IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 31 March 2023, accepted 24 June 2023, date of publication 4 July 2023, date of current version 11 July 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3292212

RESEARCH ARTICLE

Reverberation Suppression in Echocardiography Using a Causal Convolutional Neural Network

TOLLEF STRUKSNES JAHREN[1], ANDERS RASMUS SØRNES[2], BASTIEN DÉNARIÉ[2], ERIK STEEN[2], TORE BJÅSTAD[2], AND ANNE H. SCHISTAD SOLBERG[1]
[1]Department of Informatics, University of Oslo, 0316 Oslo, Norway
[2]GE Healthcare, 3183 Horten, Norway

Corresponding author: Tollef Struksnes Jahren (tollefsj@ifi.uio.no)

ABSTRACT While ultrasound imaging has seen vast technical advances over the last decades, transthoracic echocardiography still suffers from image quality degradation caused by acoustic interaction with inhomogeneous tissue layers between the transducer and the heart. The acoustic energy reflections from echogenic structures such as skin, subcutaneous fat, bone, cartilage, intercostal muscle tissue, and lungs can form a dense overlay of echoes occluding the structural information resulting in a degradation of the diagnostic value. We propose a new method for reducing this reverberational clutter inspired by how the brain addresses the problem; identifying the reverberation overlay by the way it constitutes a pattern of speckles that moves in one cohesive motion different from that of the underlying structures. With this approach, we effectively render the clutter suppression as a video separation problem. Compared to traditional clutter rejection methods that tend to specialize in either temporal or spatial qualities, we find a neural network to be more flexible in incorporating both temporal and spatial information. We generate a pseudo-paired data set using *in vivo* data by excising patches off hypo-echoic regions of strongly reverberation-affected clinical recordings and superimposing them onto clean clinical recordings. The pseudo-paired data set of beamformed in-phase and quadrature component (IQ)-data is used to train a neural network to suppress reverberations in cine-loops. We demonstrate that this post-beamformer method can enhance image quality in *in vivo* and make valuable clinical structures clearer in a commercial system. We show that the method does not display any tendency to generate false cardiac structures, and that rapid motions from e.g. valve leaflets retain high structural integrity and low levels of blurring. Our results suggest that this method can be an effective and robust tool for suppressing reverberations in transthoracic ultrasound imaging.

INDEX TERMS Reverberation, haze, clutter, ultrasound, echocardiography, neural network, deep learning.

I. INTRODUCTION

A. BACKGROUND

Echocardiography is a widely used diagnostic tool in medicine due to its non-invasive and real-time imaging capabilities. However, the quality of cardiac ultrasound images can be degraded by various factors, which can limit the accuracy of diagnosis and treatment. Dahl et al. established the two primary sources of reduction in ultrasound image quality: wavefront aberration and reverberation [1]. Wavefront aberration is caused by inhomogeneities in the sound speed, which prevent all parts of the wavefront from reaching the focal point simultaneously. This de-focusing increases the side-lobe levels and reduces the spatial resolution of the image. On the other hand, reverberation artifacts occur when the acoustic energy is reflected twice or more before returning to the transducer. The beamformer does not model these multi-paths, and all signals with equal time-of-flight are combined, representing the same location during image formation.

The origin of reverberation noise has been subjected in various studies. Fatemi et al. investigated five sources of reverberation noise in apical four-chamber view echocardiograms [2]. The scenarios included acoustic energy reflections from echogenic structures such as skin, subcutaneous fat,

The associate editor coordinating the review of this manuscript and approving it for publication was Qilian Liang.

bone, cartilage, intercostal muscle, lung, and out-of-scan-plane heart tissue. Patients with short intercostal distance or adverse relative locations of the heart and ribs were considered particularly difficult-to-scan. Connective tissue and fatty structure of the subcutaneous layer have similarly been identified as primary sources of reverberation in abdominal imaging [3], [4].

Reverberation noise can appear as replicas of proximal/nearby structures, as bright tails after stronger reflectors, or as cloud-like diffuse haze. In many cases, this clutter can be found to be close to stationary and exhibit a speckle-like texture. Reverberation noise is especially visible in hypoechoic or anechoic regions and becomes problematic when it occludes weak cardiac structures [5]. As an example, such a haze overlay can make the endocardial border challenging to localize close to the apex [6]. Together, wavefront aberration and reverberation increase noise levels and limit the image contrast and resolution. Degradation in image quality through these mechanisms may significantly reduce the diagnostic value of echocardiography. Flynn et al. [7] performed a retrospective study of cardiac surgical patients and found 17% and 37% of the cases to have inadequate left and right ventricle imaging, respectively. Moreover, clutter noise poses an additional challenge in the analysis of image segmentation, motion analysis, strain measurements, and flow estimation. In severe cases of image degradation, patients may need to be referred to more invasive and expensive forms of examination like transesophageal echocardiography, contrast injections, or alternative modalities like MRI or CT.

B. RELATED WORK

Different techniques have been proposed to suppress reverberation clutter. The most successful approach is tissue harmonic imaging (THI) [8], which is now used in clinical systems and has proven to be of crucial clinical importance [9]. THI utilizes the build-up of second harmonic energy within the body to effectively suppress reverberation clutter originating from near-field structures. Another technique that uses the nonlinear property of tissue is SURF imaging [10]. SURF uses a dual frequency band pulse-complex to separate first and multi-order scatterers and has shown promising in-vivo results [11].

Coherence-based approaches aim to obtain higher contrast by weighting each image pixel by the ratio of the coherent energy to the total incoherent energy. The coherence factor (CF) was first introduced as a quantitative measure of local image quality by Hollman et al. [12]. The concept of coherence as a weighting factor to achieve adaptive imaging has later been further elaborated on, for instance, expanding it into the concept of a generalized coherence factor (GCF) more broadly formulated as the ratio of the low-frequency spectral component to the to total aperture energy [13].

Shin et al. proposed to use multi-phase apodization with cross-correlation (MPAX) as an alternative to generating the weighting matrix. By using multiple apodization functions, their method showed robustness *in vivo* where a high level of reverberation clutter is expected [14]. Short-lag spatial coherence (SLSC) differs from the previously mentioned coherence-based methods as it was developed to generate images based on the lateral spatial coherence directly without considering the echo's brightness [15], [16]. SLSC has demonstrated superior contrast, contrast-to-noise ratio (CNR), and signal-to-noise ratio (SNR) compared to delay-and-sum B-mode images, but struggles with the detection of point-like targets in speckle-based backgrounds.

A model-based approach for the suppression of acoustic clutter was proposed by Byram et al. [17]. The method is called aperture domain model image reconstruction (ADMIRE), and it decomposes aperture domain ultrasound channel data to the modeled acoustic scattering sources. ADMIRE showed successful reduction of unwanted acoustic clutter in fundamental and harmonic imaging. Brickson et al. [18] utilized a 3D convolutional neural network to suppress diffuse reverberation noise by training it on both simulated hazy and haze-free channel data. The model demonstrated efficient reverberation suppression and generalization to *in vivo* data. Luchies et al. [19] used a deep neural network to suppress the frequency characteristics of off-axis scattering in delay-aligned channel data. One challenge with this approach is that it relies on simulated training data, which may not accurately capture general reverberation processes beyond off-axis scattering. The authors also attribute an observed reduction in speckle SNR to inaccuracies in the modeling and propose the generation of training data through experiments.

Alongside the techniques focusing on channel data and beamforming, a significant body of literature proposes post-processing filters in order to suppress acoustic clutter. These filters can be temporal, spatial, or spatiotemporal, and aim to separate the signals from cardiac structures and the clutter by decomposing the acquired data into a set of basis functions. The basis functions can either be predefined or adaptively learnt from the data.

Reverberation clutter is partially caused by slow-moving organs and parts such as ribs and lungs. Separating input data into velocity components has been extensively studied in color flow imaging (CFI), where the separation of low-velocity blood flow from non-stationary tissue has been a challenging problem. Although the acquisition in CFI differs from the B-mode acquisition as multiple pulses are transmitted in the same direction for high pulse repetition frequency, the reverberation clutter and tissue clutter separation problems share similarities. Bjærum et al. [20] examined the use of three static high-pass filters, finite impulse response (FIR), infinite impulse response (IIR), and polynomial regression filters, in order to suppress tissue clutter in CFI. Yu et al. investigated critical design parameters of principal component analysis (PCA) eigen-based clutter filters [21]. The approaches considered were the estimation of eigen-components using single-ensemble or multi-ensemble and filtering of the eigenvalues in terms of value-based or frequency-based. The value-based approach assumes clutter has higher signal power than blood signals and achieves clutter filtering by setting the k's largest eigenvalues to zero or eigenvalues above a predefined threshold (or relative threshold) to zero. In the frequency-based approach, the mean Doppler frequency of the eigenvectors is calculated using the lag-one autocorrelation estimation. The eigenvalues corresponding to the eigenvectors with a mean frequency that lies within the bandwidth of the estimated center clutter frequency are rejected. Yu et al. was unable to achieve stable filtering using the value-based algorithm. However, they found the frequency-based approach more effective than static filters in suppressing clutter in cases of substantial tissue motion.

Demené et al. investigated SVD-based clutter filtering in functional and Doppler Imaging [22]. The alternative acquisition mode of ultrafast plane-wave imaging facilitates using a longer ensemble length and brings high lateral coherence compared to conventional multi-pulse Doppler acquisition. Demené et al. argues that tissue motion is less deformable than red blood cells during motion. Therefore, signals coming from tissue are more spatially consistent over time than blood signals. The separation of tissue and blood becomes the problem of selecting sub-spaces produced by the SVD. The same group also presented a comprehensive list of adaptive estimators for the hard thresholding of the singular values in [23]. The fourteen estimators considered based their estimation on the distribution of the singular values, the right (temporal) singular vectors, and the left (spatial) singular vectors. The recommended estimator was based on a spatial similarity (correlation) matrix, and spatial context was considered equal to or more important than the information in time for clutter filtering. This is in contrast to conventional focused ultrasound imaging, where clutter filters rely on temporal information only. It is worth mentioning that the SVD was computed from the whole image and not patch-wise in this work. It was also demonstrated that imaging artifacts and the estimated CNR are highly sensitive to selected singular values.

