for image restoration tasks such as inpainting (i.e., the recovery of lost data in images and videos [30]), which can be considered analogous to the recovery of missing RF data between channels in ultrasound. Similar to inpainting, after data-driven training, the encoding stage of the neural network learns to generate a compressed representation of the most important spatiotemporal features and the decoding stage learns to regenerate the missing portions of RF data based on those encoded features.
III. CUSTOM ENCODER–DECODER IMPLEMENTATION

A. Number of Layers, Kernels per Layer, and Kernel Sizes

In a CNN, the capacity to capture complex structures is governed by the number of layers, the number of kernels per layer, and the kernel sizes. Increasing these parameters allows for increased capacity to capture complex structures; however, doing so comes at the expense of increased computational cost.

To balance this tradeoff, a nine-layer encoder–decoder network was chosen. The encoder–decoder has four encoding layers to extract the redundant structure in the RF frame, four mirrored decoding layers for inference, and one recombination layer to combine the inferred features into a single RF frame. The number of kernels per layer was also chosen to balance the tradeoff between capacity and computation. The four encoding layers, respectively, have 8, 32, 64, and 64 number of kernels; this subarchitecture was mirrored in the decoding layers. The number of kernels per layer along with the convolutional stride length helps to compress the RF data into feature maps in encoding and to infer the RF data in decoding.

The convolutional kernel sizes affect the extent of the structure captured in each layer. The sizes were selected, so that the first layer captures more of the features related to the fundamental RF frequency along the depth dimension by using a long kernel of size (5,3). The second layer captures more of the salient spatiotemporal features using a wide kernel of size (6,8). The third and fourth convolutional layers, size (3,3), add depth to the network to better capture the redundant structure. The subsequent deconvolutional layers (first four layers of decoding stage in Fig. 2) mirror the parameters of the encoding layers to form a standard encoder–decoder architecture. The convolutional layer at the end of the decoding stage (Fig. 2, in blue) used a kernel of size (5,5) to recombine the inferred features into a single image.

B. Convolutional Strides

To compress and decompress the extracted RF data features, the proposed network used a stride length of two for all layers but the second encoding layer and its corresponding decoding layer (white layers in Fig. 2). The stride of a convolutional layer represents sample “skipping” in each dimension [31]. For example, a kernel with a stride length of one performs convolution at each index. A convolutional kernel with a stride length of two performs convolution at every other index, compressing the features of the RF data. Similarly, deconvolutional strides expand the features of the RF data.

C. Skip Connections

As shown in Fig. 2, skip connections were used to connect the first and second encoding layers to the last and second

---

**Fig. 1.** Diagram of overall ultrasound imaging system (solid box), with RF data formation framework shown (dashed box). The proposed framework takes in odd-indexed channels, uses a CNN to infer the even-indexed channels, and then interleaves the channels to output a full-channel RF dataset.

**Fig. 2.** Shows the overall structure of the proposed encoder–decoder network. The network first takes in the odd-indexed RF channels as its input (light blue) and then prepares the data using zero-padding (green). Next, the prepared data enter the encoding stage of the network, where the redundant spatiotemporal features are extracted and compressed via four layers of convolutional kernels. Afterward, the compressed representation is decoded to infer the unacquired RF data via four deconvolutional layers. After the decoding stage, the resulting feature maps are combined into one using the last convolutional layer (blue). Lastly, the feature map is cropped (orange) to form the inferred even-indexed RF channel data for output (purple). In Sections II-D and III, the RF data input–output and the encoder–decoder network are discussed in detail.

**D. RF Data as an Image for Encoder–Decoder Input–Output**

The input and output of the convolutional encoder–decoder are formed by treating the RF data across channels as a single image. Channelwise concatenation of RF data, visualized as an image in Fig. 1, exhibits spatiotemporal structure across both fast time and across the channels. Even in just half the channels, the hyperbolic shape that the framework leverages for RF data inference is still present in the RF data frame, as shown in top left of Fig. 1 (under “odd-indexed channels”). Before entering the encoding–decoding stage of the network, the RF data frame is zero-padded to ensure that the appropriate sample size is used for downsampling along the depth dimension across all encoding layers. The kernels of the proposed network then capture and compress the structure in the odd-indexed channels and use that structure to infer the even-indexed channels. Finally, the RF data frame is cropped to return the output to the original size of the input RF data.
Fig. 2. Proposed architecture of the convolutional encoder–decoder. Odd-indexed RF channels (light blue) are input to the network and zero-padded (green) into a multiple of 8 in size. The encoding layers compress the information, while the decoding layers interpret the compressed information. The number and size of the kernels for each encoding and decoding layers are shown below the corresponding layers (dark and light gray). Two sets of skip connection concatenate the outputs of the encoding layers (start of arrows) to the inputs to the decoding layers (end of arrows). The last convolutional layer (dark blue) recombines the information into a single RF frame, which is cropped (orange) to ensure preservation of input and output dimensions, resulting in the inferred even-indexed RF channels (purple).

last decoding layers. These connections enable extracted features from the first two layers to have additional effect in inferring the unacquired channels in the decoder stage of the network. Skip connections also stabilize the gradient descent methods used for neural network training [32]. A drawback of using skip connections is that additional computation time is required for inference; thus, in the proposed encoder–decoder, the skip connections are limited to the outer layers with the fewer number of kernels.

D. Activation Function

Apart from the last convolutional layer, a leaky rectified linear unit (ReLU) activation function (α = 0.01) is used to better facilitate the training. While a standard ReLU takes a value of 0 for inputs below 0, the leaky ReLU has a very small slope, which has been shown to help in training by alleviating the vanishing gradient problem [33]. In contrast, the last convolutional layer uses a linear activation, since the expected even-channeled RF data outputs span both positive and negative values.

