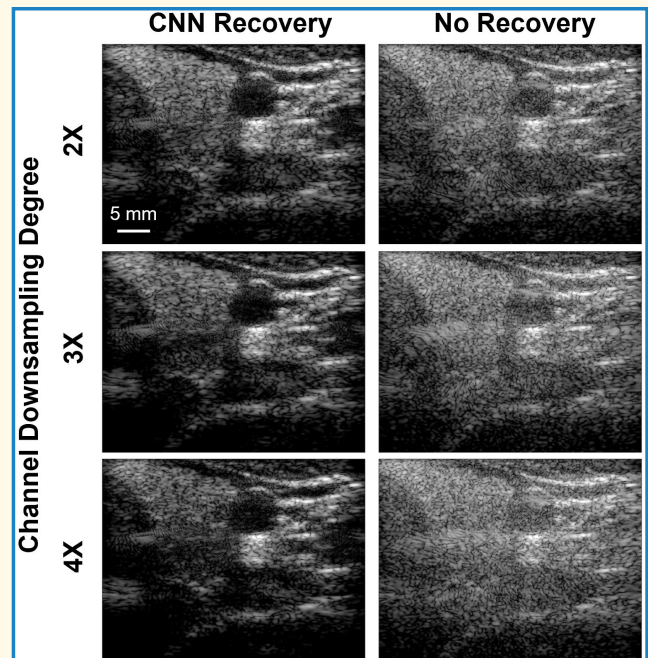# Branched Convolutional Neural Networks for Receiver Channel Recovery in High-Frame-Rate Sparse-Array Ultrasound Imaging

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

Abstract—For high-frame-rate ultrasound imaging, it remains challenging to implement on compact systems as a sparse imaging configuration with limited array channels. One key issue is that the resulting image quality is known to be mediocre not only because unfocused plane-wave excitations are used but also because grating lobes would emerge in sparse-array configurations. In this article, we present the design and use of a new channel recovery framework to infer full-array plane-wave channel datasets for periodically sparse arrays that operate with as few as one-quarter of the full-array aperture. This framework is based on a branched encoder–decoder convolutional neural network (CNN) architecture, which was trained using full-array plane-wave channel data collected from human carotid arteries (59 864 training acquisitions; 5-MHz imaging frequency; 20-MHz sampling rate; plane-wave steering angles between −15° and 15° in 1° increments). Three branched encoder–decoder CNNs were separately trained to recover missing channels after differing degrees of channelwise downsampling (2, 3, and 4 times). The framework's performance was tested on full-array and downsampled plane-wave channel data acquired from an in vitro point target, human carotid arteries, and human brachioradialis muscle. Results show that when inferred full-array plane-wave channel data were used for beamforming, spatial aliasing artifacts in the B-mode images were suppressed for all degrees of channel downsampling. In addition, the image contrast was enhanced compared with B-mode images obtained from beamforming with downsampled channel data. When the recovery framework was implemented on an RTX-2080 GPU, the three investigated degrees of downsampling all achieved the same inference time of 4 ms. Overall, the proposed framework shows promise in enhancing the quality of high-frame-rate ultrasound images generated using a sparse-array imaging setup.

Index Terms— Channel data recovery, convolutional neural network (CNN), high-frame-rate ultrasound, sparse-array imaging.

I. INTRODUCTION

ULTRASOUND is currently in the midst of two promising innovation drives: one toward high-frame-rate ultrasound and another in the direction of more compact, inexpensive systems [1]. High-frame-rate ultrasound uses unfocused transmissions to perform acquisitions with high temporal resolution, and clinical devices that monitor dynamic physiological events with this technology are beginning to be produced [2]. Concurrently, advances in transducer manufacturing, transmit/receive circuitry, and signal processing algorithms have paved the way for the development of more inexpensive and compact ultrasound scanners [3]. This endeavor has led to

Highlights

- **A deep-learning-based channel recovery framework has been devised to infer full-array plane-wave channel data from a sparse array that operates with as few as one-quarter of the full aperture.**

- **B-mode images with limited artifacts were successfully generated in vitro and in vivo from channel datasets inferred from the proposed branched encoder–decoder CNN.**

- **The proposed framework can effectively enhance ultrasound image quality in sparse imaging setups that operate with a reduced number of receiving channels.**

increased uptake of ultrasound not only within hospital settings but also in prehospital, austere, and remote environments [4], [5]. Integration of high-frame-rate ultrasound techniques into these compact and inexpensive systems can enable its more widespread use in the healthcare system, thereby improving access to the imaging paradigm.

High-frame-rate-operable ultrasound systems are difficult to make compact and inexpensive due to their system-level requirements. In particular, receiving radio frequency (RF) data for high-frame-rate acquisitions requires: 1) front-end receiving electronics for all the channels in a system and 2) high bandwidth data links between a system's front-end and back-end [6]. These requirements result in an increase in system complexity and form factor as the number of receiving channels grows. To reduce the hardware complexity of a high-frame-rate system, a possible design choice that can be made is to reduce the number of receiving channels. However, an ultrasound system must include enough receiving channels to span a sufficient aperture for image formation, while keeping the system's pitch small enough to avoid imaging artifacts due to the emergence of grating lobes in the ultrasound field profile.

Several techniques have been proposed to prevent the expected image degradation in ultrasound systems that operate with reduced sets of receiving channels. For instance, microbeamforming can be performed at the transducer front-end to send grouped data from multiple elements together on one channel [7]. Alternatively, channel multiplexing [8], [9] can be used to receive a full set of RF channel data over multiple transmissions. In addition, sparse arrays with optimized layouts [10], specialized image formation algorithms [11], [12], [13], [14], and postbeamforming artifact reduction methods [15] can be used to form high-frame-rate ultrasound images with fewer receiving channels than a fully populated array. Nevertheless, these methods are inherently incompatible with imaging algorithms that operate on prebeamforming RF channel data. For the task of raw RF recovery in high-frame-rate imaging scenarios, there has been active research aimed at recovering RF signals after subsampling [16], [17]. However, the existing methods are not designed for the task of channel data recovery after channelwise subsampling, so they cannot be applied to a system with a reduced number of receiving channels.

Methods that can enable high-frame-rate receiver channel reduction with capability to recover missing raw RF data would help integrate high-frame-rate techniques into compact ultrasound systems. Indeed, our group has previously developed a deep-learning-based RF recovery system that can recover a full set of plane-wave RF channel data from only half of receiving channels [18]. This framework can enable a system to operate with fewer receiving channels, while providing access to a full set of data during ultrasound image formation. However, it uses a static convolutional neural network (CNN) architecture that is not applicable when less than half of receiving channels are available. Furthermore, extension of the architecture toward higher degrees of channel reduction is a nontrivial task due to: 1) the spatial nonuniformity of missing channels at downsampling degrees beyond two times ($2\times$) and 2) the increase in recovery difficulty associated with highly sparse receiving scenarios. These challenges also apply to other deep-learning-based high-frame-rate channel recovery frameworks [19]. As a potential alternate solution, compressed-sensing-based channel recovery techniques may be extended to any degree of channel reduction. However, when applied to high-frame-rate synthetic aperture data, these techniques show in vitro image degradation at channelwise downsampling degrees beyond $2\times$ and there is no demonstrated feasibility for higher degrees of channel recovery with in vivo data [20], [21]. Despite these challenges, techniques that allow higher degrees of RF recovery are desired, as they can be implemented in systems with greater improvements in portability and cost. For example, a downsized system such as the US4R-Lite (us4us Ltd., Warsaw, Poland) would benefit from a framework that is capable of higher degrees of channel data recovery, as it receives on one-quarter of available channels for a given transmission [22]. Accordingly, there is a need for additional innovation in RF recovery to enable its actualization in compact and inexpensive high-frame-rate ultrasound platforms.

This article presents a novel computational solution for facilitating plane-wave channel data recovery at downsampling degrees beyond $2\times$. Our solution is based on the design of a branching encoder–decoder CNN architecture that can be trained to infer the RF data of decimated array channels in scenarios with less than half of the array channels in operation. We hypothesize that the branching encoder–decoder CNN architecture can leverage similarities in the time-delayed reflections that each channel receives to infer missing channel data from downsampled subsets. Given a uniformly downsampled set of input channels, each branch of the recovery framework outputs an equal number of uniformly downsampled output channels; the number of output branches can
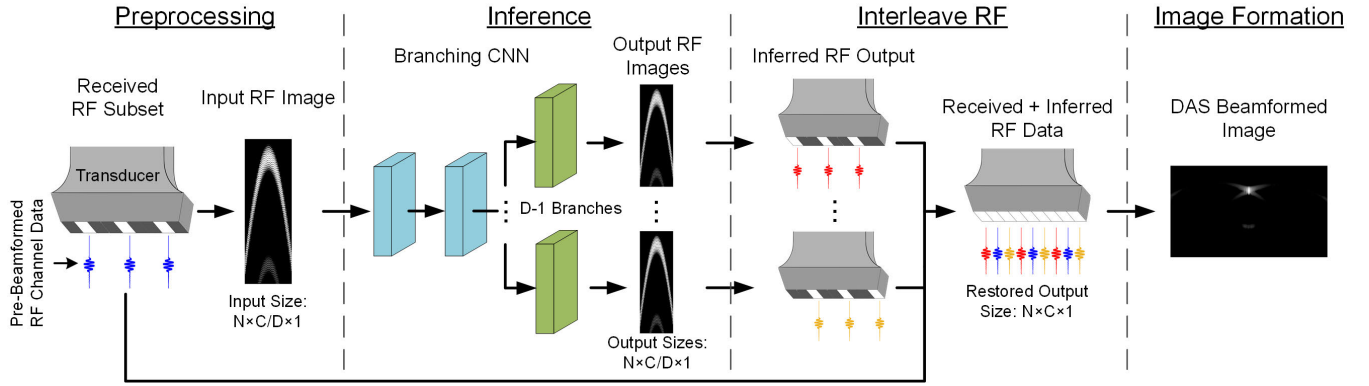
Fig. 1. Proposed recovery framework. Downsampled prebeamformed RF subsets are placed into an RF image and fed into a CNN. Outputs from the network correspond to offset subsets of prebeamformed RF data, which are interleaved with the network input to recover a full set of RF data. This recovered set of RF data can then be beamformed using standard DAS beamforming. In this figure, the RF images and the DAS beamformed image are logarithmically scaled for visualization purposes.

then be increased to accommodate higher degrees of downsampling. This work is readily distinguished from previous attempts at high-frame-rate RF channel recovery. Specifically, it represents the first demonstration of the feasibility of in vivo RF recovery at channelwise downsampling rates beyond $2\times$. It also is the first investigation to introduce a deep learning architecture that can be generalized to any degree of channelwise downsampling.