Mauldin et al. proposed a singular value filter (SVF) in the context of reverberation clutter rejection [24]. The two main findings were the importance of using complex data and a non-binary weighting scheme for the singular values. The soft thresholding of the singular values could suppress the reverberation clutter signal while avoiding artifacts generated from hard thresholding. Assuming that reverberation clutter is the dominating signal component and has a stationary characteristic, they developed a modified sigmoid function for calculating the weighting coefficients of the singular values. The insight was that signals from reverberation clutter are highly compressed, resulting in large singular values explaining most of the data. On the contrary, tissue signal has more richness and a flatter singular value spectrum. The SVF method was evaluated on experimental mouse heart data, and an average increase of 1.8dB in CNR was demonstrated. Turek et al. used Morphological Component Analysis (MCA) to demonstrate competitive performance in reverberation suppression to SVF [25]. The MCA approach uses an adaptive basis, such as the SVF, but assumes that the tissue and clutter signals can be sparsely encoded. The sparse encoding is a set of non-orthonormal vectors forming a redundant basis. The selection process of which vectors correspond to tissue and clutter was governed by the assumption that reverberation clutter signals are semi-stationary. The benefit of using MCA was reported to be less removal of tissue compared to SVF.

While it is commonly assumed that reverberation clutter is close to temporally stationary, this is not always the case. Thus, methods that rely on tissue and reverberation clutter having different motion characteristics may not always succeed. Reverberation clutter originating from reflections of cardiac tissue can be temporally connected to the tissue motion, and tissue structures can also become almost stationary in certain conditions (e.g. akinetic myocardial regions and during end-diastole). To address these issues, Sjoerdsma et al. developed a method that operates on individual frames, and no assumption regarding tissue and reverberation motion is required. This method suppresses near field clutter using complex-valued orientated fast wavelets [26]. The B-mode images were decomposed into a multiscaled pyramid of four scales and four orientations. It was shown that near field clutter was predominantly present in the highest and lowest bandpass sub-images using vertical-orientated wavelets. The motion-invariant method preserves the US speckles making it useful to apply prior to speckle tracking methods. The method showed improved contrast of 4.3dB on average on B-mode images from apical and parasternal views. Tay et al. also proposed a method for reverberation suppression based on the wavelet transform [27]. The reverberation signal is estimated from RF data by soft threshold discrete wavelet transform (DWT) coefficients. The estimated reverberation signal is subtracted from the original data to capture the obscured tissue signal.

The post-processing filters mentioned so far are based on either SVD or wavelet decomposition. These methods are limited to linear representations of data and do not accurately model spatiotemporal modeling which is essential for separating cardiac structure from reverberations. This spatiotemporal information is critical for our minds to connect speckles that move cohesively into collectively moving patterns. Learning-based methods, such as deep neural networks, are more suited to include spatiotemporal information and modeling highly non-linear systems. For instance, Tabassian et al. trained a 3D U-Net to remove synthetically superimposed artifacts in synthetic 2D echocardiographic sequences [28]. They showed that their network outperformed the SVD-based clutter filter proposed in [24]. Although the method rejected the generated artifacts with excellent performance, its generalization to *in vivo* data was not demonstrated and the superimposed artifacts lacked realism and relevance to typical reverberation patterns.

Despite the promising reverberation suppression methods proposed, the challenge of solving reverberation clutter remains an active research topic. In this study, we propose two different 3D (2D spatial plus time) convolutional neural networks namely U-Net and Causal-U-Net for filtering out reverberation clutter from tissue in cine-loops. The difference between the networks is that Causal-U-Net is temporally causal suited for real-time inference, whereas U-Net without the restriction is designed for better playback performance. As no supervised learning dataset exists for this task, neural networks are typically trained on synthetically generated data from simulations. However, modeling reverberation clutter accurately is challenging, especially because reverberations can originate from out-of-plane multiple scattering.

To address this challenge, we suggest treating the reverberation suppression task as a video separation problem and propose generating a pseudo-paired data set using *in vivo* data by superimposing patches from hazy videos onto clean videos.

Our contribution is three-fold:

1) We present an approach for creating a realistic pseudo-paired data set for reverberation suppression using *in vivo* data.
2) Using the pseudo-paired data set, we demonstrate that convolutional neural networks can learn to separate first-order scattering and multi-order scattering in a video separation problem.
3) We show that the reverberation suppression can be made causal and therefore suitable for real-time operation.

The rest of this paper is organized as follows: Section II describes the data collection, annotation process, structure of the neural networks, training setup, and evaluation of our method using *in vivo* data. Section III reports our results and compares our networks to the SVF method [24]. Section IV discusses the limitations of our method and future work, and Section V summarizes the paper and draws conclusions.

II. METHODOLOGY

A. DATA COLLECTION

The data used in this study was obtained from two sources: recordings made specifically for this study and recorded data from previously collected data sets. The data was fully anonymized and informed consent was obtained and legal agreements with the data providers ensured compliance with local requirements. The ultrasound system used was a Vivid E95 (GE Vingmed Ultrasound AS, Horten, Norway) with a GE 4Vc-D transducer. Second harmonic imaging was used with a transmit frequency between 1.56 Mhz and 1.85 Mhz. Statistics about the acquisition parameters used for the recordings are provided in Table 1.

In addition to the clinical data, we also collected cine-loops from a phantom to simulate cases of almost static reverberations. The phantom used was GAMMEX Sono403 with an attenuation of 0.7dB/cm/MHz. The cine-loops captured from the phantom were obtained with little to no movement of the probe, and consist of almost static scatters.

TABLE 1. Acquisition setup.

Parameters ↓	Min	Max	Average	std
Scan depth (cm)	10.0	20.0	14.8	1.7
Beam separation (Deg)	0.35	0.66	0.49	0.12
Number of beams [#]	100	226	154	44
Sampling rate [Mhz]	3.125	3.125	3.125	0
Sector width [Deg]	50	90	70	9
Transmit center frequency [Mhz]	1.56	1.85	1.70	0.11
Frames in cine-loop [#]	64	128		

The list of scan parameters from the recordings labeled clean and hazy. The sampling rate corresponds to the sampling rate of the demodulated IQ signals.

TABLE 2. Data collection.

Datasets ↓	Train & validation		validation-two	Test
	Clean	Cluttered	-	-
Site 1	46 (25)	35 (22)	14 (9)	14 (9)
Site 2	-	-	2 (1)	3 (1)
Site 3	13 (5)	23 (6)	4 (2)	6 (2)
Site 4	-	-	5 (2)	5 (1)
Phantom	-	6 (1)	-	-
Total	59 (30)	64 (29)	25 (14)	28 (13)

The table shows the number of recordings and the number of individuals in parenthesis. The data used for training and validation are grouped into clean and cluttered recordings. The recordings selected for the validation-two and test data sets consists of difficult-to-scan patients. The total number of available recordings was 721 from 130 subjects. The majority of the recordings (539) and subjects (105) were from site 1.

The data format used in this study is beamformed IQ data, which refers to baseband-filtered demodulated RF data prior to scan conversion. The data format will be noted as IQ data. Retrospective Transmit Beamforming (RTB) was performed on the data as part of the beamforming process.

B. ANNOTATION AND PREPROSSESSING

Subjects producing *clean* and subjects producing *hazy* images were required to generate a data set with pseudo-paired cine-loops. To efficiently annotate the cine-loops. we developed a Python GUI. We carefully went through an extensive collection of recorded data and marked each recording as *clean*, *hazy*, or in between based on the level of clarity of the images. The number of subjects and the number of recordings are given in Table 2. The *clean* and the *hazy* recordings were split into the training and validation data sets with a ratio of 80% and 20%, respectively. The split was done patient-wise to avoid corrupting the validation data set.

In this study, we selected cluttered patches from anechoic regions of the images, which were then annotated to exclude cardiac structures. The selection of cluttered patches is illustrated in Figure 1. The patches included the same spatial pixels throughout the cine-loops and were made as large

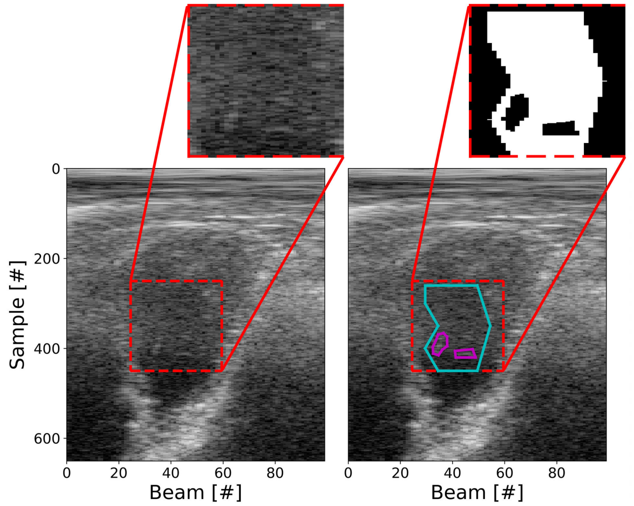

FIGURE 1. An example of a patch of haze selected from a loop is shown in the top left sub-figure. For each frame, a corresponding binary mask was annotated. The regions within the magenta polygons are set to zero and the area outside of the blue polygon is set to zero. The resulting mask is shown in the top right sub-figure.

as possible to contain spatial characteristics of clutter without including adjacent cardiac structures. However, in the end-systole phase where the left ventricle (LV) is small, we allowed some cardiac structures to enter the selected patches to retain larger patch sizes. We annotated the cardiac structures and created binary masks, where pixels corresponding to the cardiac structures were set to zero. These masks were used in the loss function during training to prevent the cardiac structures from wrongly labeling the data set. An example of a binary mask is shown in Figure 1. The dimensions of a cluttered patch are (Nt_p, Ns_p, Nb_p), where Nt_p is the number of frames, Ns_p is the number of samples (fast time), and Nb_p is the number of beams.