IV. EXPERIMENTAL METHODS

A. Acquisition of Neural Network Training Data

In order to train the CNN presented in the previous section, a total of 62,799 steered in vivo RF datasets were acquired. These datasets were captured along both the short axis and the long axis of human carotid artery from seven healthy volunteers (age: 25.9 ± 4.9) using a research imaging platform (SonixTouch; Analogic Ultrasound; Peabody, MA, USA). The research platform was programmed to transmit 31 steered plane wave pulses (2-cycle 5-MHz pulse; −15° to 15°, 1° separation; pulse-repetition frequency (PRF): 10 kHz) with an L14-5 array transducer (pitch: 0.3048 mm; Analogic Ultrasound). This research data acquisition protocol has been approved by the Clinical Research Ethics Committee of the University of Waterloo (Protocol No. 31694). The received echoes were digitized using a pre-beamformed data acquisition tool [34] with 12-bit resolution at 20-MHz sampling rate. The acquired RF data were subsequently transferred to a computer server for off-line processing. Table I summarizes the ultrasound scanner details and imaging parameters used for data acquisition.

| Parameter                          | Details                        |
|-----------------------------------|--------------------------------|
| Ultrasound Scanner               | SonixTouch                     |
| Ultrasound Probe                  | L14-5                          |
| Number of Tx/Rx Channels          | 128                            |
| Array Pitch                       | 0.3048 mm                      |
| Number of Tx Pulse Cycles         | 2                              |
| Transmit Frequency                | 5 MHz                          |
| Transmit Angle Range              | −15° to 15°                    |
| Sampling Rate                     | 20 MHz                         |
| Imaging Depth                     | 60 mm                          |
| Pulse Repetition Frequency        | 10 kHz                         |

B. Data Preprocessing

After acquisition, the data were preprocessed to prepare the training inputs and training references in MATLAB (version 2016a; MathWorks, Natick, MA, USA) using a computer server (SYS-4028-TRT; Super Micro, San Jose, CA, USA) running a Xeon E5-2620 central processing unit (Intel, Santa Clara, CA, USA) and an RTX-2080 GPU (Nvidia, Santa Clara,
CA, USA). Fig. 3 depicts the data processing pipeline that the acquired RF data follow, so that it is properly split and scaled to serve as training input and reference pairs. First, the initial 200 samples from each channel were discarded to minimize the effect of the transmit pulse on neural network training. The acquired samples were then limited to within the $-0.5$ to $0.5$ range by dividing by 4096 (the 12-bit samples acquired from the SonixTouch are stored as integers from $-2048$ to $2047$) as it is within the range that most CNN implementations work best. The odd-numbered channels and even-numbered channels were then split into $1300 \times 64$ sized input and reference RF frames for the convolutional encoder–decoder to learn.

**C. Neural Network Training and Evaluation**

The proposed encoder–decoder CNN was implemented in Python (version 3.6.7) using the TensorFlow-GPU (version 1.12.0) back end for GPU-integrated training neural network training. Keras (version 2.1.6), TensorFlow’s high-level application programming interface (API), was used to define the custom encoder–decoder architecture and to call the TensorFlow-GPU back end for GPU-accelerated training and inference. To train the implemented network, 80% of the acquired dataset was used and further split into training set (85%) and validation set (15%). The training was performed for 50 epochs by optimizing the mean-squared error (MSE) loss function using the Adam optimization algorithm [35] at a learning rate of 0.001 and a batch size of 4. Subsequently, the remaining 20% of the acquired dataset was used to evaluate the performance of the network by examining the normalized root-mean-squared error (NRMSE). The NRMSE was evaluated on three data variants: the raw RF data, the amplitude of the Hilbert transformed RF data, and the phase of the Hilbert transformed RF data.

**D. Image Quality Evaluation of the Proposed Framework**

After training, new RF datasets were acquired and beamformed with both original and inferred data to assess the performance of the overall framework. Using the same parameters as in Table I, new datasets were acquired from the carotids of healthy subjects outside of the training set subjects. Additionally, RF dataset was acquired from a multi-point target phantom (Multi-Purpose Multi-Tissue Ultrasound Phantom; Computerized Imaging Reference Systems, Inc., Norfolk, VA, USA) using a steering angle span of $-14.75^\circ$ to $+14.75^\circ$ (increments of 0.5°). This phantom test case served to assess the generalizability of our network both to angles and to imaged structure that was entirely different from our training dataset. The images formed from these acquisitions served to evaluate image quality in terms of resolution, contrast, and spatial aliasing artifacts.

To analyze the efficacy of the proposed framework in recovering image quality and suppressing spatial aliasing artifacts, three derivative RF datasets were generated from the newly acquired RF data.

1) **Odd-indexed channels**: The odd-indexed channels (64 in total) of the acquired RF data for each angle.

2) **Linearly interpolated channels**: This RF dataset represents a reference data recovery scheme for the unacquired channel data. The recovery was done by linearly interpolating the neighboring odd-indexed channels to form the unacquired even-indexed channels to restore the full-channel (128) RF dataset.

3) **CNN-inferred channels**: This dataset was formed using our proposed framework. The odd-indexed channels were extracted from the original acquisition and passed into the CNN encoder–decoder to infer the even-indexed channels for full-channel (128) restoration.

B-mode images were subsequently formed from the original full-channel dataset and the three derivative RF datasets using a conventional DAS beamforming framework according to the parameters in Table II. The RF data were first prefiltered with a 3−7-MHz finite-impulse response (FIR) bandpass filter applied along the fast-time dimension. The RF data were then converted to the analytic form using the Hilbert transform. Afterward, DAS beamforming was performed to generate the four sets of images; a constant F-number of 1.25 and a uniform (rectangular) apodization were used.

**E. Evaluation of Image Quality After Coherent Compounding**

Plane wave images are often coherently compounded to improve the image quality as compared to images beamformed from a single transmission. The image quality (spatial aliasing artifacts, contrast) was also assessed after coherently compounding increasing subsets of the beamformed images from steered transmissions. The steering angles used for compounding were selected in pairs, so that each compounded image was always formed from balanced positive and negative steering angles. The single transmit images were also selected to be evenly distributed along the angle span (e.g., $-8^\circ$ and $8^\circ$ for
two compounded angles; $-10^\circ$, $-5^\circ$, $5^\circ$, and $10^\circ$ for four compounded angles).

**F. Evaluation of Computational Capacity**

To assess the possibility for real-time inference using the proposed framework, the inference time for RF channel regeneration was calculated using Python’s built-in timing functionality (time.clock); 250 RF data inference operations were conducted and their timings were averaged.