II. THEORETICAL PRINCIPLES

A. CNNs for RF Recovery

A CNN-based approach was taken to enable RF recovery from channelwise downsampled sets of RF data. Direct downsampling of a system's receiving channels will cause spatial aliasing if the sampling pitch of the array elements exceeds half an acoustic wavelength $\lambda/2$ [23], causing spatial aliasing artifacts in a beamformed image. This loss of image clarity is a product of the beamforming process, and the information losses that channel reduction imposes on the raw RF data may be less severe. Neighboring channels should receive similar time-delayed signals from a given reflector during an imaging event, resulting in shared information between channels that can be used to predict the received signals from omitted channels [18]. CNNs are capable of learning from this type of spatiotemporal data [24], and we have previously shown that an encoder–decoder architecture can effectively produce omitted RF channel data from half of a received set [18].

To facilitate larger degrees of RF inference, we have developed novel CNN architectures that contain a more complex encoding stage followed by a branched decoding segment that produces multiple outputs. These networks encode the received RF subset into a compact feature representation, and then decode this compressed representation into additional sets of RF data. As will be explained in Sections II-B and II-C, the branching output is easily generalized to multiple levels of channelwise downsampling, and the CNN architecture is designed to be capable of RF recovery from high degrees of downsampling.

B. Generalized Uniform RF Recovery Framework

A branched recovery scheme was constructed to facilitate RF recovery from multiple levels of uniform downsampling. The overall framework is shown in Fig. 1, where N corresponds to the number of samples received on each channel, C is the number of channels in the full receiver array, and D is the degree of channelwise downsampling. First, a set of uniformly downsampled, prebeamformed RF data from a steered plane-wave transmission are placed into an $N \times (C/D) \times 1$ matrix. In this matrix, each column contains the RF data received from a given channel. This arrangement can be viewed from an image processing perspective and will be referred to as an RF image. After this preprocessing step, the input RF image is passed into an encoder–decoder CNN that branches to provide sets of output RF images, where each output RF image corresponds to a missing set of RF channels. Each set of RF channels output from a given branch has equal pitch to the original input set, but they correspond to a missing set of channels that is offset from the original input. The number of branches is dependent on D, where $D - 1$ branches are needed to recover a full set of RF data. The framework's number of branches can be easily adjusted to accommodate different degrees of uniform downsampling, where the only requirement is that the full number of channels C is divisible by the desired downsampling degree D. When this requirement is not met, extra channels can be omitted from the recovery process and optionally added back during the framework's interleave step. After inference, the branched output sets of RF data can be interleaved together to produce the full set of RF data. From here, the RF data can be used for any desired image formation or analysis. In this work, the focus is on image formation with delay-and-sum (DAS) beamforming and coherent plane-wave compounding.

C. Branching Encoder–Decoder CNN Architecture

By performing the RF inference step, the branching encoder–decoder CNN is the key component that enables operation of the RF recovery framework. A detailed diagram of this CNN architecture is given in Fig. 2, and the key features of the network are described in Sections II-C1–II-C3.
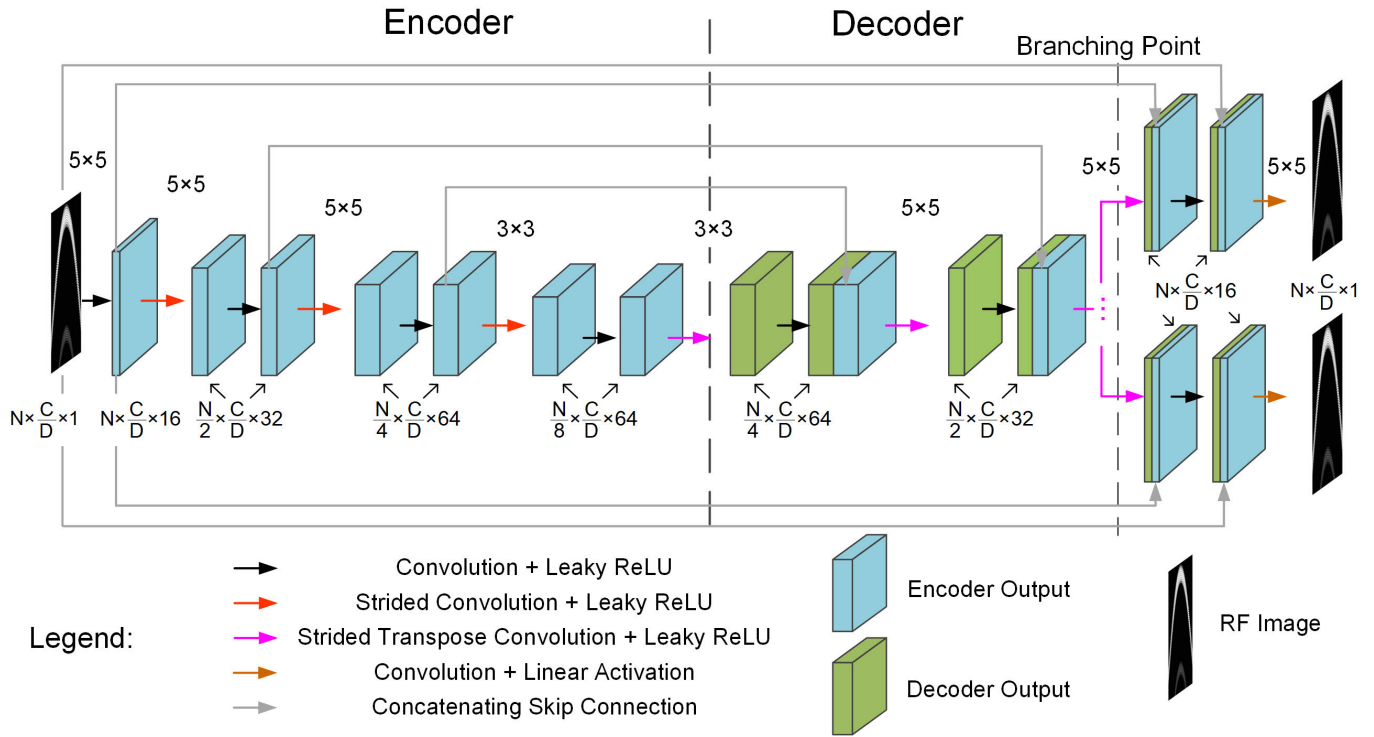
Fig. 2. Encoder–decoder architecture used for RF inference. Convolutional operations, activations, and concatenations are shown by the color-coded arrows. Filter sizes used for convolutions are given above the operation. If there is no filter size given, the filter size from the previous operation is used. Feature maps are represented by blocks, with their dimensions indicated below each block (with an exception in the branched region where the label is between branches). Feature map dimensions for concatenated maps refer to the convolutional output only. Note that the RF images are logarithmically scaled for visualization purposes.

1) Network Input—Downsampled RF Image: The RF image input to the CNN only contains the received RF channels, arranged side by side. If downsampled RF channels are selected with a uniform sampling scheme, each column is separated by an equal pitch and there is no need to encode information on missing channel location with additional columns in the RF image.

2) Network Depth, Filters, and Nonlinearity: The depth, filter sizes, and filter depths of the architecture were chosen to balance computational cost and network complexity. The encoder segment for each network uses seven layers, with strided convolutions used to compress the features in the vertical direction (along each channel). The earlier layers use a filter size of 5 × 5 to capture a large receptive field [25] early, and later layers use a filter size of 3 × 3 to reduce computational cost. This architecture results in a lateral receptive field of 25 channels at the output of each network's encoder. By design, the transducer coverage of the CNN's receptive field naturally increases with greater degrees of downsampling. This increased coverage is caused by the increased effective pitch of the subsampled array at greater degrees of downsampling. If RF data are being inferred for a 128-element probe, this means that the compressed features output by an encoder segment have a receptive field that covers 38% (49/128 elements) of the transducer width for 2× downsampling, 58% (74/128 elements) for 3× downsampling, and 77% (99/128 elements) for 4× downsampling. The expanding coverage of the receptive field acts to bolster a network's inference ability in the face of larger degrees

of downsampling. To increase the complexity of features learned throughout the networks, nonlinear activations are used after convolutional operations, using leaky rectified linear unit (leaky ReLU) activations [26] with $\alpha = 0.01$. Feature depth is grown from 1 to 64 throughout the encoder segment to gradually increase the number of learned features alongside their complexity. After encoding, the decoder segments of each network use an additional seven layers to infer RF sets from these encoded features. Filter sizes/depths are constructed in a pattern that mirrors the encoder, and strided transpose convolutions are used to upsample the network in the vertical direction.