The cine-loops selected for the validation-two and test data sets were challenging due to excessive clutter or small cardiac structures obscured by reverberations. These cine-loops were not annotated as *hazy* subjects because not a large region was free from cardiac structures throughout the cardiac cycles. In addition, the cine-loops were also needed for selecting and evaluating the neural networks. The recordings in the validation-two data set were used as an extension to the validation data set to select the final model. This was necessary since the validation data set does not include accurate modeling of true haze.

To generate cluttered cine-loop overlays, we augmented and spatially stacked cluttered patches to form realistic clutter overlays of the same spatial size as the clean cine-loops. Most of the cluttered patches were selected to be rectangular, making them easy to stack. Figure 2 shows an example of a cluttered overlay superimposed onto a clean recording, and the left and right subfigures constitute the pseudo-paired data. The data used in this study is complex IQ data, where each pixel has a real and an imaginary component. The images are displayed using the complex envelope on a logarithmic scale.

The signal model can be described as

$$\mathbf{x} = \mathbf{y} + \mathbf{c} \tag{1}$$

where $\mathbf{y}$ is the *clean* reference data, $\mathbf{c}$ is the cluttered overlay, and $\mathbf{x}$ is the corresponding pseudo-cluttered data. Both $\mathbf{y}$ and $\mathbf{c}$ contain electrical noise, which is assumed to be weaker than the acoustic noise present in the cluttered signals. However, the modeling of the pseudo-paired data has a few challenges that need to be addressed: (1) There is a misalignment between the frames corresponding to events such as end-diastole and end-systole in the clean reference data and the superimposed clutter overlays. Since ECG recordings are not available, it is difficult to event align the data; (2) The data is in beam space and, therefore, anatomically morphed as a function of the acquisition parameters. The selected cluttered patches often originate from depth samples between 200 and 450. Stacking these patches randomly to form the cluttered overlaid breaks the geometrical correspondence to the *clean* data, and (3) stacking the cluttered patches creates discontinuities.

Most of the data suitable for the categories *clean* and *hazy* were obtained from apical two and four-chamber recordings. These images were generally cleaner, and the LV extended over a large semi-stationary region, which was well-suited for extracting *hazy* patches. In the training data set, 44 out of the 48 *clean* recordings are from an apical view.

The amplitude of the complex envelope was annotated for cardiac structures and hypo-echoic regions in all selected cine-loops. This was done to maintain control over the signal levels when superimposing cluttered data onto the clean reference images. Histograms of the contrast between cardiac structures and hypo-echoic regions are shown in Figure 3. On average the contrast was found to be 23.5 dB higher in the *clean* recordings compared to the *hazy* recordings. This contrast can serve as an upper bound for what to aim for during clutter suppression.

C. NEURAL NETWORK ARCHITECTURES

We used two 3D convolutional neural networks based on the U-Net architecture presented in [29]: U-Net and Causal-U-Net. The networks are identical except that Causal-U-Net uses causal convolutions along the frame dimension, as illustrated in Figure 4. The benefit of causal convolutions is that the model does not impose any frame delay during inference. The drawback is that the causal model may perform inferior to the model utilizing information in future frames.

During training, the dimension of a single input sequence ($\mathbf{x}$) is (n_c=2, n_f=30/45, n_s=576, n_b=100), where n_c is the number of channels, n_f is the number of frames (30 for U-Net and 45 for Causal-U-Net), n_s is the number of samples (fast time), and n_b is the number of beams. The neural network is real-valued, and the real and imaginary components of the complex IQ data are stacked in the channel dimension (n_c). Figure 5 illustrates the network structure, which is the same for both U-Net and Causal-U-Net. The neural networks can be described as follows:

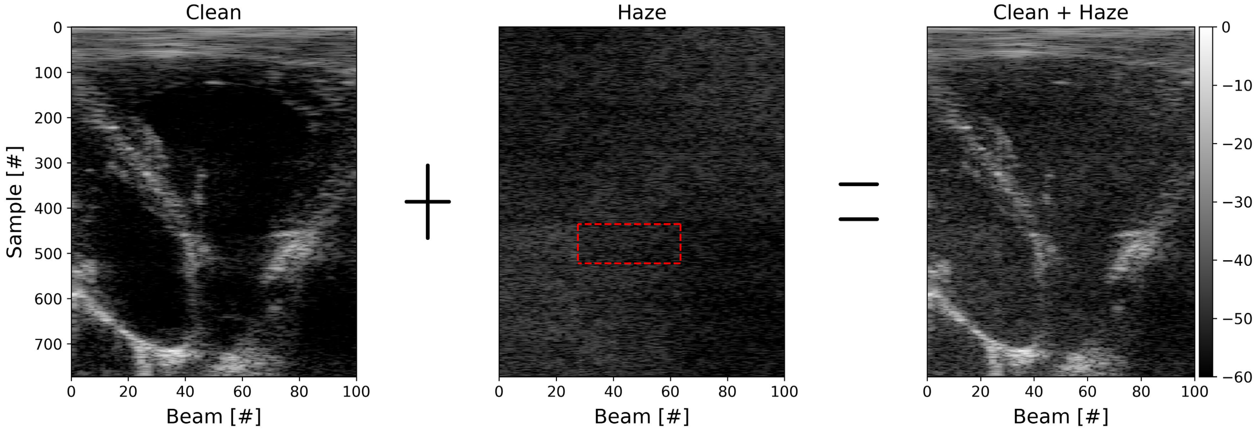

FIGURE 2. The figure shows how a pseudo-paired recording is produced. The dashed red box illustrates the size of the *hazy* patch used to form the *hazy* overlay.

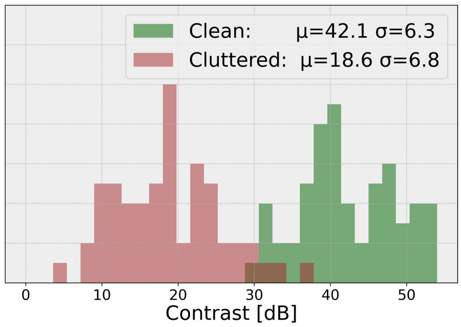

FIGURE 3. Histograms showing the distribution of the contrast between cardiac structures and cavity for the *clean* and *hazy* recordings in the training and validation datasets.

- Each level of the encoder and decoder consists of a residual block followed by a down-sampling layer or an up-sampling, respectively. The number of channels is doubled in each layer.
- The residual blocks are constructed from 3D convolutions (conv3D), group normalization layers (GN) [30], and the Leaky-ReLU (LReLU) activation function [31]. The layers are combined in the following order: conv3D → GN → LReLU → conv3D → GN → LReLU, and is noted as *ReLU before addition* in [32].
- All 3D convolutions use a kernel size of (3 × 3x3), a stride of (1 × 1x1), and no dilation. Padding is applied along the spatial dimensions (samples, beams) to keep the original input shape. U-Net also uses padding along the frame dimension n_f.
- The stride and kernel sizes of the max pooling operations (given as n_f, n_s, n_b) are: (1,2,1), (1,2,2), and (1,2,2), for the encoder levels 1, 2, and 3. The transposed convolutions mirror the max pooling operations.

D. LOSS FUNCTION

To train the neural networks, we use an element-wise loss function (L) based on the Huber loss (H) with $\delta = 1$.

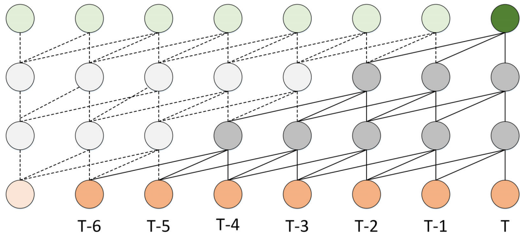

FIGURE 4. Illustration of causal convolutions with kernel size 3. The frame indices are noted as $T, T-1, T-2, \ldots, T-6$.

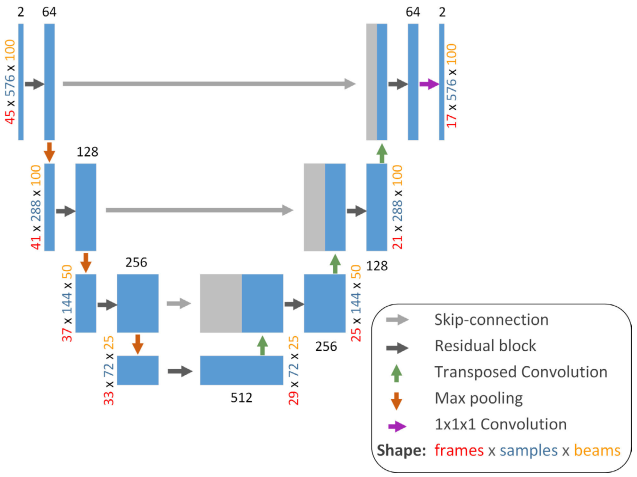

FIGURE 5. The U-Net and Causal-U-Net have 4 levels and include residual connections within each block, and skip connections (concatenation) between the encoder and the decoder for each level. The noted number of frames corresponds to Causal-U-Net.