**V. RESULTS**

**A. Trained CNN Encoder–Decoder Can Recover the Unacquired Channel Structure**

The proposed CNN encoder–decoder was able to recover the structure of even-channeled RF data for the unseen test set data (i.e., not included for training and validation). As an example, Fig. 4(a) shows an in vivo human carotid B-mode image (40-dB dynamic range) generated from the test set along with the corresponding reference even-channel RF frame [see Fig. 4(b)] and the inferred RF frame [see Fig. 4(c)] based on the odd-channeled RF input. The overall structure in the inferred RF data generally matched the reference one; this observation is substantiated by the low NRMSE (0.012) between the two RF images.

To more closely examine the performance of the network at different depths, Fig. 4(d)–(i) shows the enlarged regions of RF data at increasing depths [depicted by the corresponding dotted boxes in Fig. 4(b) and (c)]. As can be observed, the local structures in the inferred RF data (right column) resembled that of reference RF data (left column) for shallow and medium depths [see Fig. 4(d)–(g)] but were dissimilar for deep imaging depths [see Fig. 4(h) and (i)]. Table III quantifies the NRMSE (normalized to RF data range) in the image for each region and compares the CNN inference approach with linear interpolation of the channels. Additionally, the Hilbert transform of the RF data is taken, so that the NRMSE of the amplitude (normalized to RF amplitude range) and phase (normalized to $2\pi$) can be evaluated. RF data inferred with our CNN show that the CNN better matches the reference RF data in every scenario compared to linear interpolation. The higher NRMSE in deeper regions [see Fig. 4(h) and (i)] is due to the lack of structured scattering from that region as shown in the corresponding B-mode images [bottom of Fig. 4(a)].

**B. Grating Lobe Reduction and Preservation of Resolution in Multiple Point Target Phantom Across Transmit Angles**

To demonstrate that our data recovery scheme does not result in a loss of post-beamformed image quality, Fig. 5 displays point target images generated from the full-channel and the derivative datasets (left to right, see Section IV-D) with $-14.75^\circ$ (top row) and $-0.25^\circ$ (bottom row) steering transmissions. It can be observed that the images beamformed with the proposed method (second column) contain reduced spatial aliasing artifacts (highlighted by yellow squares) from the channel reduction as compared to images from other derivative datasets (third and fourth columns). This observation holds across the two steering angles shown in Fig. 5.

To quantify the reduction in grating lobe power across transmission angles, Fig. 6 plots the mean grating lobe power at different transmission angles, averaged over a $4 \times 4$ mm$^2$ region centered on the peak spatial aliasing artifacts on the left

**TABLE III**

| Region               | Entire RF | Shallow Depth | Medium Depth | Deep Depth |
|----------------------|----------|--------------|--------------|------------|
| CNN Interpolated (Int) | 0.012    | 0.050        | 0.026        | 0.137      |
| CNN Int. Amplitude   | 0.017    | 0.099        | 0.049        | 0.195      |
| CNN Int. Phase       | 0.206    | 0.165        | 0.092        | 0.239      |
| Linear Interpolated  | 0.022    | 0.075        | 0.037        | 0.138      |
| Linear Int. Amplitude| 0.025    | 0.142        | 0.096        | 0.198      |
| Linear Int. Phase    | 0.269    | 0.247        | 0.254        | 0.264      |

Fig. 4. (a) B-mode image of human carotid in vivo and the corresponding acquired even-indexed RF data (left column) and inferred even-indexed RF data (right column). Whole RF data “images” are shown in second row, (b) and (c). Snippets corresponding to respective orange dashed boxes in (d)–(i).
Fig. 5. Single plane wave transmission B-mode images (40-dB dynamic range) of point target phantom beamformed from original full-channel RF dataset (first column), CNN-inferred channel dataset (second column), linearly interpolated channel dataset (third column), and the original odd-indexed channel dataset (fourth column). Images are shown for $-14.75^\circ$ transmission [top row, (a)-(d)] and for $-0.25^\circ$ transmission [bottom row, (e)-(h)]. Yellow boxes highlight spatial aliasing regions used to calculate artifact power in Fig. 6.

Fig. 6. Measured spatial aliasing artifact power in B-mode images across transmission angles for original full-channel RF dataset (dark blue), CNN-inferred channel dataset (orange), linearly interpolated channels dataset (light purple), and the original odd-indexed channel dataset (light blue). Thumbnails of B-mode images beamformed from CNN-inferred channels shown at bottom.

C. Recovery of Carotid In Vivo Lumen-to-Tissue Contrast

Images beamformed from RF data inferred with our method also show improved image quality for in vivo scenarios. Specifically, it can be observed in Fig. 8 that the images of the human carotid in vivo generated from the full-channel (first column) and the CNN-derived (second column) datasets demonstrate clear visibility of the common carotid artery.
lumen (center-top of each image). In contrast, in the linearly interpolated (third column) and odd-indexed (fourth column) datasets, the carotid lumen is hidden by artifacts from channel count reduction for both carotids, regardless of the transmit steering.

Examining the lumen-to-tissue contrast across transmission angles for each compared dataset, Fig. 9 plots the measured contrast difference over $4 \times 4$ mm$^2$ regions of lumen and tissue [yellow and green squares, respectively, in Fig. 8(e)]. On average, the lumen-to-tissue contrast of images beamformed from the CNN-derived dataset (orange line) is within 3 dB of the contrast of images beamformed from the full-channel dataset (dark blue line), while demonstrating between 8- and 10-dB contrast improvement over images beamformed from the odd-indexed channels (light blue line) alone. The lumen-to-tissue contrast of images generated from linearly interpolated data (light purple line) shows minor improvement but does not reach the contrast improvement of the proposed method.