3) Network Outputs, Branching Decoders, and Feature Sharing: A branching scheme is used in the decoder segment to: 1) output RF sets with the same dimensions as the input and 2) provide each output RF set with specialized upsampling and inference filters. With the branching scheme, the path from the input RF set to an individual branch's output RF set forms a symmetrical encoder–decoder CNN. Orientation of the CNN in this manner enables the inputs to the CNN to only include the downsampled channels, while also enabling symmetrical sharing of encoder/decoder features via concatenating skip connections. This feature sharing restores some of the information lost in the encoding scheme of the network, and it also promotes more stable training by enabling direct pathways for the gradient back through the network [27]. The network's branching point is placed midway through the decoder segment to enable feature sharing between outputs during the first half of the decoder while still allowing each

TABLE I
DATA ACQUISITION PARAMETERS

Parameter	Details
Ultrasound Scanner	SonixTouch
Ultrasound Probe	L14-5
RF Data Range and Resolution	-2,048 to 2,047, 12 bits
Number of Tx/Rx Channels	128
Array Pitch	0.3048 mm
Transmit Frequency	5 MHz
Transmit Angle Range	-15°:1°:15°, -0.5° substitute for 0°
Sampling Rate	20 MHz
Imaging Depth	60 mm
Pulse Repetition Frequency	10 kHz

individual output RF set to be decoded with its own set of specialized upsampling/inference filters. In accordance with the overall framework, the number of branches is dependent on the downsampling degree and is given by $D - 1$. At the end of each branch, the output sets of RF data are attained through a final convolution followed by a linear activation, allowing positive and negative RF values as outputs.

III. EXPERIMENTAL METHODS

To effectively evaluate the recovery framework at multiple levels of channelwise downsampling, CNN architectures were created for downsampling levels of $2\times$, $3\times$, and $4\times$ ($D = 2$, 3, and 4). This involved the creation of CNNs with one, two, and three branches, respectively. In contrast to our previous work [18], the CNNs were trained using data that were subject to a cleaning process, and the mean-absolute-error (MAE) loss function was used instead of the mean-squared error (MSE) for optimization. Details related to the training and evaluating of the recovery framework are described as follows.

A. Dataset Acquisition, Cleaning, and Preprocessing

A dataset of in vivo carotid artery scans from a SonixTouch research scanner (SonixTouch; Analogic Ultrasound; Peabody, MA, USA) was used to train the RF recovery architectures. This dataset was acquired for our previous work [18] and it consisted of 67 299 steered plane-wave acquisitions from seven healthy volunteers (age: 25.9 ± 4.9 years). The research scanner was programmed to acquire batches of 31 steered angles from $-15°$ to $15°$, with a $-0.5°$ transmission used instead of $0°$ due to an inability of the scanner to transmit on all the elements simultaneously. A total of 67 299 separate training frames were acquired for the networks, where each frame's RF values ranged between -2048 and 2047 with 12-bit resolution. The acquisition was performed with a 128-element L14-5 probe ($C = 128$); the system operated with a 5-MHz transmission frequency, 2-pulse transmissions, a 10-kHz pulse repetition frequency, and a 20-MHz sampling rate. A summary of acquisition parameters can be found in Table I. Acquired data were transferred to a computer server (SYS-4028-TRT; Super Micro, San Jose, CA, USA) with a Xeon E5-2620 central processing unit (Intel, Santa Clara, CA, USA) to be preprocessed in MATLAB (ver. 2020b; MathWorks, Natick, MA, USA).
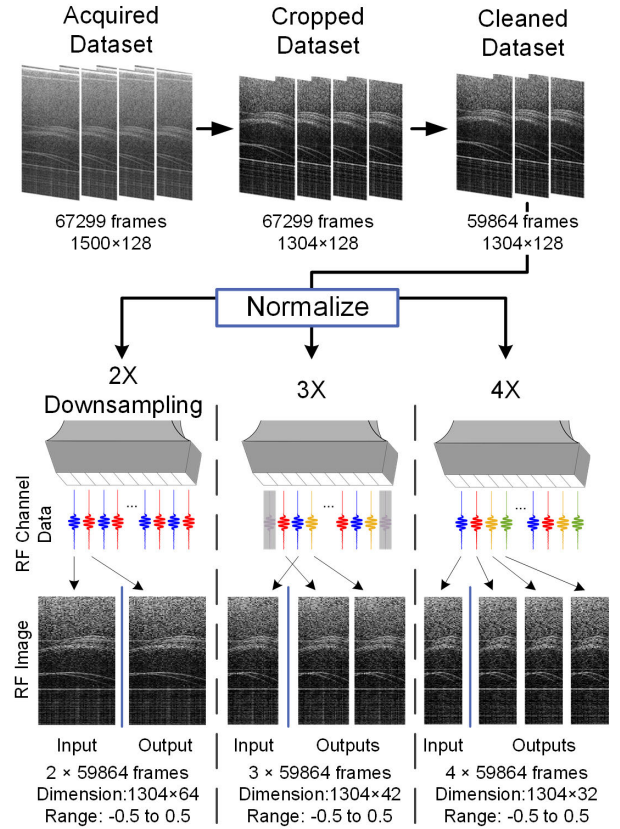

Fig. 3. CNN training dataset preprocessing pipeline. Data were cropped, cleaned to remove heavily clipped frames, normalized, and parsed by channel for different downsampling scenarios. Note that the RF images are logarithmically scaled for visualization purposes.

The acquired dataset was cleaned and preprocessed to facilitate a stable and effective training process, with steps shown in Fig. 3. First, the initial 196 samples of each channel were removed to reduce the number of near-field reflections present in training RF images and to conform the input height to a multiple of eight. Ensuring a multiple of eight on input height allowed the three strided convolutional layers in the network's encoder segment to be performed without the need to zero pad the signal. This cropping resulted in $N = 1304$ samples per channel, where all the training samples covered a depth of 11.4–61.6 mm. Second, the dataset was cleaned by removing RF frames that had an excess of clipped samples. Training frames with more than 50 samples valued at 2047 (the maximum output value of the Sonixtouch's analog-to-digital converter) were removed from the training dataset. The data cleaning resulted in 59 864 training frames being retained (90% of the original acquired set) while reducing the potential negative impact of clipped samples on the training data. This subset of training data was then normalized to be between -0.5 and 0.5 prior to being input into the network.

Network inputs and outputs were formed by selecting uniformly spaced channels from an RF frame and placing them into smaller RF images. For $2\times$ downsampled data, odd-numbered channels were selected as inputs with even channels selected as network outputs (each sized as $N \times C/D = 1304 \times 64$). For $3\times$ downsampled data, three

sets of uniformly spaced channels with two-element separation were formed ($N \times C/D = 1304 \times 42$), with the middle set as a network input and the remaining sets as output. The first and last channels of the full RF set were discarded to ensure a uniform downsampling scheme where inputs and outputs to the network all had the same size ($C = 126$ for $3\times$ downsampling). Finally, for $4\times$ downsampled data, four sets of uniformly spaced channels with three-element separation were formed ($N \times C/D = 1304 \times 32$), with the first set as network inputs and the remaining three sets as network outputs.

B. CNN Training

Networks for each downsampling level were developed and trained in Python (ver. 3.6.7), using TensorFlow-GPU (ver. 1.12.0) with the Keras (ver. 2.1.6) application user interface. The training of the networks was facilitated using a GTX 1080 graphical processing unit (GPU; Nvidia, Santa Clara, CA, USA). Network weights were initialized according to a He-uniform distribution, and then the Adam optimization algorithm [28] was used for training; each network used a learning rate of 0.001, a batch size of 32, and 50 total epochs. For training, the MAE was used to reduce the impact that clipped RF samples had on the training process. The MAE from each branch's output was added together with equal weighting to form the overall loss function. About 90% of the cleaned dataset was used for training with the remaining 10% used for validation of each network. The loss curves from the individual branches for the three investigated degrees of downsampling are shown in Fig. 4. The validation loss closely follows the training loss for all the branches, and the overall loss in different branches increases as the downsampling degree is increased. The comparatively higher loss for branch 2 (channels 3, 7, 11, …) in the $4\times$ downsampling case occurs because the channels composing this branch are two pitch-lengths away from the input RF (channels 1, 5, 9, …) set, while channels composing branches 1 (channels 2, 6, 10, …) and 3 (channels 4, 8, 12, …) are only one pitch-length from channels on the input RF set.

C. Recovery Performance Assessment

The effectiveness of the RF recovery framework was assessed using both raw RF signal analysis and quality comparisons of DAS beamformed images. For image analysis, images were beamformed with: 1) only the downsampled subset of RF channels; 2) the downsampled RF channels and CNN-inferred RF channels; and 3) the full original set of RF channels. The image quality improvement provided by CNN-inferred RF data was then assessed by comparing images that were beamformed from each of these three sets of channels.