For a single element (i) in the real-valued output data $\hat{\mathbf{y}} \in \mathbb{R}^{n_c \times n_f \times n_s \times n_b}$, the Huber loss function is defined as:

$$H(\hat{y}_i, y_i) = \begin{cases} \frac{1}{2}(\hat{y}_i - y_i)^2, & \text{if } \left|(\hat{y}_i - y_i)\right| < 1 \\ (\hat{y}_i - y_i) - \frac{1}{2}, & \text{otherwise.} \end{cases} \quad (2)$$

While the primary goal is to suppress clutter, inadvertently suppressing of cardiac structures has a more negative impact from a clinical perspective. Therefore, we propose using an asymmetrical loss function (A) to balance the clutter suppression and the preservation of cardiac structures. The asymmetrical loss function is defined as:

$$A(\hat{y}_i, y_i, \beta) = \begin{cases} H(\hat{y}_i, y_i) \cdot \beta, & \text{if } ||\hat{y}_i||_{2,n_c} < ||y_i||_{2,n_c} \\ H(\hat{y}_i, y_i), & \text{otherwise.} \end{cases} \tag{3}$$

The value of β determines the level of asymmetry. A higher value results in the model preserving more cardiac structures at the expense of reduced clutter suppression. The conditioning in Eq. (3) is the L2 norm along the channel dimension and compares the predicted and the reference complex envelopes.

To compute the loss function for a single recording, we take the weighted average over all dimensions as:

$$L(\hat{\mathbf{y}}, \mathbf{y}, \mathbf{w}, \beta) = \frac{1}{\sum_{i=1}^{n_c n_f n_s n_b} w_i} \sum_{i=1}^{n_c n_f n_s n_b} w_i \, A(\hat{y}_i, y_i, \beta). \tag{4}$$

where the binary mask, $\mathbf{w}$, is zero for pixels considered cardiac structures within the *hazy* patches and one for all other pixels. Refer to Figure 1 for an example of the mask.

Ultrasound data has a high dynamic range, and using the proposed loss function directly could introduce a bias towards high-amplitude pixels, potentially reducing the level of clutter suppression. To mitigate this, we apply a nonlinear scaling similar to the one presented in [18] to minimize the loss function's sensitivity to the amplitude.

$$g(\mathbf{y}) = \mathbf{y} \odot \frac{log_{10}(||\mathbf{y}||_{2,n_c} + 1)}{||\mathbf{y}||_{2,n_c}} \tag{5}$$

The $+1$ is used to avoid the steep fall of the logarithm for values below 1. The alternative loss function based on logarithmic compression of the networks' output is given as $L(g(\hat{\mathbf{y}}), g(\mathbf{y}), \mathbf{w}, \beta)$.

E. TRAINING AND AUGMENTATION

To optimize the model parameters, we used Kaiming initialization [33] and the AdamW [34] optimizer. The learning rate was warmed up linearly from $5 \cdot 10^{-6}$ to $5 \cdot 10^{-4}$ during the first 15 epochs. We employed a learning rate scheduler, which divided the learning rate by five every 180 epochs, with a total of 450 epochs. We did not apply any weight decay and used a batch size of 12 through gradient accumulation.

The loss curves for the training and validation data sets gradually reached a plateau over the 450 epochs. We did not use early stopping and instead selected the model weight configuration after the last epoch.

The *hazy* patches need to be stacked spatially to form diverse and realistic overlays. The following list describes the augmentation performed to create the pseudo-paired data set:

1) A single *hazy* patch is selected, and the amplitude is randomly adjusted to lie between −50 dB to −20 dB below the annotated tissue amplitude of the selected *clean* recording.
2) The *hazy* patch is spatially smaller than the *clean* recording. We duplicated the *hazy* patch and stacked it in the sample and beam direction. Each duplicated patch was randomly flipped with a probability of 50% along one or more of the axes (frame, sample, beam), and assigned a different gain ratio between 0.5 and 1.5.
3) To increase the diversity of the haze amplitude, we randomly scaled regions of the spatial extent of 100 samples and 40 beams with a value between 1/6 and 6. This resulted in unrealistic edges, so we applied lowpass filtering using a Gaussian kernel (sigma: 70 samples, 15 beams) to smooth the gain map.
4) We randomly flipped the *clean* data along the frame and beam axis with a probability of 50%. We did not flip along the sample (fast time) direction, as the near field is different from the fear field and could result in unrealistic data.
5) We resampled the *clean* and *hazy* data within the range of ±20% with a probability of 0.5%.

TABLE 3. Hyperparameter search.

Hyperparameter	Search space
Skip connection	[add, concatenation]
Base number of channels	[16, 32, 64]
Apply log compression on output	[True, False]
Asymmetrical loss weight (β)	[1, 5, 10]

F. HYPERPARAMETER SEARCH

We performed a grid search over the hyperparameters given in Table 3. We fixed the seeds of the pseudo-random number generators to make the comparison between hyperparameters as fair as possible. Other hyperparameters related to the optimization and the neural network structure were set to reasonable values based on initial experiments during code development. The final model selected yielded low validation loss and visually pleasing filtered images on the validation-two data set.

G. EVALUATION

To evaluate the performance of our method, we will use recordings from the test dataset. As our pseudo-paired data is a modeling of reverberations and does not include real reverberations, using recordings from the test dataset will allow us to assess the method's effectiveness on real-world data. However, using only the test dataset means we do not have a clean reference for comparison.

The following metrics will be used to evaluate the effectiveness of the neural networks: contrast, contrast-to-noise ratio (CNR), and generalized contrast-to-noise ratio (gCNR) [35]. Contrast and CNR are defined by

$$\text{Contrast} = 10 \log_{10}\left(\frac{\mu_t}{\mu_b}\right) \tag{6}$$

$$\text{CNR} = \frac{|\mu_t - \mu_b|}{\sqrt{\sigma_t^2 + \sigma_b^2}}, \tag{7}$$

where μ_t and μ_b denote the means, and σ_T^2 and σ_b^2 denote variances of the signal power of tissue and background (hypoechoic and anechoic regions), respectively.

As a neural network is a nonlinear filter that can alter the data's dynamic range, we will also use gCNR to measure the overlapping probability density functions of the cardiac structures ($p_t(x)$) and the hypoechoic and anechoic regions ($p_b(x)$). gCNR is defined as

$$\mathrm{gCNR} = 1 - \int \min\{p_t(x), p_b(x)\}\, dx. \tag{8}$$

In our experiment, the prior probabilities were set to 0.5. Although gCNR is robust to nonlinear modifications of the amplitudes, changes in texture, such as low-pass filtering, may alter the shape of the probability density functions and the value of gCNR.

Temporal consistency is crucial in medical image processing. Therefore, videos in mp4 format are available in the supplementary material to demonstrate the method's performance on challenging subjects with excessive clutter, thin structures overlapping with reverberations, and other complexities. Additionally, we will evaluate whether the IQ signal phase is altered after filtering. We will assess the method's stability through an adversarial attack based on [36]. An unstable model can be described as

$$||nn(\mathbf{x}+\mathbf{r}) - nn(\mathbf{x})||_2 \text{ is large, while } ||\mathbf{r}||_2 \text{ is small} \tag{9}$$

where nn denotes the model, and $\mathbf{r}$ denotes the perturbation tensor. By fixing the upper bound on the L^2 norm of $\mathbf{r}$, gradient ascent is used to search for the worst-case scenario.

H. COMPARISON METHOD

To compare the performance of the proposed models, we will use the Singular Value Filter (SVF) [24] as a benchmark method. The SVF was chosen due to its competitive performance and straightforward implementation. The SVF uses the singular value decomposition (SVD) to decompose the image data into a spatiotemporal matrix, which is then filtered using a modified sigmoid function. To find the optimal SVF parameters, we performed an extensive parameter search. However, we found that the configurations that yielded the highest contrast ratio also resulted in severe flickering artifacts. After trial and error, we selected a shared configuration for all recordings with the following parameters: $\tau = 0.15$, $\alpha = 20$, and the SVD was computed using 15 frames and the full spatial image. It should be noted that the filtered frame is the center frame.

III. RESULTS

A. CONFIGURATION OF THE NETWORKS

After conducting a hyperparameter search, a common set of hyperparameters yielded the best results for both U-Net and Causal-U-Net. The configuration consists of a skip connection of type concatenation, a base number of channels equal to 64, no logarithmic compression of the output, and an asymmetrical loss weight of five. The results presented in this section are from the test data set, except for subject twelve, which is part of the validation data set.

TABLE 4. Summary - image quality metrics.

Metric	Reference	SVF	Causal-U-Net	U-Net
C [dB]	16.2 ± 1.6	18.3 ± 1.4	22.5 ± 1.7	24.1 ± 1.9
CNR	0.37 ± 0.03	0.40 ± 0.04	0.39 ± 0.03	0.38 ± 0.03
gCNR	0.68 ± 0.05	0.76 ± 0.05	0.85 ± 0.03	0.85 ± 0.04

The table shows the mean and standard error of the mean on the test subjects one to eleven. The *reference* corresponds to the unfiltered images.

B. REVERBERATION SUPPRESSION ON TEST DATA

Figure 6 shows a comparison between the (unfiltered) reference image and the filtered images using SVF, Causal-U-Net, and U-Net for subject one. Subject one exhibits an excessive amount of haze throughout the imaging sector, which is especially problematic in the right ventricle as it occludes the underlying cardiac structure, as indicated by the yellow arrow. All three filtering methods successfully bring forward the cardiac structure and increase the contrast and the gCNR, but yield no real improvement in CNR.

Results from subjects two to eleven are displayed in Figure 7 and Figure 8. The first column shows the reference images, the second column displays the SVF images used as a benchmark, and the third and fourth columns show the two presented models Causal-U-Net and U-Net. All images are displayed with a dynamic range of 60 dB, and no s-curve is applied (Figure 1 in [35] shows an s-curve example). Each image sets the reference amplitude (dB=0) separately, as SVF may alter the signal amplitude.