### D. Consistent Image Quality Improvement After Coherent Compounding of Single Transmit Images

B-mode images formed from steered plane wave transmissions were coherently compounded to assess the image quality improvement compared to using only a single transmission. Figs. 10 and 11 demonstrate the effect of compounding images beamformed with CNN-inferred data using the same imaging scenarios as Figs. 6 and 9, respectively. Fig. 10 shows that coherent compounding reduces the spatial aliasing artifacts even without inference (light blue line), but our proposed CNN inference scheme results in reduced artifacts with fewer
XIAO et al.: MINIMIZING IMAGE QUALITY LOSS AFTER CHANNEL COUNT REDUCTION FOR PLANE WAVE ULTRASOUND

Fig. 9. Measured lumen–tissue contrast in B-mode images across transmission angles for original full-channel RF dataset (dark blue), CNN-inferred channel dataset (orange), linearly interpolated channels dataset (light purple), and the original odd-indexed channel dataset (light blue). Thumbnails of B-mode images beamformed from CNN-inferred channels shown at bottom.

angles compounded (left half of red line). The artifact power is measured via a $4 \times 4 \text{mm}^2$ region centered on the peak spatial aliasing artifacts on the left side for each compounded image. It is notable that the images beamformed with CNN-inferred data have lower average power in the artifact region when many angles are compounded (right half of red line). This result is attributed to the lack of structured scattering (see Section V-A) changing the speckle pattern. Fig. 11 shows that coherent compounding also improves visibility of the lumen without inference (light blue line), but CNN inference (red line) results in consistently improved (6–8 dB) lumen–tissue contrast overall. The contrast was measured over the same $4 \times 4 \text{mm}^2$ regions of lumen and tissue [yellow and green squares, respectively, in Fig. 8(e)]. In both Figs. 10 and 11, linear interpolation (light purple line) slightly improves the measured metrics compared to no channel inference (light blue line), but not to the degree of the proposed method.

E. Real-Time Computational Capacity

Using the TensorFlow-GPU framework (version 1.12.0) running on an Intel Xeon E5-2620 CPU (Intel, Santa Clara, CA, USA) and an Nvidia RTX 2080 GPU (Nvidia, Santa Clara, CA, USA), an inference rate of 4 ms is achieved per 64 channels of RF data. This rate corresponds to a 250-fps inference throughput on our system.

VI. DISCUSSION

A. Convolutional Encoder–Decoder as a New Framework for Channel Data Recovery

A primary advantage of ultrasound over other medical imaging modalities, such as CT or MRI, lies in its ability to perform point-of-care imaging [36]. Given its bedside applicability, the novel imaging insights provided by high-frame-rate imaging [3] are most effective when they are available to the sonographers in real time. However, two barriers in enabling real-time high-frame-rate imaging are: 1) the immense data transfer rate to move the data from the sampling front end to the processing back end within the imaging system and 2) the

Fig. 10. Measured spatial aliasing artifact power in B-mode images across a number of coherently compounded single-transmit images for original full-channel RF dataset (dark blue), CNN-inferred channel dataset (orange), linearly interpolated channels dataset (light purple), and original odd-indexed channel dataset (light blue). Thumbnails of compounded B-mode images beamformed from CNN-inferred channels shown at bottom.

Fig. 11. Measured lumen–tissue contrast in B-mode images across a number of coherently compounded single-transmit images for original full-channel RF dataset (dark blue), CNN-inferred channel dataset (orange), linearly interpolated channels dataset (light purple), and original odd-indexed channel dataset (light blue). Thumbnails of compounded B-mode images beamformed from CNN-inferred channels shown at bottom.
computational cost of real-time beamforming at high frame rates.

Our work addresses the data transfer challenge of real-time high-frame-rate imaging via channel count reduction without sacrificing image quality. Simply reducing the transducer receive channel count results in a loss of post-beamformed image quality due to the introduction of artifacts and the loss of contrast. To compensate for the channel reduction, a neural network was trained, using over 60,000 in vivo sets of plane wave data (see Table I), to infer unacquired RF channel data from the acquired channel RF data. An encoder–decoder network structure (see Fig. 2) was designed to leverage the hyperbolic structure of the acoustic echoes to regenerate data from unacquired channels. Examining only the inferred RF data (see Fig. 4), our CNN correctly captures the expected spatiotemporal structure. After CNN inference, a set of full-channel RF data can be formed by interleaving the acquired and inferred channel data (see Fig. 1). B-mode images beam-formed from the regenerated RF data closely matched the images beamformed from the full-channel RF data in grating lobe artifact power (see Fig. 6), point target resolution (see Fig. 7), and lumen–tissue contrast (see Fig. 9), even when coherent compounding was applied (see Figs. 10 and 11).

To minimize CNN inference time for real-time performance, only nine convolutional layers were employed, leading to an inference rate of 250 RF fps on our system when executed on a GPU. Such inference rate can support both a live imaging mode and replay mode of a real-time high-frame-rate imaging system, such as the one described in [37]. To form a functional system that involves the entire signal processing chain from channel data recovery to image formation, the computational challenge of real-time high-frame-rate beamforming must also be addressed. Previous works have already demonstrated leveraged GPUs for real-time image formation [4]. Thus, we anticipate that a two-GPU system, where one GPU performs the inference before passing the full set of RF data to the second GPU for beamforming, can help realize such a system.

Our proposed data reduction and recovery framework has several advantages. First, the network demonstrated angle independence on data from steering angles ranging from −15° to +15°. As such, the same network can be used for inference from all the acquired angles. This capability represents an improvement over our preliminary results [38] that only inferred unsteered plane wave data using the same number of layers. Second, the trained CNN showed high generalizability across imaging scenarios. Reconstructed images were derived from a wire phantom in vitro (see Fig. 5), but the training data consisted solely of in vivo samples that contained no clear point targets, highlighting the generalizability of our framework. Third, the ease of training data formation is another important advantage of the proposed framework. While our method falls under the umbrella of supervised learning within the topical theme of deep learning, our training input–output pairs (odd/even indexed RF channels) are formed without any user-assisted labeling (see Fig. 3). One consequence of this feature is that for a given ultrasound imaging system, all plane wave RF data acquired without data reduction can then form a new training pair to further improve the network.