1) Evaluation Scenarios: Additional RF datasets were acquired from several different imaging scenarios to evaluate the recovery framework's success. First, nine additional carotid/thyroid scans from nine volunteers separate to the training set were acquired to evaluate RF reconstruction success for different in vivo tissue types, namely, the hyperechogenic
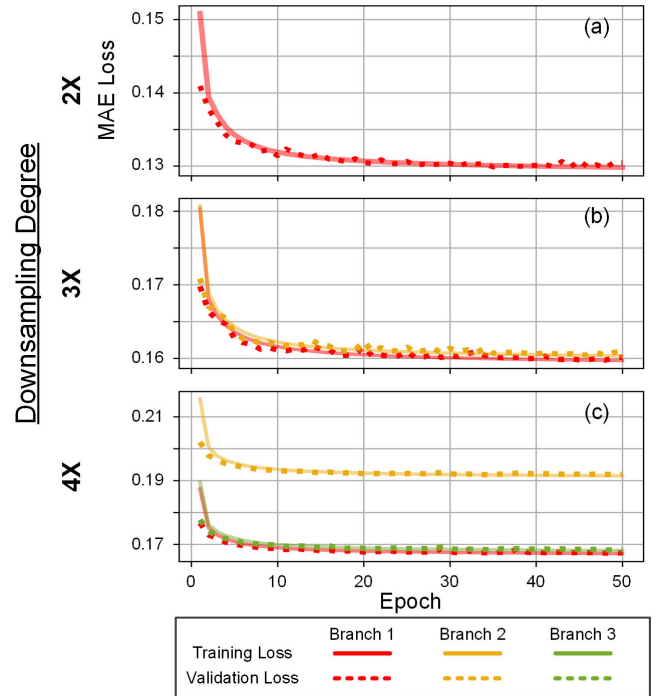

Fig. 4. Loss curves for each branch at the three investigated degrees of downsampling. (a) Loss curve for the $2\times$ downsampling case. (b) Loss curves for each branch in the $3\times$ downsampling case. (c) Loss curves for each branch in the $4\times$ downsampling case.

carotid wall and the homogeneous, less echogenic thyroid. These in vivo datasets were acquired with approval from the Clinical Research Ethics Committee of the University of Waterloo (Protocol No. 31694). Second, an in vitro scan of a point target phantom was acquired for an interpretable evaluation of RF inference success along a hyperbola. Finally, an in vivo plane-wave brachioradialis acquisition from the Challenge on Ultrasound Beamforming with Deep Learning (CUBDL) [29] Task 1 data was used to test the generalizability of the network to a new imaging scenario and a new ultrasound system. All the acquisitions were taken using imaging parameters from Table I, with the exception of the CUBDL acquisition, which used a Verasonics Vantage 256 with a transmit frequency of 7.5 MHz and a sampling frequency of 25 MHz [29]. All the scenarios were evaluated at downsampling/recovery levels of $2\times$, $3\times$, and $4\times$.

2) Evaluation Data Preparation: The testing RF data were cropped to have a length of 1304 to ensure they could be fed into the trained CNN. Each set of RF data was cropped to ensure that the beamformed B-mode image contained regions of interest for image quality assessment. RF data were normalized and partitioned in the same manner as the training data, and input RF images were then fed into the trained networks to recover a full set of RF data for each level of downsampling.

3) Image Formation Pipeline: For subsequent image analysis, RF data were first bandpass-filtered between 3 and 7 MHz and converted into an analytic signal with the Hilbert transform. The analytic data were then DAS beamformed with an F# of 1 and rectangular apodization, and envelope detection was performed by taking the absolute value of the

TABLE II
IMAGE BEAMFORMING PARAMETERS

Parameter	Details
Prefilter Passband	3 MHz – 7 MHz
Filter Design Method	Equiripple (30th order)
Apodization	Rectangular
Focusing	Constant $F_\# = 1$
Image Resolution (Axial and Lateral)	0.1 mm/pixel

beamformed analytic signal. The final step in the ultrasound image formation procedure was to then logarithmically scale the beamformed envelope values. The key parameters from the beamforming process are summarized in Table II.

4) Beamformed Image Quality Metrics: Beamformed images were assessed through a combination of contrast evaluation and full-reference image quality measurement. The contrast ratio (CR) and the generalized contrast-to-noise ratio (gCNR) [30] of the carotid artery lumen to the surrounding tissue regions were taken to evaluate the reduction in spatial aliasing artifacts. For each gCNR calculation, probability density functions were estimated using a 256-bin histogram that covered the image's full dynamic range. To evaluate the overall image quality restoration with a full reference metric, the structural similarity measure (SSIM) [31] was used. For SSIM calculation, pixel values from the DAS operation were logarithmically scaled and clipped to values within the displayed dynamic range. Images beamformed with a full set of receiving RF data were used as reference. As performed in the original SSIM paper, windows of size 11×11 were used for SSIM window calculations [31], and the overall SSIM was calculated using the entire beamformed image.

5) RF Characterization and Recovery Metrics: The root mean squared (rms) and normalized root MSE (NRMSE) were used to compare RF data characteristics and RF data recovery success for different tissue types. The rms allows for evaluation of overall RF magnitude when characterizing reflections from different regions of interest. In addition, the NRMSE metric allows RF recovery comparison between multiple tissue types, despite potentially differing overall amplitudes due to each tissue's varying depth and echogenicity characteristics.

6) Compounded Image Quality Evaluation: The compatibility of the recovery framework with the plane-wave compounding process was evaluated by tracking image metric changes as beamformed images were compounded. Images were compounded by adding $1°$ transmissions sequentially, starting from the centermost $-0.5°$ acquisition. Therefore, the compounding pattern went as follows: one-angle: $-0.5°$; two-angle: $-0.5°$, $1°$; three-angle: $-1°$, $-0.5°$, $1°$; four-angle: $-1°$, $-0.5°$, $1°$, $2°$; etc.

D. Inference Speed Evaluation

To evaluate the inference speed of RF channel inference, the recovery framework was timed using Python's time module. The same system as described in Section III-B was used, except that an RTX-2080 GPU (Nvidia) was used for inference instead of the GTX-1080 GPU. Total inference time

over 2500 RF frames was calculated and averaged for the three investigated degrees of downsampling.

IV. RESULTS

A. Image Structure Is Recovered When CNN-Inferred RF Data Are Used for Beamforming

Images beamformed with CNN-inferred RF data showed suppression of spatial aliasing artifacts in multiple imaging scenarios and at multiple levels of downsampling. As shown in Fig. 5, the underlying structures in beamformed images of a $-0.5°$ point target phantom acquisition [Fig. 5(a)], a $-0.5°$ carotid artery/thyroid acquisition [Fig. 5(b)] and a $0°$ brachioradialis muscle acquisition [Fig. 5(c)] became harder to discern when channels are removed during the receiving process [Fig. 5(d)–(f)]. Only the $3\times$ downsampling case is shown for the sake of brevity, but the spatial aliasing artifacts would be less obtrusive for the $2\times$ case and more prominent in the $4\times$ case. Highlighted by the point target image Fig. 5(d), the most obtrusive artifacts are caused by the strongly echogenic structures in the image. Beamforming with CNN-inferred RF data resulted in a reduction in these artifacts at all the levels of downsampling [Fig. 5(g)–(o)]. This artifact reduction revealed the underlying image structure, as the point targets are visible and isolated, the carotid lumen is revealed, and the brachioradialis muscle fibers can be distinguished. The artifact reduction was accompanied by a slight degradation in image quality, quantified with the reduced SSIM values as downsampling degrees are increased. In addition to the improvement in image quality, the recovery framework achieved an average inference speed of 4 ms for the three investigated degrees of downsampling, corresponding to an average speed of 250 frames/s.

B. Inclusion of CNN-Inferred RF Data Increases Contrast of Beamformed Images

The suppression of spatial aliasing artifacts resulted in an increased contrast of the carotid artery when CNN-inferred RF data were used for beamforming. Using the hand-segmented reference regions on the carotid artery given in Fig. 6(a), the single-transmit images beamformed from 31 steered plane-wave acquisitions experienced CR improvement when CNN-inferred RF data were used [Fig. 6(b) and (c)] across all the transmit steering angles. The CR was improved when both the hyperechogenic carotid wall was used as a reference and when the less echogenic, homogeneous thyroid was used as a reference. Relatively higher recovery in contrast was achieved when the hyperechogenic carotid wall was used as a reference, and this is reflected in the images of Fig. 5(h), (k), and (n), as some thyroid content is lost at higher levels of downsampling. The gCNR of the images followed a similar trend when the carotid wall was used as a reference [Fig. 6(c)], but when the thyroid was used as a reference the downsampled cases see higher gCNR than the fully sampled and recovered cases for some positive transmission angles [Fig. 6(d)]. These higher gCNR values in the purely downsampled acquisitions can be explained by the increasingly negative CR values for these transmission angles [Fig. 6(c)], since the gCNR will increase
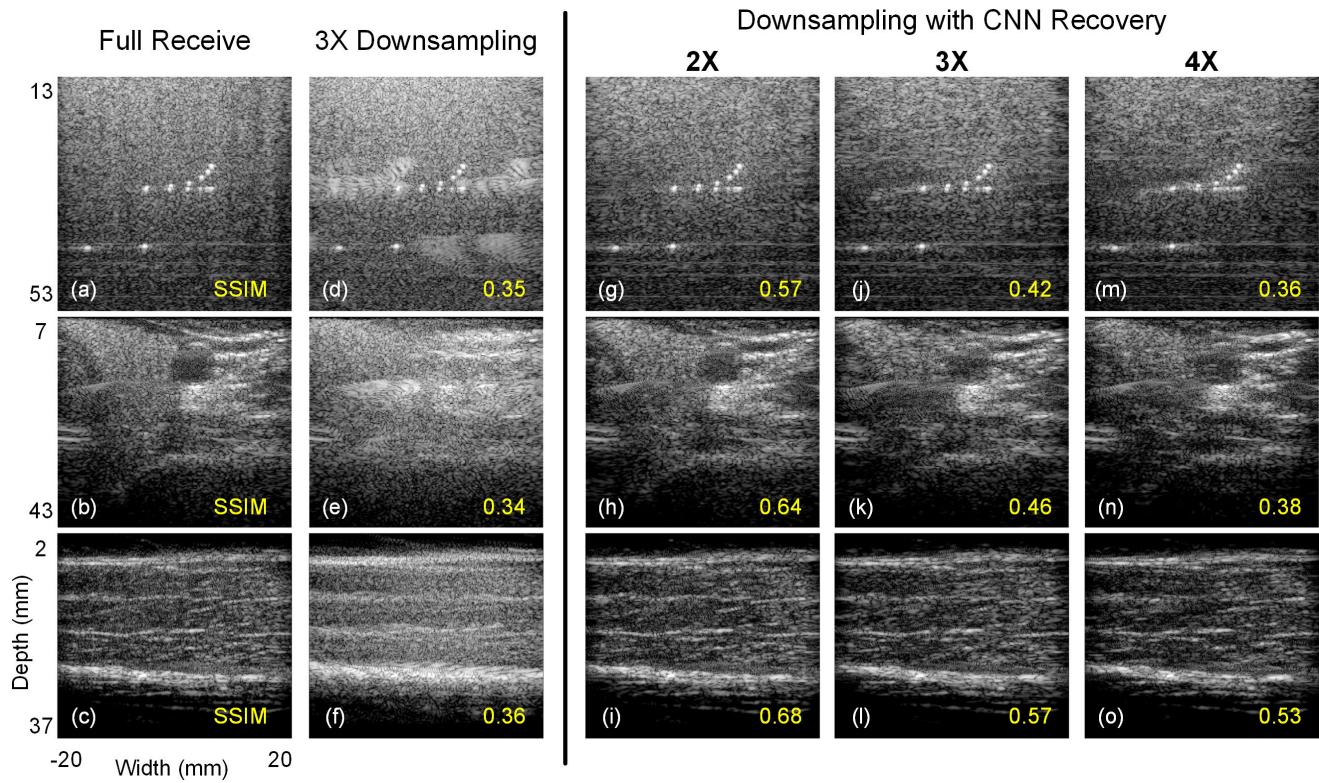
Fig. 5.   Single-angle B-mode images. All the images are displayed with a dynamic range of 50 dB. (a)–(c) B-mode images beamformed with all the original receiving channels for a −0.5° point target phantom acquisition, a −0.5° carotid artery acquisition, and a 0° brachioradialis acquisition, respectively. (d)–(f) B-mode images beamformed with only 1/3 of receiving channels. (g)–(i) B-mode images beamformed with 1/2 original channels + 1/2 inferred channels. (j)–(l) B-mode images beamformed with 1/3 of receiving channels + 2/3 inferred channels. (m)–(o) B-mode images beamformed with 1/4 of receiving channels + 3/4 inferred channels.
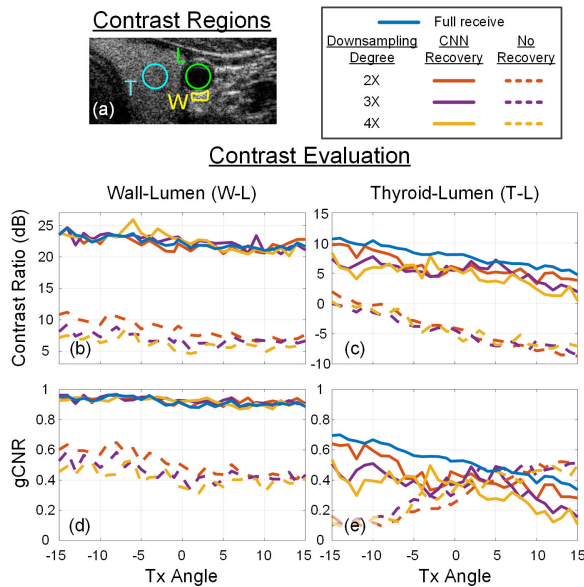