Table 4 summarizes the image quality metrics for subjects one to eleven. The two presented models exceed the reference and SVF images in contrast and gCNR, but there is little to no improvement in CNR. Evaluating the filtered cine-loops for each subject, the following observations are valid for both of the two presented models:

- In subjects four, seven, eight, nine, and ten, the two presented models offer mainly visual enhancement. In these cases, the haze in the reference images is not of a severity that obscures diagnostic information, but merely impairs it. The two presented models still offer a reduction in clutter levels and enhancement of structural information that clearly outperforms that of the SVF method.
- Cardiac structures occluded by haze are more visible in subjects one, two, three, and six. Details such as endocardial borders and papillary muscles that are hardly detectable in the reference image stand out clearly with the two presented models. There is a perceived contrast between the heart walls and chambers achieved with these models that seem superior to that obtained with the SVF method.
- Subject five can be seen as a worst-case example where cardiac structures are strongly attenuated at end-diastole. The models are erroneously filtering the

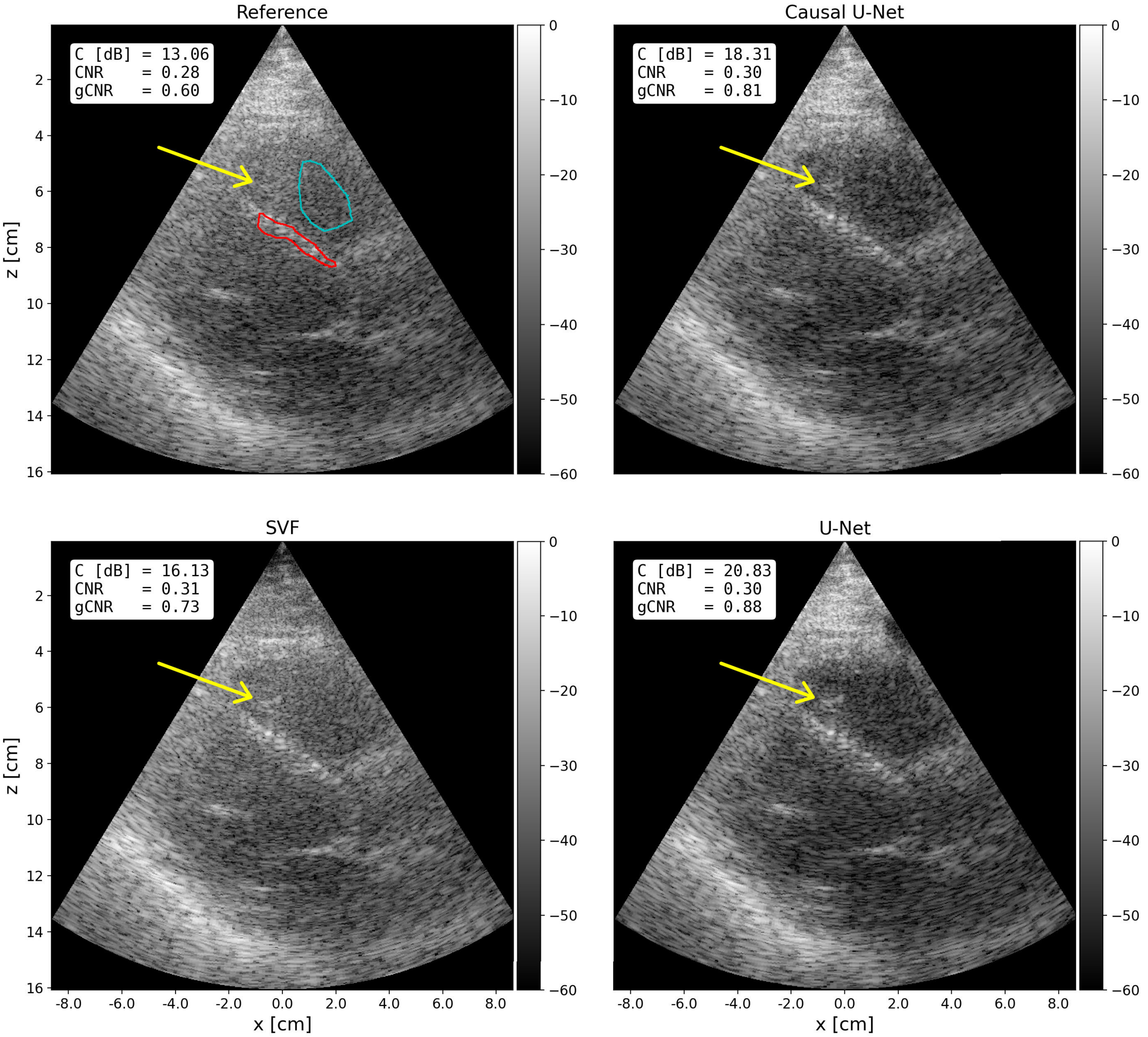

FIGURE 6. Comparison between filtered and unfiltered images of subject one. The blue (hypo-echoic region) and red (cardiac structure) marked regions are used for calculating the image quality metrics. The yellow arrow highlights a cardiac structure that is covered by haze. The four images are displayed using a dynamic range of 60 dB.

cardiac structures as they are stationary for a prolonged period of time (15 frames).

- Subjects eight and eleven show signs of cardiac structure suppression at end-diastole.
- In none of these cases do the two presented models exhibit the high temporal flickering that is commonly present in the output of the SVF method.

All cine-loops for the eleven subjects described above are available as supplementary material in the form of mp4 files.

C. STABILITY

To evaluate the stability of the models, we compared their filtered output with reference signals and also conducted an adversarial attack. While the results from Causal-U-Net will be presented here for convenience, we want to note that U-Net has also been evaluated using the same recordings and produced nearly identical results.

Figure 9 shows a comparison between a trace from Causal-U-Net and the reference data for a hazy subject (subject six) and a clean subject (subject twelve). The yellow lines in the top row display the location of the traces. In the middle row right column, we observe that the real and the imaginary components of the complex IQ data are passed through Causal-U-Net with minimal modification. As for the hazy data in the middle row left column, the filtered trace follows the reference data well, except for the low-value samples, which are predominantly considered to contain clutter. The phase difference between the fil-

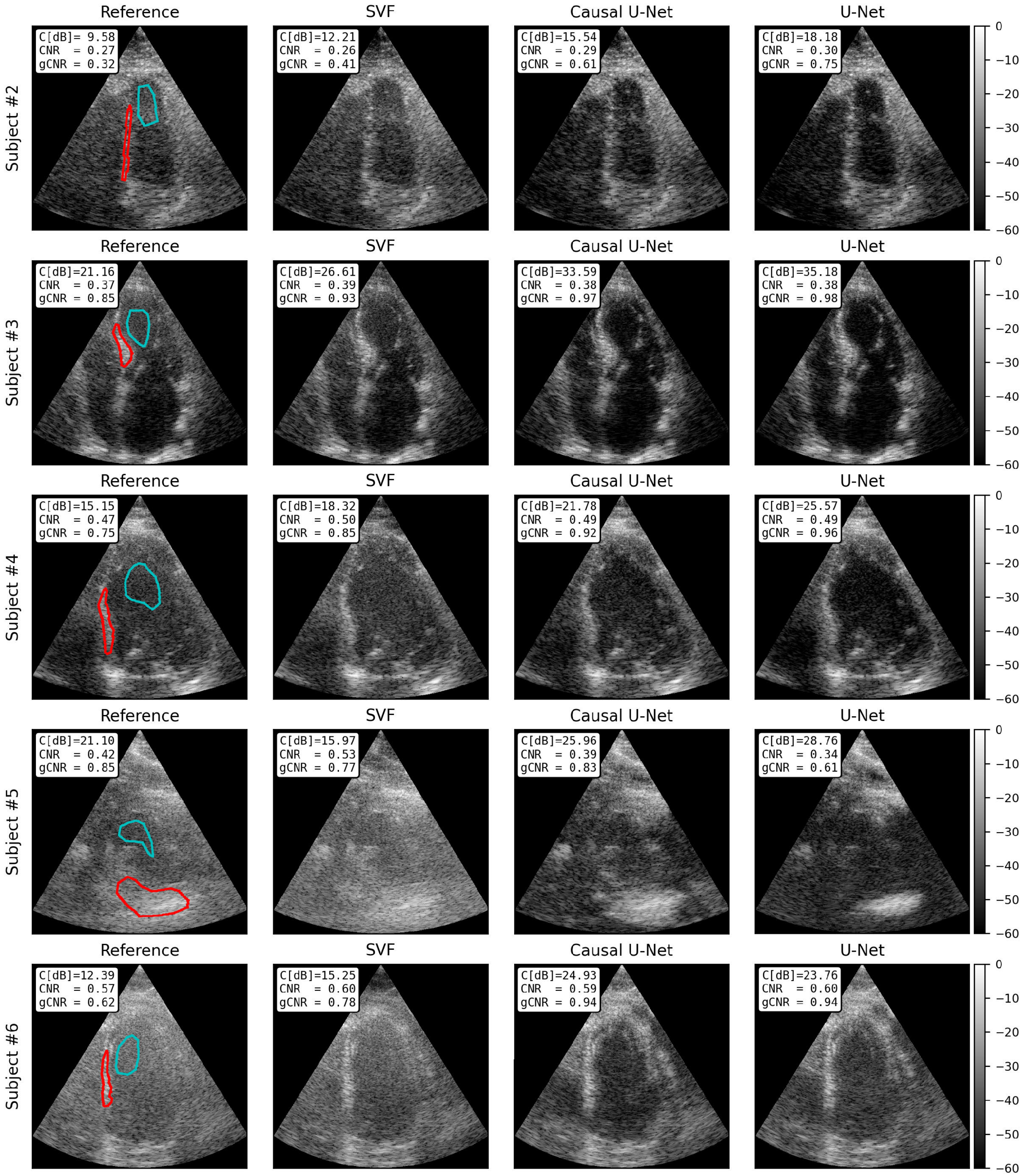

FIGURE 7. **Comparison between the reference images and the filtered images. The blue (hypo-echoic region) and red (cardiac structure) marked regions are used for calculating the image quality metrics. The images are displayed using a dynamic range of 60 dB.**

tered and reference signals is plotted in the bottom row. The filtered and reference signal phases seem more aligned for high-amplitude samples, but the phase differences are difficult to interpret elsewhere due to the numerous crossings. Overall, the filtered signals seem to lay one sample in front of the reference signals.