B. Impact of a Pure RF-to-RF Framework

Conventionally, when receiving acoustic echoes, the ultrasound transducer pitch obeys the Nyquist spatial sampling criterion. Violation of the Nyquist sampling rate leads to spatial aliasing, but other works have shown that the Nyquist rate can be overcome given knowledge of signal structure [28], [29]. In the Appendix, we show based on physical principles that spatiotemporal structure exists in received acoustic echoes from a plane wave ultrasound transmission. Our experimental results showed that the proposed deep learning method successfully inferred the RF data structure using convolutional kernels. According to the parameters of system used for this study (see Table I), the fundamental frequency was 5 MHz (wavelength of 0.308 mm at 1540-m/s sound speed), while the effective receive pitch was double of the wavelength (i.e., 0.61 mm for the pitch of odd-indexed channels). Given this discrepancy, our results suggest that the Nyquist limit for transducer element pitch is too conservative in ultrasound, given an appropriate compensation technique. This conclusion is in the vein of similar insights from works, such as [19] or [39]. Additionally, because CNNs are universal function approximators [40], there may exist undiscovered alternative nonlearning methods, which can also perform RF-to-RF data regeneration.

The value of our framework’s RF-to-RF nature is its ability to output raw RF data. This benefit can potentially enable the derivation of high-frame-rate imaging insights outside of B-mode imaging. Some high-frame-rate ultrasound methods, such as color Doppler [41], vector flow imaging [42], [43], or shear wave elastography [44], rely on the availability of the raw RF data or the phase to implement custom signal processing techniques to extract the desired information. Our proposed framework attempts to regenerate the raw RF data, unlike other methods to reduce channel count, which focus only on improving B-mode image quality, such as [19] or [45]. The fidelity of the reconstructed RF data (see Fig. 4) lends credence to our data reduction framework finding applications in non-B-mode ultrasound imaging and evaluating our framework’s applicability to color Doppler is an ongoing effort. The ability to reduce the data on acquisition while maintaining a full RF dataset for image formation is essential for real-time display in a clinical setting.

C. Limitations and Perspectives on Future Work

Our proposed framework demonstrated millisecond inference frame rates (Section V-E). However, true real-time high-frame-rate ultrasound requires submillisecond inference to achieve 1000-fps imaging. There are several avenues to pursue to further improve the inference speed without compromising the fidelity of the regenerated RF data. Without modifying the network architecture, inference frameworks such as Nvidia’s TensorRT (Nvidia, Santa Clara, CA, USA) can perform network optimizations and have shown up to twofold speedup of inference. Apart from software acceleration, next-generation hardware, such as Google’s tensor processing unit (Google, Mountain View, CA, USA), is being produced and specifically tailored for neural network operations. These hardware
advances can be integrated into an ultrasound open platform to add efficient CNN operations to the ultrasound imaging pipeline. Another possibility is to reduce the number of layers or kernels and correspondingly decrease the inference time. Nonetheless, in doing so, data recovery performance may suffer as the network learns to capture less.

High-frame-rate ultrasound imaging is also not limited to the imaging system used in this study. While this work was demonstrated using a specific probe (L14-5) and imaging system (SonixTouch), the proposed deep learning framework can be generalized to other linear probes and systems through acquisition of new data and network retraining. The existence of spatiotemporal redundancy in the RF data and the insights of applying an encoder–decoder architecture by treating the RF data as an image frame (i.e., a 2-D data array) are widely applicable. A more challenging barrier is expanding the framework beyond linear array probes as the current framework relies on applying spatially invariant convolutional kernels on raw RF data. Curvilinear probes, for example, show an increase in physical distance in neighboring channels with increased imaging depth. One suggestion to extend our work for curvilinear arrays is through scan conversion of the raw RF data before inference. Yet, this approach may degrade the fidelity of the data through excess interpolation.

In this work, the total data transfer was halved. Further reduction may be achieved via two means. First, the receive channels can potentially be decimated three or even four times with some modifications to the neural network, as the principles of data structure redundancy (see Appendix) are still valid for those scenarios. Second, as our framework is purely RF-to-RF, it can be used in conjunction with other data reduction schemes. The proposed method does not currently require hardware modifications but could potentially be applied together with a hardware-based compression scheme [26]. It remains to be investigated what the practical limit for channel reduction is for ultrasound imaging.

High-volume-rate 3-D ultrasound imaging is an area where data transfer reduction is even more necessary and is a natural application direction for our framework. Densely populated matrix probes can have upward of 1000 elements, while sparsely populated probes can suffer from low image quality [22], [46]. Our work provides a possible framework for bridging the gap by acquiring from a sparse matrix probe and then regenerating the RF data to emulate a fully populated matrix probe. This strategy can be realized either by applying the 2-D convolutional network architecture presented in our work one dimension at a time or by redevelopment of a network using 3-D convolutional layers [47], [48] with the possibility of using simulated data for network training.

VII. Conclusion

Reducing the data transfer rate can alleviate one bottleneck in real-time high-frame-rate ultrasound imaging. In this article, RF data transfer was halved by acquiring data from only half the transducer channels. The subsequent loss of post-beamformed image quality due to channel reduction was mitigated in vitro and in vivo using a deep learning approach to regenerate the RF data from the unacquired channel set. This method not only helps to enable real-time high-frame-rate imaging but also has a natural extension to the 3-D ultrasound imaging domain, where the data transfer rate remains an even larger barrier.

Appendix A

Explanation of Structural Redundancy of Channel-Domain Signals

To explain the origin of the redundant structure present in the pulse-echo signals received over different array channels, let us consider the basic case of pulse-echo sensing for a point target. Fig. 12 illustrates a setup of five channels and a reflecting point source on a 2-D plane. Each receiving element has coordinates \((0, x_0), (0, x_1), (0, x_2), (0, x_3),\) and \((0, x_4)\), respectively, and the point target has coordinates \((z_p, x_p)\). Assuming an ideal plane wave transmission at time \(t_0 = 0\) and a constant speed of sound \(c_0\), each transducer element (indexed by \(e\)) would receive the backscattered signal from the point source at the following time:

\[
t_e = \frac{z_p + \sqrt{z_p^2 + (x_p - x_e)^2}}{c_0}.
\] (A1)

Rearranging the terms to remove the radical and forming the equation of a conic section, we arrive at

\[
\frac{z_p^2}{c_0^2}
\left(t_e - \frac{z_p}{c_0}\right)^2 - \frac{1}{z_p^2}(x_p - x_e)^2 = 1.
\] (A2)

In this scenario, the signal received by each element is identical apart from the time of reception. Because of the isotropic reflection of the point source, the delays form a hyperbola in relation to the spacing between elements as in (A2). This delay calculation is the fundamental principle behind DAS beamforming, but this spatial structure can also be leveraged in a novel way for data reduction.