Fig. 6.   Contrast evaluation of a carotid artery over multiple transmission angles. (a) Segmented regions for contrast assessment. A 31-angle compounded image was used to choose reference regions. (b) and (c) CR evaluations using the carotid wall and thyroid as a tissue reference points, respectively. (d) and (e) gCNR evaluations using the carotid wall and thyroid as tissue reference points, respectively.

as the separability between tissue regions increases, regardless of the direction of separation.

Evaluation of the carotid CR and gCNR on the rest of the acquisitions in the test set indicated a similar trend to

the acquisition in Fig. 6. Fig. 7 shows box plots for the CR [Fig. 7(a) and (b)] and gCNR [Fig. 7(c) and (d)] of 31-angle carotid acquisitions from nine separate volunteers (9 × 31 = 279 total transmissions). Relatively higher recovery in CR and gCNR was achieved when the carotid wall was used as reference compared with the thyroid at all the downsampling degrees. Similar to Figs. 6(d) and 7(d) is skewed to have higher gCNR for the downsampling cases due to the increasingly negative CR values [Fig. 7(b)].

C. Successful RF Inference in Hyperbolic Regions of RF Images

Relatively higher RF recovery success was seen across all the downsampling degrees when hyperechogenic data were being inferred. The hyperechogenicity of the carotid wall in region 1 of Fig. 8(a) manifested itself in the hyperbolic reflections in region 1 of Fig. 8(b). Conversely, the homogeneous, less echogenic thyroid in region 3 of Fig. 8(a) did not provide hyperbolic structure in region 3 of Fig. 8(b). Consequently, the NRMSE values [Fig. 8(c)] at each level of downsampling were >2× lower for inferred RF from the hyperechogenic carotid region compared with the homogeneous, less echogenic thyroid region.

The accuracy of inference was not directly dependent on higher amplitudes that are associated with hyperechogenic reflections, but the hyperbolic RF structure that is due to the scattering properties of these regions. Reflections from region
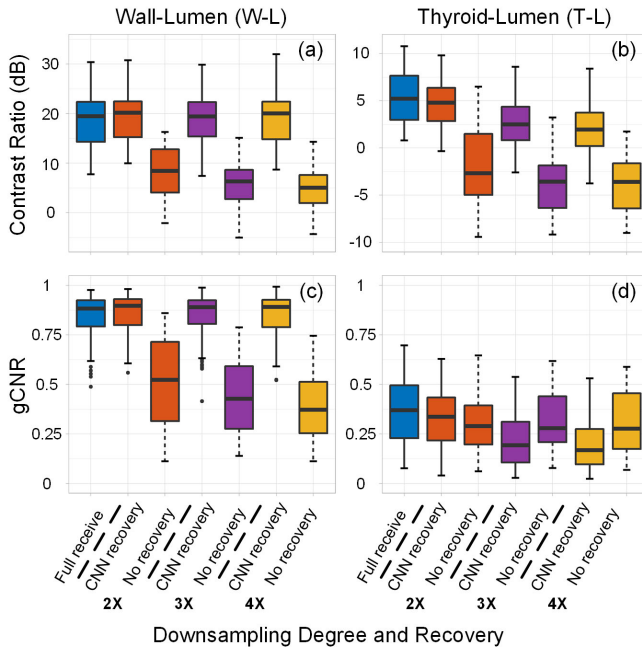
Fig. 7. Contrast evaluation of multiple carotid arteries over multiple transmission angles. (a) and (b) CR evaluations using the carotid wall and thyroid as tissue reference points, respectively. (c) and (d) gCNR evaluations using the carotid wall and thyroid as tissue reference points, respectively.
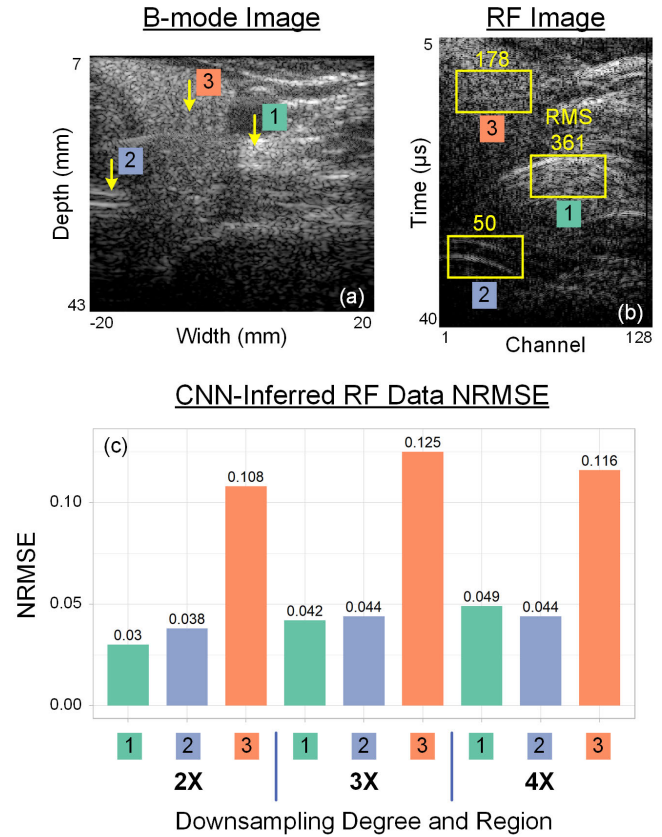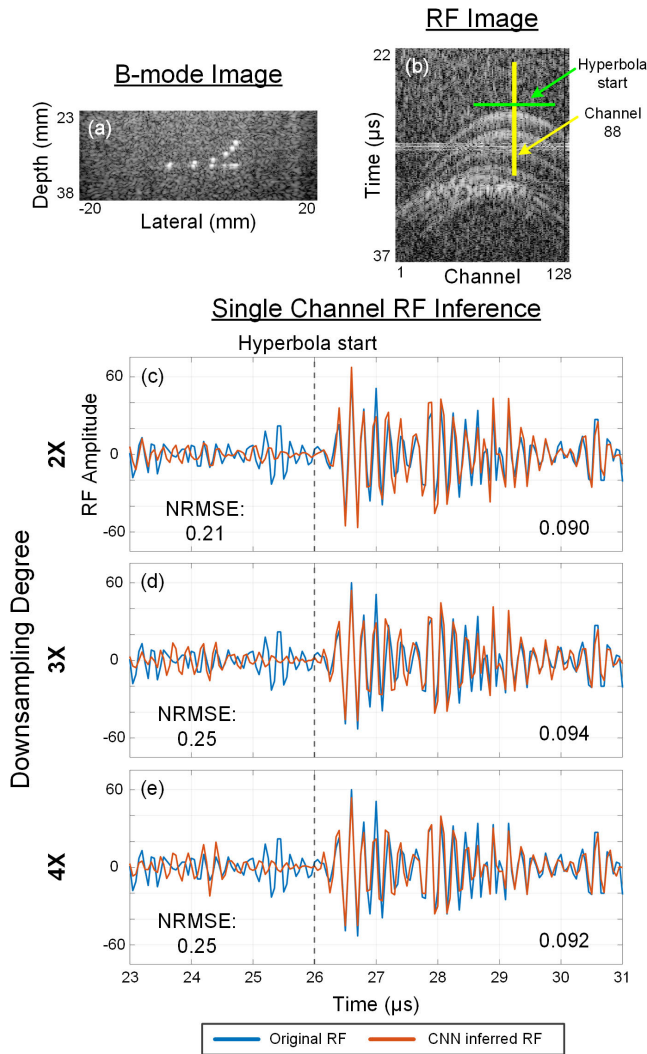

Fig. 8. In vivo RF reconstruction evaluation for different regions of a carotid artery. (a) Highlighted tissue regions on a B-mode image. 1 is the hyperechogenic carotid wall, 2 is a lower amplitude hyperechogenic region, and 3 is the homogeneous, less echogenic thyroid. The B-mode image is displayed with a dynamic range of 50 dB. (b) RF reflections from each highlighted tissue region. The region # is shown below its yellow bounding box, and the rms of the region is shown above the box. The RF image is logarithmically scaled and displayed with a dynamic range of 40 dB. (c) NRMSE for the CNN-inferred RF data from each region.