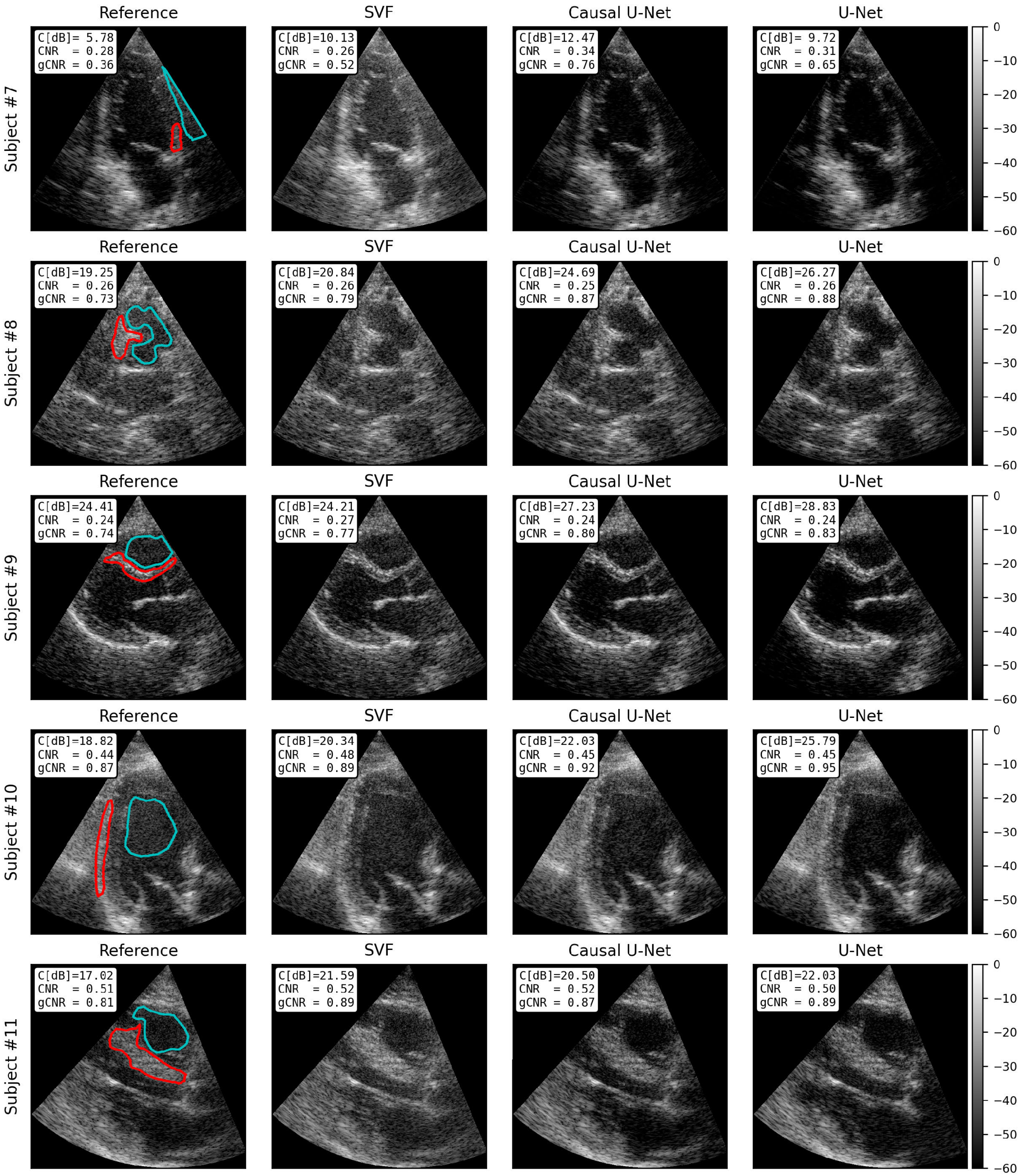

FIGURE 8. Comparison between the reference images and the filtered images. The blue (hypo-echoic region) and red (cardiac structure) marked regions are used for calculating the image quality metrics. The images are displayed using a dynamic range of 60 dB.

We also conducted an adversarial attack on the hazy recording from subject six and the clean recording from subject twelve, as described in section II-G. Subjects six and twelve were chosen as they represent the extremes of image quality. Figure 10 shows that the perturbation tensor **r** alters the input data for subject six such that the model does not filter the haze and instead removes some of the cardiac structures. This attack is considered effective as the input

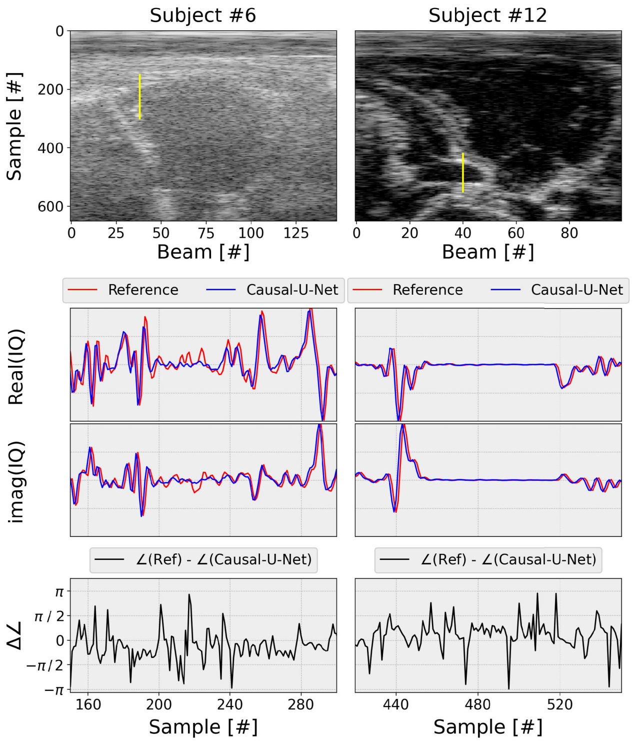

FIGURE 9. A comparison between the reference and filtered IQ data from Causal-U-Net. The top row shows the location of the traces, the middle row shows the real and the imaginary components, and the bottom row shows the phase differences.

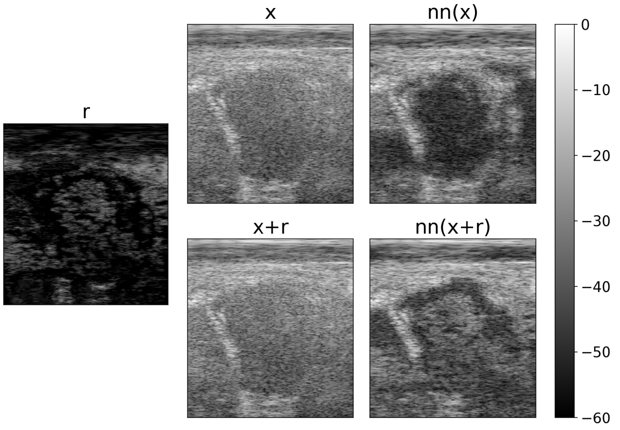

FIGURE 10. Results of an adverserial attack on *hazy* subject # six. The perturbation tensor is denoted as r, the input data x, and the Causal-U-Net as *nn*.

image shows little change. On the other hand, as shown in Figure 11, the attack on the clean recording from subject twelve was unsuccessful. The alteration of the input by the perturbation tensor resulted in a comparable change in the output.

IV. DISCUSSION

In this study, we have presented a novel approach to suppress unwanted reverberations in cardiac ultrasound imaging using deep learning. We have demonstrated that treating reverberation artifact suppression as a video separation problem can effectively leverage the network's ability to learn spatio-temporal features that are key to distinguishing between the two raw data components. Our method is based on a pseudo-paired dataset generated from *in vivo* data acquired from a high-end commercial scanner, which ensures that the clinical image quality is relevant and that the artifact generation is realistic. Additionally, we have shown that our method can be made causal with insignificant loss of impact, which is an important finding that can facilitate the adoption of the proposed method in real-time settings.

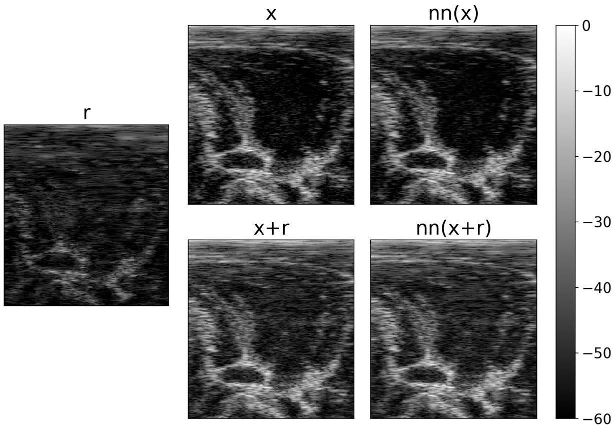

FIGURE 11. Results of an adverserial attack on *clean* subject # twelve. The perturbation tensor is denoted as r, the input data x, and the Causal-U-Net as *nn*.