From algebra and geometry, given only two pairs of values \((t_e, x_e)\) and knowing the delays form a hyperbolic shape, the equation becomes fully determined and can be obtained by solving a system of equations with two variables. Therefore, in the single point-source, five-transducer element example without noise or other adverse effects, three of the five delayed signals can be considered to carry redundant information and
can be perfectly reconstructed from the other two. Extrapolating from this example to a fully populated linear array, we argue that there exist many such spatiotemporal redundancies between fast-time sets of RF data collected from neighboring channels. Each channel receives time-shifted version of echo pulses from the same scattering sources in the insonified medium.

Let us now extend the explanation to the case of a multi-point target phantom. Fig. 13 shows a snippet of the spatiotemporal structure of the RF data for such a phantom (the same one reported in Section IV-D; data were acquired using the same imaging parameters as those reported in Table I). For each point target in the phantom, three of which are highlighted in yellow in Fig. 13(a), the hyperbolic shape of the corresponding echoes is shown in Fig. 13(b). While the hyperbolic relationship in (A2) still exists, the signals can suffer from the nonidealities of attenuation and transducer angle dependence, as depicted by the reduction in signal strength away from the center of the hyperbolic structure, and potential superposition from other scatterers. Additionally, the signals from human tissue rarely manifest as those from true echogenic point sources. These reasons prohibit the exact use of (A2) to recover the unacquired channels in a typical imaging scenario. Instead, we take the data-driven approach presented in this work to learn an optimal method for leveraging the transducer channel redundancy to recover unacquired channels.

Fig. 13. (a) B-mode image of a multi-point target phantom and (b) corresponding 128 channel by 200 sample snippet of raw RF data highlighted in dashed yellow box.

References

[1] T. A. Whittingham, “Medical diagnostic applications and sources,” Prog. Biophys. Mol. Biol., vol. 83, pp. 84–110, Jan. 2007.

[2] E. Boni, A. C. H. Yu, S. Freear, J. A. Jensen, and P. Tortoli, “Ultrasound open platforms for next-generation imaging technique development,” IEEE Trans. Ultrason., Ferroelectr., Freq. Control, vol. 65, no. 7, pp. 1078–1092, Jul. 2018.

[3] M. Tanter and M. Fink, “Ultrafast imaging in biomedical ultrasound,” IEEE Trans. Ultrason., Ferroelectr., Freq. Control, vol. 61, no. 1, pp. 102–119, Jan. 2014.

[4] B. Y. S. Yu, I. K. H. Tsang, and A. C. H. Yu, “GPU-based beamformer: Fast realization of plane wave compounding and synthetic aperture imaging,” IEEE Trans. Ultrason., Ferroelectr., Freq. Control, vol. 58, no. 8, pp. 1698–1705, Aug. 2011.

[5] G. Montaldo, M. Tanter, J. Bercoff, N. Benech, and M. Fink, “Coherent plane-wave compounding for very high frame rate ultrasonography and transient elastography,” IEEE Trans. Ultrason., Ferroelectr., Freq. Control, vol. 56, no. 3, pp. 489–506, Mar. 2009.

[6] J. A. Ketterling et al., “High-speed, high-frequency, in utero vector-flow imaging of mouse embryos,” Sci. Rep., vol. 7, Nov. 2017, Art. no. 16658.

[7] R. Urs, J. A. Ketterling, A. C. H. Yu, H. O. Lloyd, B. Y. S. Yu, and R. H. Silverman, “Ultrasonic imaging and measurement of choroidal blood flow,” Translational Vis. Sci. Technol., vol. 7, no. 5, p. 5, Sep. 2018.

[8] J. Udesen, F. Gran, K. L. Hansen, J. A. Jensen, C. Thomsen, and M. B. Nielsen, “High frame-rate blood vector velocity imaging using plane waves: Simulations and preliminary experiments,” IEEE Trans. Ultrason., Ferroelectr., Freq. Control, vol. 55, no. 8, pp. 1729–1743, Aug. 2008.

[9] M. Tanter, J. Bercoff, L. Sandrin, and M. Fink, “Ultrasonic compound imaging for 2-D motion vector estimation: Application to transient elastography,” IEEE Trans. Ultrason., Ferroelectr., Freq. Control, vol. 49, no. 10, pp. 1363–1374, Oct. 2002.

[10] E. Macé, G. Montaldo, I. Cohen, M. Baulac, M. Fink, and M. Tanter, “Functional ultrasound imaging of the brain,” Nature Methods, vol. 8, pp. 662–664, May 2011.

[11] H. S. J. Chen, B. Y. S. Yu, and A. C. H. Yu, “Medical ultrasound imaging: To GPU or not to GPU?” IEEE Micro, vol. 31, no. 5, pp. 54–65, Sep. 2011.

[12] M. Hansen, D. Schaa, and J. A. Jensen, “Synthetic aperture beamformation using the GPU,” in Proc. IEEE Int. Ultrason. Symp., Oct. 2011, pp. 373–376.

[13] T. Chernyakova and Y. Eldar, “Fourier-domain beamforming: The path to compressed ultrasound imaging,” IEEE Trans. Ultrason., Ferroelectr., Freq. Control, vol. 61, no. 8, pp. 1252–1267, Aug. 2014.

[14] H. Liebgott, R. Prost, and D. Friboulet, “Pre-beamformed RF signal reconstruction in medical ultrasound using compressive sensing,” Ultrasound, vol. 53, pp. 525–533, Feb. 2013.

[15] J. Liu, Q. He, and J. Luo, “A compressed sensing strategy for synthetic transmit aperture ultrasound imaging,” IEEE Trans. Med. Imag., vol. 36, no. 4, pp. 878–891, Apr. 2017.

[16] T. Chernyakova et al., “Fourier-domain beamforming and structure-based reconstruction for plane-wave imaging,” IEEE Trans. Ultrason., Ferroelectr., Freq. Control, vol. 65, no. 10, pp. 1810–1821, Oct. 2018.