2 had lower rms values compared with region 3, but they still experienced a considerably lower NRMSE for inference compared with region 3. This success can be associated with the low-amplitude hyperbolic reflections that are clearly visible in region 2 of Fig. 8(b). The scattering properties of the region's reflectors resulted in a clear hyperbola in the RF image, and consequently, there was a NRMSE that is $>2\times$ lower than the homogeneous, less echogenic thyroid (region 3) for each level of downsampling.

Relatively higher RF inference accuracy was also achieved on hyperbolic tissue types that were different from the CNNs' training set. The examination of a single channel's inference from an in vitro point target phantom is shown in Fig. 9, where a $>2\times$ reduction in NRMSE was seen in the region that contains hyperbolas from the point target. This distinction was present at all the levels of downsampling. The recovery framework was not trained on any in vitro acquisitions, but the trend of higher hyperechogenic RF recovery success was present for this type of data as well.

D. Coherent Compounding of Images Beamformed With CNN-Inferred RF Data Further Improves Image Quality

The contrast and SSIM of images beamformed with CNN-inferred RF data were further improved with coherent plane-wave compounding. The compounding process of the carotid artery in Fig. 5(b) is shown in Fig. 10, where compounding resulted in a progressive improvement of contrast and SSIM for all the beamforming scenarios. Similar to the single-angle case, the inclusion of CNN-inferred data in the beamforming process enabled improved image quality compared with when only the subset of receiving channels

were used. First, both the CR and gCNR between the lumen and carotid wall were enhanced beyond the case with all the receiving channels when CNN-inferred RF data were used in beamforming. Second, both the CR and gCNR between the thyroid and the lumen were consistently improved when CNN-inferred data were included during beamforming, despite the less hyperbolic RF data provided from thyroid reflections. Finally, while the SSIM of all the beamformed images improved at higher degrees of compounding, higher resultant SSIMs were achieved when CNN-inferred RF data were included during beamforming.

The recovery-attributed enhancement of compounded image quality can be observed in Fig. 11, where seven-angle-compounded images are displayed. While the compounding of images formed with downsampled RF data [Fig. 11(e)–(g)] resulted in a higher quality image compared with the single-transmit case [Fig. 5(e)], the images were still obstructed by spatial aliasing artifacts. The visibility of the carotid structure was enhanced when CNN-inferred data were also used in beamforming [Fig. 11(b)–(d)]. This increased carotid visibility was seen over the carotid test set, as shown
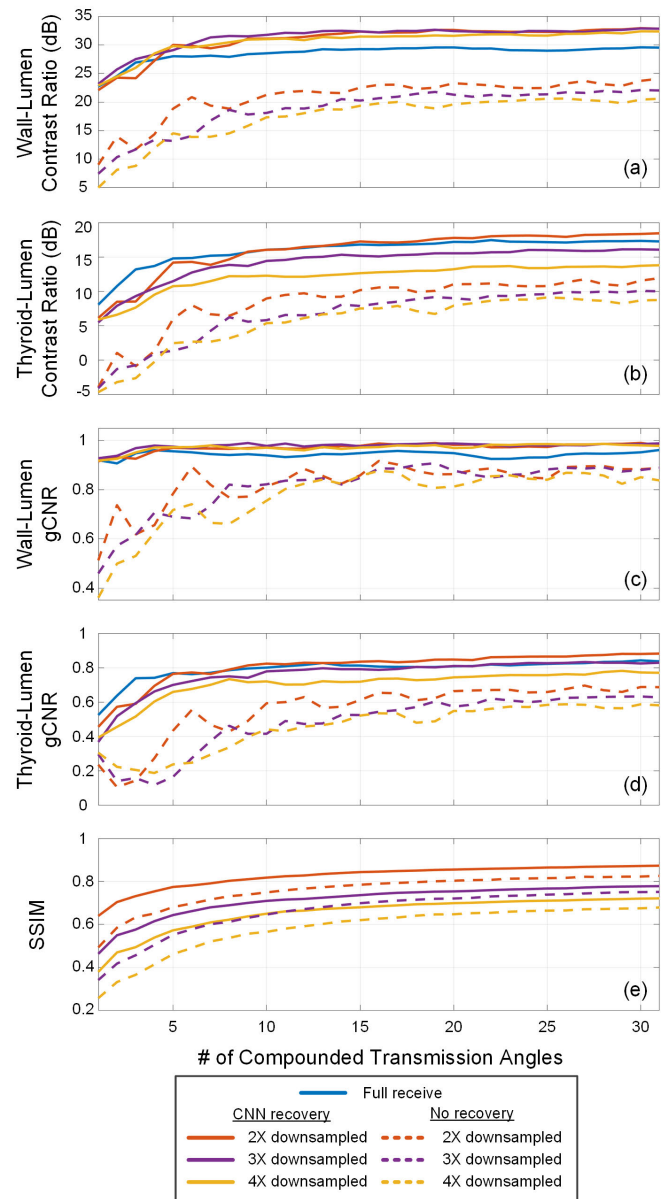
Fig. 9. In vitro RF reconstruction evaluation for a point target phantom. (a) B-mode image of the point targets being examined, displayed with a 50-dB dynamic range. (b) RF image of the point target with the channel examined highlighted in yellow, and the start of the hyperbolic point target reflections denoted in green. The RF image is logarithmically scaled and displayed with a dynamic range of 40 dB. (c)–(e) Comparison of the original RF data to the inferred RF data for downsampling levels of 2, 3, and 4, respectively. The NRMSE is given for the prehyperbolic region and the hyperbolic region.

Fig. 10. Changing carotid artery image quality metrics throughout the compounding process. Regions used for contrast evaluation are highlighted in Fig. 6. (a) Wall to lumen CR. (b) Thyroid to lumen CR. (c) Wall to lumen gCNR. (d) Thyroid to lumen gCNR. (e) SSIM.

by the box plot images that summarize the contrast of the seven-angle compounded images for each carotid in the test set (Fig. 12).

V. DISCUSSION

A. Summary of Contributions

Each receiving channel in an ultrasound system imposes a tangible increase in the system's complexity, whether it is through an increased data transfer bandwidth, or through the requirement of additional receiver electronics such as analog-to-digital converters or low-noise amplifiers. To enable high-frame-rate ultrasound systems with lower channel counts and lower system complexity, we have developed an RF recovery framework (Fig. 1) that can be applied to uniformly channelwise downsampled RF data. Our framework leverages

novel branching encoder–decoder CNN architectures (Fig. 2) to directly recover RF channels that were omitted during the receive process. This novel CNN architecture is an improvement upon the static nature of our previous framework [18], and its generalizability to multiple degrees of downsampling provides higher flexibility for its implementation in high-frame-rate ultrasound systems with differing specifications. Furthermore, the inference time of the framework was found to be held constant at 4 ms regardless of the downsampling degree (Section IV-A). Correspondingly, the proposed framework achieved a throughput speed of 250 frames/s for each of the three investigated degrees of downsampling with a RTX-2080 GPU. Further improvements in the processing throughput can be expected with the use of more advanced microprocessor technology [32].
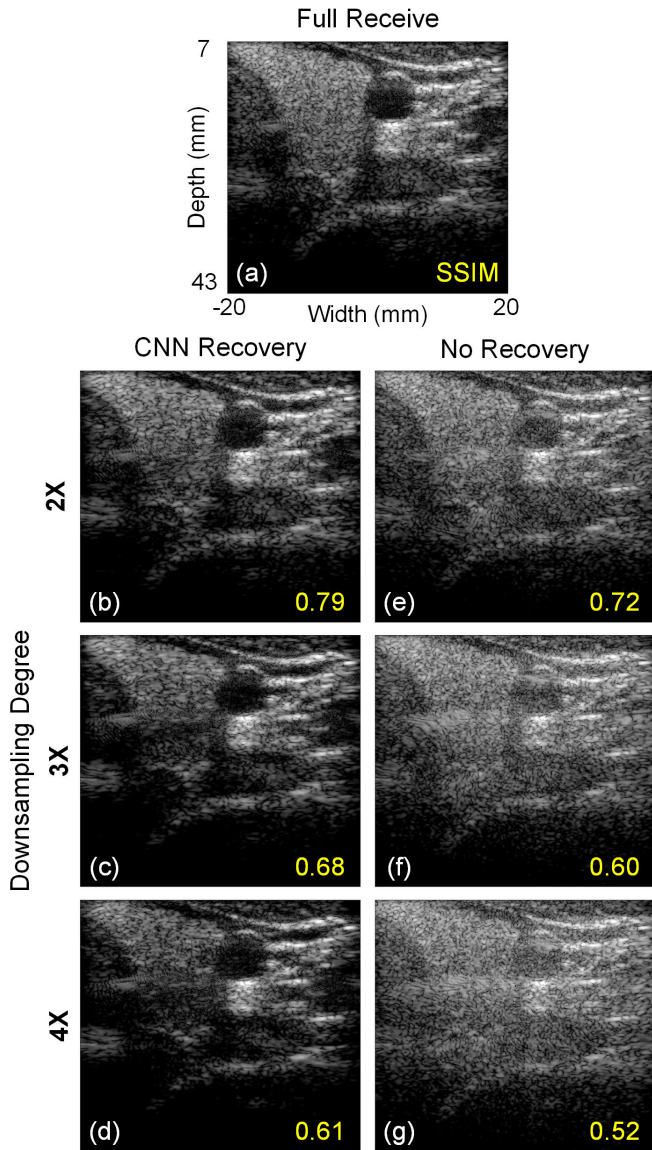
Fig. 11. Seven-angle compounded images of a carotid artery. All the images are displayed with a dynamic range of 50 dB. (a) Beamforming performed with all the receiving channels. (b)–(d) Beamforming performed with 1/2 received + 1/2 inferred channels, 1/3 received + 2/3 inferred channels, and 1/4 received + 3/4 inferred channels, respectively. (e)–(g) Beamforming performed with 1/2 received channels, 1/3 received channels, and 1/4 received channels, respectively.