A. EVALUATION OF CAUSAL-U-NET AND U-NET

The performance of U-Net and Causal-U-Net in enhancing the quality of hazy ultrasound image sequences has been evaluated. Evaluating image quality objectively is challenging, and we selected the frequently used metrics of contrast, CNR, and gCNR for our evaluation. The results presented in Table 4 are based on representative selections of patients, frames, and regions where the metrics were computed.

Overall, the two neural networks perform heavier filtering than the SVF method, as reflected by the increased contrast and gCNR values, typically resulting in improved detectability of structural details such as endocardial borders and papillary muscles that would otherwise be occluded. The neural networks are particularly effective in filtering non-stationary haze, and their output has superior temporal and spatial consistency compared to the SVF method, which often has a flickering appearance. This may be due to the networks' ability to track larger patches of the image over time, which helps preserve spatio-temporal consistency.

However, we also observed limitations in the performance of the neural networks. Haze close to the apex in the apical view is less suppressed than in other regions. We believe it occurs because this haze has a fluctuating component during the heart cycle, most likely caused by interactions with nearby structures. That type of haze is not modeled correctly in the pseudo-paired data set as a result of the random frame alignment between the *clean* reference and the *hazy* overlay. In addition, the number of *hazy* patches close to the apex is underrepresented in the training data set. The reason is their proximity to cardiac structures, which should not be included.

In the evaluation of image quality, we are limited to using metrics based on contrast as the ground truth clean data is unknown. The structural similarity index measure (SSIM) would have been helpful to grade, e.g., the preservation of speckles. The selected image quality metrics do not evaluate all aspects of image quality, such as resolution, temporal consistency, removal of cardiac structures, and generation of false cardiac structures. Therefore, visual assessment is considered the best solution for evaluating image quality.

The filtered output of subjects one to four is a good representation of the typical visual enhancement achieved using Causal-U-Net and U-Net on hazy cine-loops. On average, we find U-Net to perform a better haze suppression compared to Causal-U-Net, especially in cases of non-stationary haze. Both models exhibit failure in preserving cardiac structures that become stationary over a prolonged period of time. Subjects five, eight, and eleven are examples of this yet-to-be-solved challenge. Note that subjects five, eight, and eleven are specifically selected from the test set to illustrate this particular challenge and constitute worst-case examples. The suppression of cardiac structures is most prominent at end-diastole. In pathological cases like cardiomyopathy, portions of the myocardium can appear almost static, which may cause the models to suppress the myocardium. The training data set (clean) mainly includes data from the apical view (44/48), and the cardiac structures in the apical view are typically in motion. The phantom data used as *hazy* patches can also increase the effect of suppression of stationary objects and should be considered dropped. We believe this is the reason why the models have a tendency to suppress stationary objects.

Despite these limitations, none of the models show any tendency to generate false cardiac structures, which is crucial for a filter designed for clinical practice. This is seen in connection with the used loss function, as discussed in section IV-D. The models are not trained to make the cavities black, and blood flow can be seen in subjects four and ten. The rapid motion of the leaflets gives them low temporal consistency and should thus make them vulnerable to blurring. However, the models preserve the structure of the leaflets well, which is vital as the anatomy of the leaflets is of high clinical significance.

The recordings used in this study are from a high-end commercial system, already utilizing various methods for image enhancement. The proposed approach is a post-beam-former method complementary to other reverberation suppression methods, which typically do not include information across frames.

Finally, we tested the stability of the neural networks using an adversarial attack, which provides a semi-quantitatively evaluation of the worst-case scenario. We found that Causal-U-Net is stable in cases of clean images. However, in cases of cluttered data, the perturbation tensor was able to alter the clutter and revert the filtration. This is expected behavior and does not significantly affect the performance of Causal-U-Net.

Overall, our results demonstrate the potential of the Causal-U-Net and U-Net models as effective post-beam-former methods for enhancing the quality of hazy ultrasound images, and we believe that future work can further improve their performance.

B. MODELLING OF REVERBERATIONS

In this study, we created a pseudo-paired data set using *in vivo* data from subjects producing hazy and subjects producing clean images. This approach has the strength of providing realistic clutter, potentially covering all types and aspects of clutter generation, which is not possible with simulated data. Simulated reverberations are limited to those originating from within the simulation domain, and it is questionable whether a simulation can ever be made sufficiently realistic to be used for training purposes, given the complex and diverse mechanisms for reverberation clutter in cardiac imaging.

However, there are some limitations to our approach. Firstly, the cluttered patches' origin is limited to hypoechoic regions, and it was challenging to extract cluttered patches close to the apex without including cardiac structures. Secondly, reverberations created from interaction by nearby moving tissue are not modeled correctly. To alleviate this concern, we could sync the reference data and the superimposed data at specific cardiac events (e.g. end-diastole and end-systole) or use the information from a simultaneously recorded ECG trace.

Creating a pseudo-paired data set is not easily accessible, and we had to go through an extensive collection of recordings to acquire a reasonable amount of useful clean and cluttered recordings. Using the Site 1 database given in Table 2 as a reference, we found that only 21% of patients produce recordings useful for cluttered patch extraction, and 24% of patients produce recordings useful as clean reference images. Therefore, it is a practical challenge to capture data covering all aspects found in clinical settings, such as beam depth, heart rates, transmit frequency, beam separation, view, and variations within patents.

C. NETWORK STRUCTURE

The U-Net architecture was chosen due to its wide success in various image processing tasks such as segmentation [29], image-to-image translation [37], and super-resolution [38]. We used 3D convolutions instead of convolutional recurrent neural networks like ConvLSTM [39] as it allows for better parallelization during training.

Of the two models presented, Causal-U-Net is more practical for clinical use as it does not introduce a frame delay and has lower memory consumption by processing frames one at a time. U-Net was primarily used to evaluate the maximum haze suppression achievable with this class of methods. The structure of Causal-U-Net can also be used in a non-causal way, allowing the model to output a filtered frame at different frame delays. While introducing frame delays may be acceptable if it substantially improves haze suppression, the largest reasonable frame delay is half of the theoretical receptive field, making Causal-U-Net equivalent to U-Net.

Neural networks using causal convolutions often employ dilation to increase the receptive field, as seen in [40]. However, in our use case, we did not use dilation along the frame axis as the theoretical receptive field of 27 frames was assumed to be sufficient. Alternatively, a kernel of size two in injunction with dilation can be used to improve inference speed and memory consumption. The U-Net architecture's fully convolutional structure makes it flexible and can handle various spatial input sizes, which is useful as the number of beams and samples can vary between acquisitions.

D. LOSS FUNCTION

The Huber loss was selected first and foremost for its stability. During training, the L1 norm reduces the gradients compared to the squared L2 norm for inaccurate estimates. The pixel-wise regression loss is also less prone to generating false structures compared to other loss functions such as perceptual losses [41] and adversarial losses [42]. When the network is unsure of what to output, it outputs a blurred version of the structures. We used this as feedback when tuning the training setup. Critical hyperparameters were the value of the asymmetrical loss weight β and the use of logarithmic compression on the model's output. We found the filtered cine-loops made by a network trained using $\beta = 10$ and logarithmic compression to perform similarly to the presented configuration of $\beta = 5$ and no logarithmic compression. Although the filtered cine-loops looked similar, the values of the loss functions were different, making automated hyperparameter searches difficult.

E. FUTURE WORK

There are several areas where future work can be done to improve the performance of the neural network. One potential area of improvement is the data set. Including more diverse set of views could potentially mitigate the suppression of cardiac structures observed in the current data set. Another area of exploration is modifications to the neural network architecture. Dropping normalization layers and using pixel shuffling, which have shown success in the literature on super-resolution [43], [44], can be considered. Another promising alternative is converting the network to use complex-valued convolutions.

While the pixel-wise regression loss was chosen for its stability, including perceptual or adversarial type losses in the loss function can improve the visual aspects of the filtered cine-loops. Additionally, an alternative to applying logarithmic scaling on the output of the network prior to calculating the loss function, as been evaluated, is to apply logarithmic scaling on the input to the network. Reducing the large dynamic range within the neural network may be beneficial.

Furthermore, it is interesting to investigate the role of spatial and temporal information in separating clutter from tissue. Our visual system is capable of distinguishing reverberations from cardiac structures in cases of mild to medium levels of reverberations, but this process is more difficult in a static frame. One of our arguments for using neural networks is that combining spatial and time information is trivial compared to approaches based on SVD or wavelets. The methods using SVD often rely on change through time, whereas methods based on wavelet decomposition often rely on spatial information. Based on the present work, it is unclear whether spatial or time information is the predominant feature for separating clutter from tissue. Training networks only based on spatial or time information can provide insight into this question.

V. CONCLUSION

In this paper, we have presented an approach for generating a realistic pseudo-paired data set for the specific task of reverberation suppression. Our approach involves extracting patches from hypoechoic regions, obtained from subjects producing hazy videos, and overlaying them onto recordings from subjects producing haze-free videos. Furthermore, we demonstrate that both causal and non-causal convolutional neural networks can be trained on this pseudo-paired data set to perform reverberation suppression in cine-loops successfully. Our experimental results show that the post-beamformer method can enhance contrast, gCNR, and improve the visibility of important clinical structures on a high-end commercial system.

ACKNOWLEDGMENT

The authors would like to thank Børge Solli Andreassen for help in improving the manuscript and Jan Yee for the valuable discussions related to image quality.

REFERENCES

[1] G. F. Pinton, G. E. Trahey, and J. J. Dahl, "Erratum: Sources of image degradation in fundamental and harmonic ultrasound imaging: A nonlinear, full-wave, simulation study [Apr 11 754–765]," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 58, no. 6, pp. 1272–1283, Jun. 2011.

[2] A. Fatemi, E. A. R. Berg, and A. Rodriguez-Molares, "Studying the origin of reverberation clutter in echocardiography: In vitro experiments and in vivo demonstrations," *Ultrasound Med. Biol.*, vol. 45, no. 7, pp. 1799–1813, Jul. 2019.