[17] I. A. M. Huijben, B. S. Veelings, K. Janse, M. Misschi, and R. J. G. van Sloun, “Learning sub-sampling and signal recovery with applications in ultrasound imaging,” IEEE Trans. Med. Imag., vol. 39, no. 12, pp. 3955–3966, Dec. 2020.

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

[19] R. Cohen and Y. C. Eldar, “Sparse convolutional beamforming for ultrasound imaging,” IEEE Trans. Ultrason., Ferroelectr., Freq. Control, vol. 65, no. 12, pp. 2390–2406, Dec. 2018.

[20] O. T. von Ramm, S. W. Smith, and F. L. Thustone, “Grayscale imaging with complex TGC and transducer array,” in Proc. SPIE, J. E. Gray and W. R. Hendee, Eds., vol. 70, Jan. 1975, pp. 266–270.

[21] A. Austeng and S. Holm, “Sparse 2-D arrays for 3-D phased array imaging-design methods,” IEEE Trans. Ultrason., Ferroelectr., Freq. Control, vol. 49, no. 8, pp. 1073–1086, Aug. 2002.

[22] B. Diarra, M. Robini, P. Tortoli, C. Cachard, and H. Liebgott, “Design of optimal 2-D nongrid sparse arrays for medical ultrasound,” IEEE Trans. Biomed. Eng., vol. 60, no. 11, pp. 3093–3102, Nov. 2013.

[23] S. K. Mitra, K. Mondal, M. K. Tchobanou, and G. Jovanovic, “General polynomial factorization-based design of sparse periodic linear arrays,” IEEE Trans. Ultrason., Ferroelectr., Freq. Control, vol. 57, no. 9, pp. 1952–1966, Sep. 2010.

[24] J. D. Larson, “2-D phased array ultrasound imaging system with distributed phasing,” U.S. Patent 5 229933, Jul. 20, 1993.

[25] S. C. Leavitt, B. J. Savord, B. M. Herrick, J. R. Fallon, and M. Mooney, “Sub-beamforming apparatus and method for a portable ultrasound imaging system,” U.S. Patent 6 491 634 B1, Dec. 10, 2002.

[26] U. Lok, G. Fan, and P. C. Li, “Lossless data compression for improving the performance of a GPU-based beamformer,” Ultrason. Imag., vol. 37, no. 2, pp. 135–151, Apr. 2014.

[27] F. W. Kremkau and K. Taylor, “Artifacts in ultrasound imaging,” Ultrasonic Med., vol. 5, no. 4, pp. 227–237, Apr. 1986.

[28] E. J. Candès, J. Romberg, and T. Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” IEEE Trans. Inf. Theory, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[29] M. Mishali and Y. C. Eldar, “From theory to practice: Sub-Nyquist sampling of sparse wideband analog signals,” IEEE J. Sel. Topics Signal Process., vol. 4, no. 2, pp. 375–391, Apr. 2010.

[30] J. Xie, L. Xu, and E. Chen, “Image denoising and inpainting with deep neural networks,” in Proc. Adv. Neur., 2012, pp. 341–349.

[31] L. Springerberg, A. Desovich, D. H. Brock, and M. Riedmiller, “Stirring for simplicity: The all convolutional net,” in Proc. Workshop Int. Conf. Learn. Represent., 2015, pp. 1–14.
XIAO et al.: MINIMIZING IMAGE QUALITY LOSS AFTER CHANNEL COUNT REDUCTION FOR PLANE WAVE ULTRASOUND 2861

[32] M. Drozdal, E. Vorontsov, G. Chartrand, S. Kadoury, and C. Pal, “The importance of skip connections in biomedical image segmentation,” in Proc. Int. Workshop Deep Learn. Med. Image Anal., 2015, pp. 234–241.

[33] A. L. Maas, A. Y. Hannun, and A. Y. Ng, “Rectifier nonlinearities improve neural network acoustic models,” presented at the ICML Workshop Deep Learn. Audio Speech Lang. Process., Jun. 2013, pp. 1–6.

[34] C. C. F. Cheung et al., “Multi-channel pre-beamformed data acquisition system for research on advanced ultrasound imaging methods,” IEEE Trans. Ultrason., Ferroelectr., Freq. Control, vol. 59, no. 2, pp. 243–253, Feb. 2012.

[35] D. Kingma and J. Ba, “Adam: A method for stochastic optimization,” in Proc. Int. Conf. Learn. Represent. (ICLR), May 2015, pp. 1–11.

[36] P. N. T. Wells, “Ultrasound imaging,” Phys. Med. Biol., vol. 51, no. 13, pp. R83–R99, Jun. 2006.

[37] B. Y. S. Yu, M. Walczak, M. Lewandowski, and A. C. H. Yu, “Live ultrasound color encoded speckle imaging platform for real-time complex flow visualization in vivo,” IEEE Trans. Ultrason., Ferroelectr., Freq. Contr., vol. 66, no. 4, pp. 656–668, Jan. 2019.

[38] D. Xiao, B. Y. S. Yu, A. J. Y. Chee, and A. C. H. Yu, “Channel count reduction for plane wave ultrasound through convolutional neural network interpolation,” in Proc. Int. Conf. Image Anal. Recognit., 2019, pp. 442–451.

[39] P. D. Wilcox and J. Zhang, “Quantification of the effect of array element pitch on imaging performance,” IEEE Trans. Ultrason., Ferroelectr., Freq. Control, vol. 65, no. 4, pp. 603–616, Apr. 2018.

[40] D.-X. Zhou, “Universality of deep convolutional neural networks,” Appl. Comput. Harmon. Anal., vol. 48, no. 2, pp. 787–794, Mar. 2020.

[41] C. Demené et al., “Spatiotemporal clutter filtering of ultrafast ultrasound data highly increases Doppler and ultrasound sensitivity,” IEEE Trans. Med. Imag., vol. 34, no. 11, pp. 2271–2283, Nov. 2015.