Note that our CNN's constant inference time may seem counterintuitive since, in principle, the complexity of the architecture is increased due to the addition of decoder branches at higher degrees of downsampling. Nevertheless, the impact of the increased CNN complexity is practically offset because higher degrees of downsampling naturally result in fewer input channels for the CNN's inferencing operations. As such, it is feasible for our branched CNN to restore, with a constant inference time, a full set of RF channel data from a downsampled subset.

Our experiments with DAS beamforming showed that including CNN-inferred channels in the beamforming process improved image quality (Fig. 5). These improvements from
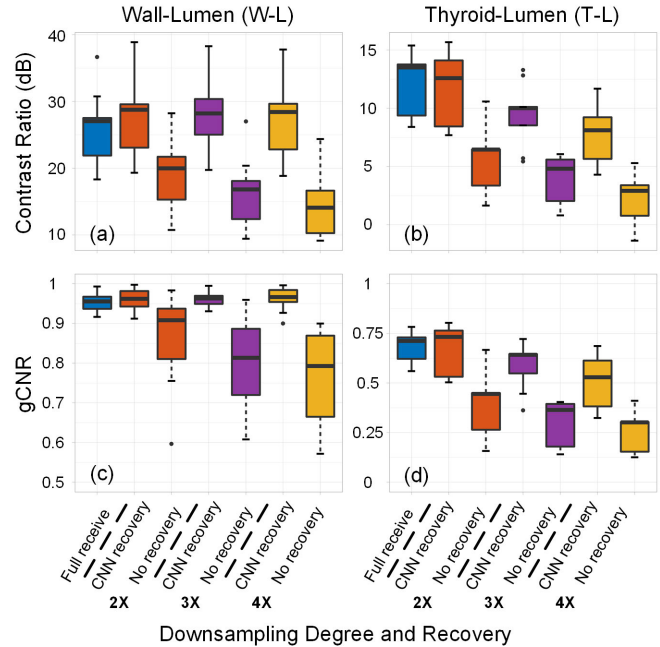
Fig. 12. Contrast evaluation of seven-angle compounded carotid arteries. (a) and (b) CR evaluations using the carotid wall and thyroid as tissue reference points, respectively. (c) and (d) gCNR evaluations using the carotid wall and thyroid as tissue reference points, respectively.

our framework were generalizable to multiple imaging mediums, namely, in vivo carotid arteries, an in vitro point target, and an in vivo brachioradialis muscle. These improvements were also generalizable across two imaging systems, namely, a SonixTouch research scanner and a Verasonics scanner, and they were achieved with different center/sampling frequencies from the training acquisitions. Furthermore, inferred RF channels provided similar improvements when the network's inputs were from varying transmission angles (Figs. 6 and 7). This angle independence of the framework's inputs enables its use with more advanced RF-processing techniques; when inferred RF data from steered transmissions were used for coherent plane-wave compounding, progressive image quality improvement was also achieved (Figs. 10–12). Overall, the DAS improvements yielded by the inferred RF data indicate the proposed framework's feasibility for RF recovery. This feasibility was demonstrated for recovery from downsampling degrees beyond 2×, surpassing the in vivo recovery rates demonstrated in our previous work [18], and by other high-frame-rate channel recovery techniques [19], [20], [21].

B. Spatial Aliasing Artifact Reduction Stems From Strong Hyperbolic RF Inference

The image quality improvements that CNN-inferred RF data provided to DAS beamformed images can be attributed to a reduction in spatial aliasing artifacts. The inhibiting features present in the downsampled images of Fig. 5 are the spatial aliasing artifacts that hide the imaging region's underlying structure. The strong reduction in these aliasing artifacts is expected due to two reasons. First, the most prevalent artifacts stem from insufficient spatial sampling of the

hyperechogenic structures in the imaging medium [highlighted by the point target artifacts in Fig. 5(d)]. Second, reflections from more echogenic scatterers manifest themselves in RF image hyperbolas, enabling higher accuracy in RF inference (Figs. 8 and 9). Therefore, the RF samples provided by the inference framework are expected to be highly effective at suppressing the most prominent spatial aliasing artifacts in a medium. This reduction in aliasing artifacts is visible in the images of Fig. 5, as the spatial aliasing artifacts are minimal in the images beamformed with recovered RF data.

A lack of hyperbolic RF structure results in a more difficult RF inference scenario. The spatial aliasing artifacts are suppressed in Fig. 5(h), (k), and (n); however, there is a notable loss in thyroid content, quantified by the decrease in contrast in Figs. 6 and 7. This loss in image quality can be explained by the difficulty inferring reflections from the nonhyperbolic RF structure of the thyroid (Fig. 8). Less accurate RF inferences will have a less coherent sum during DAS beamforming, and as higher portions of the beamformed samples are provided by inference, the beamformed signal will be more subdued. This effect had a significant impact on SSIM measures, as any inconsistencies in speckle amplitudes caused a large degradation of the SSIM, even if the overall image structure was captured. Due to the different degrees of RF recovery success for different types of tissues, there is not a well-defined limit to the amount of CNN-based recovery that can be achieved, as it will depend on the medium being imaged.

When imaging a more challenging medium with less echogenic scatterers, the quality of beamformed images can be improved with plane-wave compounding (Fig. 10). In addition, images beamformed with CNN-inferred RF data had greater degrees of enhancement compared with those beamformed with only a subset of channels (Figs. 10–12). The image quality improvements observed indicate a general coherence between pixels beamformed with CNN-inferred RF data, allowing progressive image quality improvement through the compounding process.

C. Implications on System Design

Using a CNN-based solution for RF recovery, the proposed framework is well-suited for integration into a compact system's signal processing pipeline. Convolutional operations are easily parallelized using GPUs, and recent innovations have seen significant downsizing of this technology in products such as the NVIDIA Jetson platform [33]. The ability for the GPU to execute convolutional operations complements other known capabilities of the GPU, such as the parallelization of computing tasks in an ultrasound processing pipeline (beamforming being a pertinent example). With these technical advances, the GPU is well-poised as a strong candidate to be included in modern ultrasound scanners [34].

D. Perspectives for Future Work

With feasibility of the RF recovery framework demonstrated within the context of plane-wave acquisitions on 1-D arrays, additional research should be pursued for extending the framework to additional imaging schemes. The appearance of hyperbolic structure in RF images is not exclusive to plane-wave acquisitions; therefore, it is expected that similar results would be seen when applying the proposed framework to other imaging schemes such as synthetic aperture imaging [35]. In addition, the demonstrated feasibility of $4\times$ RF recovery raises the question of whether the framework could be extended toward 2-D matrix arrays, since 2-D sparse arrays typically require a larger reduction in channel count to adequately reduce system complexity [36], [37]. This extension could be through a row/columnwise application of the framework to a downsampled matrix array, or through the use of 3-D convolutional kernels to infer reflections from 3-D hyperbolic structure in RF matrices. Finally, the system-agnostic channelwise RF recovery and cross-system generalizability beyond the examples provided with two imaging systems remains to be fully explored. Techniques such as transfer learning [38] could be used to tune the proposed framework to different acquisition parameters and alternative acquisitions systems.

VI. Conclusion

An effective receiver channel recovery scheme can facilitate the uptake of high-frame-rate ultrasound techniques into compact ultrasound systems. To this end, we have devised a CNN-based channel recovery scheme and demonstrated its ability to recover a full set of RF data given multiple degrees of channelwise downsampling. The channel recovery framework is expected to improve beamformed image quality in high-frame-rate ultrasound systems that operate with a reduced number of receiving channels. This work can thus aid the adoption of high-frame-rate principles into compact ultrasound systems, extending the imaging paradigm into more remote and austere healthcare environments.

References

[1] G. M. Lanza, "Ultrasound imaging: Something old or something new?" *Investigative Radiol.*, vol. 55, no. 9, pp. 573–577, 2020.

[2] 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.

[3] J. M. Baran and J. G. Webster, "Design of low-cost portable ultrasound systems: Review," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Sep. 2009, pp. 792–795.

[4] B. P. Nelson and A. Sanghvi, "Out of hospital point of care ultrasound: Current use models and future directions," *Eur. J. Trauma Emergency Surg.*, vol. 42, no. 2, pp. 139–150, Apr. 2016.

[5] S. Sippel, K. Muruganandan, A. Levine, and S. Shah, "Review article: Use of ultrasound in the developing world," *Int. J. Emergency Med.*, vol. 4, no. 1, pp. 1–11, Dec. 2011.

[6] 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.

[7] J. D. Larson, "2-D phased array ultrasound imaging system with distributed phasing," U.S. Patent 5 933 229, Jul. 20, 1993.