[42] J. A. Jensen, S. I. Nikolov, A. C. H. Yu, and D. Garcia, “Ultrasound vector flow imaging—Part II: Parallel systems,” IEEE Trans. Ultrason., Ferroelectr., Freq. Control, vol. 63, no. 11, pp. 1722–1732, Nov. 2016.

[43] M. Lenge, A. Ramalli, P. Tortoli, C. Cachard, and H. Liebgott, “Plane-wave transverse oscillation for high-frame-rate 2-D vector flow imaging,” IEEE Trans. Ultrason., Ferroelectr., Freq. Control, vol. 62, no. 12, pp. 2126–2137, Dec. 2015.

[44] J. Bercoff, M. Tanter, and M. Fink, “S supersonic shear imaging: A new technique for soft tissue elasticity mapping,” IEEE Trans. Ultrason., Ferroelectr., Freq. Control, vol. 51, no. 4, pp. 396–409, Apr. 2004.

[45] J.-F. Synnevåg, A. Austeng, and S. Holm, “Benefits of minimum-variance beamforming in medical ultrasound imaging,” IEEE Trans. Ultrason., Ferroelectr., Freq. Control, vol. 56, no. 9, pp. 1868–1879, Sep. 2009.

[46] J. Provost et al., “3D ultrafast ultrasound imaging in vivo,” Phys. Med. Biol., vol. 59, no. 19, pp. L1–L10, Oct. 2014.

[47] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, “Learning spatiotemporal features with 3D convolutional networks,” in Proc. IEEE Int. Conf. Comput. Vis., Dec. 2015, pp. 4449–4457.

[48] H. Shan et al., “3-D convolutional encoder-decoder network for low-dose CT via transfer learning from a 2-D trained network,” IEEE Trans. Med. Imag., vol. 37, no. 6, pp. 1522–1534, Jun. 2018.

Di Xiao (Graduate Student Member, IEEE) received the bachelor's degree in mathematics from the University of California at Berkeley, Berkeley, CA, USA, in 2014, and the master's degree in electrical engineering from the University of Minnesota–Twin Cities, Minneapolis, MN, USA, in 2016. He is currently pursuing the Ph.D. degree in electrical and computer engineering with the University of Waterloo, Waterloo, ON, Canada.

Previously, he was a Software Engineer at IBM Canada, Markham, ON, Canada, and has also interned at the Hitachi Central Research Laboratory, Tokyo, Japan. He is a member of the Laboratory on Innovative Technology in Medical Ultrasound (LITMUS), University of Waterloo. His current research is focused on beamforming and speed-of-sound estimation.

William M. K. Pitman (Graduate Student Member, IEEE) received the B.A.Sc degree in electrical engineering from Queen’s University, Kingston, ON, Canada, in 2019. He is currently pursuing the M.A.Sc. degree in electrical and computer engineering with the University of Waterloo, Waterloo, ON, Canada.

From 2017 to 2018, he was an Electrical Engineering Intern at L3Harris Wescam, Hamilton, ON, Canada. His current research is on applying artificial intelligence tools to develop new ultrasound imaging techniques.

Mr. Pitman was a recipient of the NSERC Alexander Graham Bell Canada Graduate Master’s Scholarship in 2021 and the Waterloo Engineering Excellence Master’s Fellowship in 2020.

Billy Y. S. Yiu (Member, IEEE) received the B.Eng. degree (Hons.) in medical engineering and the M.Phil. degree in electrical and electronic engineering from The University of Hong Kong, Hong Kong, in 2007 and 2010, respectively, and the Ph.D. degree in electrical and computer engineering from the University of Waterloo, Waterloo, ON, Canada, in 2019.

Before joining the University of Waterloo, he was a Research Staff with the Biomedical Ultrasound Laboratory, The University of Hong Kong, from 2010 to 2016. He is currently a Research Assistant Professor and a Co-Principal Investigator of LITMUS at the University of Waterloo. His research interest is in advanced ultrasound imaging techniques and systems.

Dr. Yiu was a past recipient of the USE Young Scientist Award, the KSUM Young Investigator Gold Prize Award, and the ASA Biomedical Acoustics Best Student Paper Award.

Adrian J. Y. Chee (Member, IEEE) received the B.Eng. degree in electrical and computer systems engineering from Monash University, Melbourne, VIC, Australia, in 2011, and the Ph.D. degree in mechanical engineering from The University of Hong Kong, Hong Kong, in 2016.

He was a Corporate Intern at Hitachi Central Research Laboratory, Tokyo, Japan, in 2016. He is currently an Associate Scientist with the Laboratory on Innovative Technology in Medical Ultrasound (LITMUS), University of Waterloo, Waterloo, ON, Canada. His research interests are in designing new vascular ultrasound techniques and related experimental tools.

Alfred C. H. Yu (Senior Member, IEEE) received the B.Sc. degree in electrical engineering from the University of Calgary, Calgary, AB, Canada, in 2002, and the M.A.Sc. and Ph.D. degrees in biomedical engineering from the University of Toronto, Toronto, ON, Canada, in 2004 and 2007, respectively.

From 2007 to 2015, he was the founding Principal Investigator of the Biomedical Ultrasound Laboratory, The University of Hong Kong, Hong Kong. In 2015, he relocated to the University of Waterloo, Waterloo, ON, Canada, where he is currently a Professor of biomedical engineering and the Director of the Laboratory on Innovative Technology in Medical Ultrasound (LITMUS). He is also the Director of the Natural Sciences and Engineering Research Council of Canada (NSERC) Collaborative Research Training Program on “Next-Generation Innovations in Ultrasomics” in Canada. He has long-standing research interests in ultrasound imaging and therapeutics.

Dr. Yu has been named as a Stacie Fellow by NSERC. He is a fellow of the American Institute of Ultrasound in Medicine. Earlier in his career, he was a recipient of the IEEE Ultrasonics Early Career Investigator Award, the Frederic Lizzi Award, and the Ontario Early Researcher Award. He was the Past Chair of the Medical Ultrasound Group of the IEEE Ultrasonics Symposium and was an Elected AdCom Member of the IEEE UFFC Society. He is currently the Editor-in-Chief of the IEEE TRANSACTIONS ON ULTRASONICS, FERROELECTRICS, AND FREQUENCY CONTROL.