[8] J. Yu, H. Yoon, Y. M. Khalifa, and S. Y. Emelianov, "Design of a volumetric imaging sequence using a vantage-256 ultrasound research platform multiplexed with a 1024-element fully sampled matrix array," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 67, no. 2, pp. 248–257, Feb. 2020.

[9] T. M. Carpenter, M. W. Rashid, M. Ghovanloo, D. M. J. Cowell, S. Freear, and F. L. Degertekin, "Direct digital demultiplexing of analog TDM signals for cable reduction in ultrasound imaging catheters," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 63, no. 8, pp. 1078–1085, Aug. 2016.

[10] G. R. Lockwood, P.-C. Li, M. O'Donnell, and F. S. Foster, "Optimizing the radiation pattern of sparse periodic linear arrays," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 43, no. 1, pp. 7–14, Jan. 1996.

[11] 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.

[12] A. Besson, R. E. Carrillo, O. Bernard, Y. Wiaux, and J.-P. Thiran, "Compressed delay-and-sum beamforming for ultrafast ultrasound imaging," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 2509–2513.

[13] G. David, J. L. Robert, B. Zhang, and A. F. Laine, "Time domain compressive beamforming of ultrasound signals," *J. Acoust. Soc. Am.*, vol. 137, no. 5, pp. 2773–2784, 2015.

[14] I. A. M. Huijben, B. S. Veeling, K. Janse, M. Mischi, 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.

[15] D. Perdios, M. Vonlanthen, F. Martinez, M. Arditi, and J.-P. Thiran, "Single-shot CNN-based ultrasound imaging with sparse linear arrays," in *Proc. IEEE Int. Ultrason. Symp. (IUS)*, Sep. 2020, pp. 1–4.

[16] D. Perdios, A. Besson, M. Arditi, and J.-P. Thiran, "A deep learning approach to ultrasound image recovery," in *Proc. IEEE Int. Ultrason. Symp. (IUS)*, Sep. 2017, pp. 1–4.

[17] H. Liebgott, R. Prost, and D. Friboulet, "Pre-beamformed RF signal reconstruction in medical ultrasound using compressive sensing," *Ultrasonics*, vol. 53, no. 2, pp. 525–533, Feb. 2013.

[18] D. Xiao, W. M. K. Pitman, B. Y. S. Yiu, A. J. Y. Chee, and A. C. H. Yu, "Minimizing image quality loss after channel count reduction for plane wave ultrasound via deep learning inference," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 69, no. 10, pp. 2849–2861, Oct. 2022.

[19] V. Kumar, P.-Y. Lee, B.-H. Kim, M. Fatemi, and A. Alizad, "Gap-filling method for suppressing grating lobes in ultrasound imaging: Experimental study with deep-learning approach," *IEEE Access*, vol. 8, pp. 76276–76286, 2020.

[20] A. Ramkumar and A. K. Thittai, "Compressed sensing approach for reducing the number of receive elements in synthetic transmit aperture imaging," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 67, no. 10, pp. 2012–2021, Oct. 2020.

[21] R. Anand and A. K. Thittai, "Towards practical implementation of the compressed sensing framework for multi-element synthetic transmit aperture imaging," *Ultrasonics*, vol. 112, Apr. 2021, Art. no. 106354.

[22] D. Cacko and M. Lewandowski, "Shear wave elastography implementation on a portable research ultrasound system: Initial results," *Appl. Sci.*, vol. 12, no. 12, p. 6210, Jun. 2022.

[23] R. S. C. Cobbold, "Properties of imaging arrays," in *Foundations of Biomedical Ultrasound*, vol. 2. London, U.K.: Oxford Univ. Press, 2007, ch. 7, pp. 427–460.

[24] S. Wang, J. Cao, and P. S. Yu, "Deep learning for spatio-temporal data mining: A survey," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 8, pp. 3681–3700, Aug. 2022.

[25] A. Araujo, W. Norris, and J. Sim, "Computing receptive fields of convolutional neural networks," *Distill*, vol. 4, no. 11, pp. 1–10, Nov. 2019.

[26] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. ICML*, 2013, pp. 1–18.

[27] X. Mao, C. Shen, and Y. B. Yang, "Image restoration using very deep convolutional encoder–decoder networks with symmetric skip connections," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 1–16.

[28] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent.*, 2015, pp. 1–13.

[29] D. Hyun et al., "Deep learning for ultrasound image formation: CUBDL evaluation framework and open datasets," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 68, no. 12, pp. 3466–3483, Dec. 2021.

[30] A. Rodriguez-Molares et al., "The generalized contrast-to-noise ratio: A formal definition for lesion detectability," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 67, no. 4, pp. 745–759, Apr. 2020.

[31] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[32] F. H. Khan, M. A. Pasha, and S. Masud, "Advancements in microprocessor architecture for ubiquitous AI—An overview on history, evolution, and upcoming challenges in AI implementation," *Micromachines*, vol. 12, no. 6, p. 665, Jun. 2021.

[33] S. Mittal, "A survey on optimized implementation of deep learning models on the NVIDIA Jetson platform," *J. Syst. Archit.*, vol. 97, pp. 428–442, Aug. 2019.

[34] H. So, J. Chen, B. Yiu, and A. Yu, "Medical ultrasound imaging: To GPU or not to GPU?" *IEEE Micro*, vol. 31, no. 5, pp. 54–65, Sep. 2011.

[35] J. A. Jensen, S. I. Nikolov, K. L. Gammelmark, and M. H. Pedersen, "Synthetic aperture ultrasound imaging," *Ultrasonics*, vol. 44, pp. e5–e15, Dec. 2006.

[36] E. Roux, F. Varray, L. Petrusca, C. Cachard, P. Tortoli, and H. Liebgott, "Experimental 3-D ultrasound imaging with 2-D sparse arrays using focused and diverging waves," *Sci. Rep.*, vol. 8, no. 1, p. 9108, Jun. 2018.

[37] P. Mattesini, A. Ramalli, L. Petrusca, O. Basset, H. Liebgott, and P. Tortoli, "Spectral Doppler measurements with 2-D sparse arrays," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 67, no. 2, pp. 278–285, Feb. 2020.

[38] H. Shin et al., "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1285–1298, May 2016.

William M. K. Pitman received the B.A.Sc. degree in electrical engineering from Queen's University, Kingston, ON, Canada, in 2019, and the M.A.Sc. degree in electrical and computer engineering from the University of Waterloo, Waterloo, ON, in 2022.

From 2017 to 2018, he was an Electrical Engineering Intern at L3Harris Wescam, Hamilton. He is currently a Biomedical Signal Engineer at FluidAI Medical, Kitchener. His research interest is in applying artificial intelligence tools to ultrasound and medical diagnostics.

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.

Di Xiao (Graduate Student Member, IEEE) received bachelor's degree in mathematics from the University of California, Berkeley, CA, USA, in 2014, the master's degree in electrical engineering from the University of Minnesota—Twin Cities, Minneapolis, MN, USA, in 2016, and the Ph.D. degree in electrical and computer engineering from the University of Waterloo, Waterloo, ON, Canada, in 2024.

He was a Software Engineer at IBM Canada, Markham, ON, Canata, and has also interned at the Hitachi Central Research Laboratory, Tokyo, Japan. He is a Postdoctoral Fellow 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.
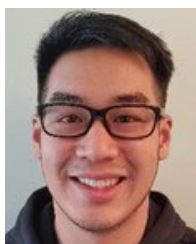
Billy Y. S. Yiu (Member, IEEE) received the B.Eng. degree (Hons.) in medical engineering and the M.Phil. degree in electrical and electronic engineering from the University of Hong Kong, Hong Kong, in 2007 and 2010, respectively, and the Ph.D. degree in electrical and computer engineering from the University of Waterloo, Waterloo, ON, Canada, in 2019.

Before joining the University of Waterloo, he was a Research Staff with the Biomedical Ultrasound Laboratory, University of Hong Kong. From 2021 to 2023, he was a Research Assistant Professor and the Co-Principal Investigator of the Laboratory on Innovative Technology in Medical Ultrasound (LITMUS), University of Waterloo. In 2023, he joined the Technical University of Denmark, Lyngby, Denmark, as an Associate Professor. He is also the Co-Instructor for the "Ultrasound Signal Processing with GPUs" short course at the IEEE Ultrasonics Symposium. His research interests are in advanced ultrasound imaging techniques and systems.

Dr. Yiu was a 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, ON, Canada. His research interests are in designing new vascular ultrasound techniques and related experimental tools.
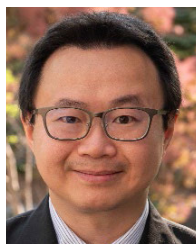
Alfred C. H. Yu (Fellow, IEEE) received the B.Sc. degree in electrical engineering from the University of Calgary, Calgary, AB, Canada, in 2002, and the M.A.Sc. and Ph.D. degrees in biomedical engineering from the University of Toronto, Toronto, ON, Canada, in 2004 and 2007, respectively.

From 2007 to 2015, he was the Founding Principal Investigator of the Biomedical Ultrasound Laboratory, University of Hong Kong, Hong Kong. In 2015, he relocated to the University of Waterloo, Waterloo, ON, Canada, where he is currently the Assistant Vice-President, Research and International, and Professor of Biomedical Engineering. He has long-standing research interests in ultrasound imaging and therapeutics.

Dr. Yu has been named as a Steacie Fellow by NSERC. He is a Fellow of the Canadian Academy of Engineering (CAE), the Engineering Institute of Canada (EIC), and the American Institute of Ultrasound in Medicine (AIUM). 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 previously the Program Chair of the 2023 IEEE Ultrasonics Symposium, the Chair of the Medical Ultrasound Group of the IEEE Ultrasonics Symposium, and an Elected AdCom Member of the IEEE UFFC Society. He is now the Editor-in-Chief of the IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control.