[3] J. J. Dahl and N. M. Sheth, "Reverberation clutter from subcutaneous tissue layers: Simulation and in vivo demonstrations," *Ultrasound Med. Biol.*, vol. 40, no. 4, pp. 714–726, Apr. 2014.

[4] S. Choudhry, B. Gorman, J. W. Charboneau, D. J. Tradup, R. J. Beck, J. M. Kofler, and D. S. Groth, "Comparison of tissue harmonic imaging with conventional US in abdominal disease," *RadioGraphics*, vol. 20, no. 4, pp. 1127–1135, Jul. 2000.

[5] M. A. Lediju, M. J. Pihl, J. J. Dahl, and G. E. Trahey, "Quantitative assessment of the magnitude, impact and spatial extent of ultrasonic clutter," *Ultrason. Imag.*, vol. 30, no. 3, pp. 151–168, Jul. 2008.

[6] G. Zwirn and S. Akselrod, "Stationary clutter rejection in echocardiography," *Ultrasound Med. Biol.*, vol. 32, no. 1, pp. 43–52, Jan. 2006.

[7] B. C. Flynn, J. Spellman, C. Bodian, and V. K. Moitra, "Inadequate visualization and reporting of ventricular function from transthoracic echocardiography after cardiac surgery," *J. Cardiothoracic Vascular Anesthesia*, vol. 24, no. 2, pp. 280–284, Apr. 2010.

[8] M. A. Averkiou, D. N. Roundhill, and J. E. Powers, "A new imaging technique based on the nonlinear properties of tissues," in *Proc. IEEE Ultrason. Symp.*, Oct. 1997, pp. 1561–1566.

[9] F. Chirillo, A. Pedrocco, A. D. Leo, A. Bruni, O. Totis, P. Meneghetti, and P. Stritoni, "Impact of harmonic imaging on transthoracic echocardiographic identification of infective endocarditis and its complications," *Heart*, vol. 91, no. 3, pp. 329–333, Mar. 2005.

[10] S. P. Nasholm, R. Hansen, S.-E. Masoy, T. Johansen, and B. A. J. Angelsen, "Transmit beams adapted to reverberation noise suppression using dual-frequency SURF imaging," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 56, no. 10, pp. 2124–2133, Oct. 2009.

[11] J. M. Rau, S.-E. Masøy, R. Hansen, B. Angelsen, and T. A. Tangen, "Methods for reverberation suppression utilizing dual frequency band imaging," *J. Acoust. Soc. Amer.*, vol. 134, no. 3, pp. 2313–2325, Sep. 2013.

[12] K. W. Hollman, K. W. Rigby, and M. O'Donnell, "Coherence factor of speckle from a multi-row probe," in *Proc. IEEE Ultrason. Symp.*, Oct. 1999, pp. 1257–1260.

[13] P.-C. Li and M.-L. Li, "Adaptive imaging using the generalized coherence factor," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 50, no. 2, pp. 128–141, Feb. 2003.

[14] J. Shin, Y. Chen, H. Malhi, and J. T. Yen, "Ultrasonic reverberation clutter suppression using multiphase apodization with cross correlation," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 63, no. 11, pp. 1947–1956, Nov. 2016.

[15] M. A. Lediju, G. E. Trahey, B. C. Byram, and J. J. Dahl, "Short-lag spatial coherence of backscattered echoes: Imaging characteristics," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 58, no. 7, pp. 1377–1388, Jul. 2011.

[16] J. J. Dahl, M. Jakovljevic, G. F. Pinton, and G. E. Trahey, "Harmonic spatial coherence imaging: An ultrasonic imaging method based on backscatter coherence," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 59, no. 4, pp. 648–659, Apr. 2012.

[17] B. Byram, K. Dei, J. Tierney, and D. Dumont, "A model and regularization scheme for ultrasonic beamforming clutter reduction," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 62, no. 11, pp. 1913–1927, Nov. 2015.

[18] L. L. Brickson, D. Hyun, M. Jakovljevic, and J. J. Dahl, "Reverberation noise suppression in ultrasound channel signals using a 3D fully convolutional neural network," *IEEE Trans. Med. Imag.*, vol. 40, no. 4, pp. 1184–1195, Apr. 2021.

[19] A. C. Luchies and B. C. Byram, "Deep neural networks for ultrasound beamforming," *IEEE Trans. Med. Imag.*, vol. 37, no. 9, pp. 2010–2021, Sep. 2018.

[20] S. Bjaerum, H. Torp, and K. Kristoffersen, "Clutter filter design for ultrasound color flow imaging," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 49, no. 2, pp. 204–216, Feb. 2002.

[21] A. Yu and L. Lovstakken, "Eigen-based clutter filter design for ultrasound color flow imaging: A review," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 57, no. 5, pp. 1096–1111, May 2010.

[22] C. Demené, T. Deffieux, M. Pernot, B. Osmanski, V. Biran, J. Gennisson, L. Sieu, A. Bergel, S. Franqui, J. Correas, I. Cohen, O. Baud, and M. Tanter, "Spatiotemporal clutter filtering of ultrafast ultrasound data highly increases Doppler and fUltrasound sensitivity," *IEEE Trans. Med. Imag.*, vol. 34, no. 11, pp. 2271–2285, Nov. 2015.

[23] J. Baranger, B. Arnal, F. Perren, O. Baud, M. Tanter, and C. Demené, "Adaptive spatiotemporal SVD clutter filtering for ultrafast Doppler imaging using similarity of spatial singular vectors," *IEEE Trans. Med. Imag.*, vol. 37, no. 7, pp. 1574–1586, Jul. 2018.

[24] F. W. Mauldin, D. Lin, and J. A. Hossack, "The singular value filter: A general filter design strategy for PCA-based signal separation in medical ultrasound imaging," *IEEE Trans. Med. Imag.*, vol. 30, no. 11, pp. 1951–1964, Nov. 2011.

[25] S. J. Turek, M. Elad, and I. Yavneh, "Clutter mitigation in echocardiography using sparse signal separation," *Int. J. Biomed. Imag.*, vol. 2015, Jun. 2015, Art. no. 958963.

[26] M. Sjoerdsma, S. Bouwmeester, P. Houthuizen, F. N. Van De Vosse, and R. G. P. Lopata, "A spatial near-field clutter reduction filter preserving tissue speckle in echocardiography," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 68, no. 4, pp. 979–992, Apr. 2021.

[27] P. C. Tay, S. T. Acton, and J. A. Hossack, "A wavelet thresholding method to reduce ultrasound artifacts," *Computerized Med. Imag. Graph.*, vol. 35, no. 1, pp. 42–50, Jan. 2011.

[28] M. Tabassian, X. Hu, B. Chakraborty, and J. D'hooge, "Clutter filtering using a 3D deep convolutional neural network," in *Proc. IEEE Int. Ultrason. Symp. (IUS)*, Oct. 2019, pp. 2114–2117.

[29] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation,' in *Medical Image Computing and Computer-Assisted Intervention—MICCAI*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham, Switzerland: Springer, 2015, pp. 234–241.

[30] Y. Wu and K. He, "Group normalization," *Int. J. Comput. Vis.*, vol. 128, no. 3, pp. 742–755, Mar. 2020.

[31] A. L. Maas, A. Y. H. Annun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. ICML*, vol. 30, Atlanta, GA, USA, 2013, pp. 1–6.

[32] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *Computer Vision—ECCV*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Cham, Switzerland: Springer, 2016, pp. 630–645.

[33] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1026–1034.

[34] I. Loshchilov and F. Hutter, "Fixing weight decay regularization in Adam," 2017, *arXiv:1711.05101*.

[35] A. Rodriguez-Molares, O. M. H. Rindal, contrast-to-noisJ. D'hooge, S. Masøy, A. Austeng, M. A. L. Bell, and H. Torp, "The generalizede ratio: A formal definition for lesion detectability," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 67, no. 4, pp. 745–759, Apr. 2020.

[36] V. Antun, F. Renna, C. Poon, B. Adcock, and A. C. Hansen, "On instabilities of deep learning in image reconstruction and the potential costs of AI," *Proc. Nat. Acad. Sci. USA*, vol. 117, no. 48, pp. 30088–30095, Dec. 2020.

[37] P. Isola, J. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5967–5976.

[38] X.-J. Mao, C. Shen, and Y.-B. Yang, "Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections," in *Proc. 30th Int. Conf. Neural Inf. Process. Syst. (NIPS)*, Barcelona, Spain, vol. 29, D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, Eds. Red Hook, NY, USA: Curran, 2016, pp. 2810–2818

[39] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Proc. 28th Int. Conf. Neural Inf. Process. Syst. (NIPS)*, Montreal, QC, Canada, vol. 1. Cambridge, MA, USA: MIT Press, 2015, pp. 802–810.

[40] A. . D. Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu, "WaveNet: A generative model for raw audio," in *Proc. 9th ISCA Workshop Speech Synth. Workshop*, 2016, p. 125.

[41] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Computer Vision—ECCV*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Cham, Switzerland: Springer, 2016, pp. 694–711.

[42] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 105–114.

[43] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 1132–1140.

[44] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1874–1883.

TOLLEF STRUKSNES JAHREN was born in Oslo, Norway, in 1986. He received the M.Sc. degree in applied physics and mathematics from the Norwegian University of Science and Technology, Trondheim, Norway, in 2013. He is currently pursuing the joint Ph.D. degree in machine learning with the Department of Informatics, University of Oslo, Norway, in collaboration with GE Healthcare. From 2013 to 2016, he was with Petroleum Geo-Services, Oslo.

ANDERS RASMUS SØRNES was born in Sarpsborg, Norway, in 1970. He received the M.Sc. degree in physics and the Ph.D. degree in magnetic resonance from the University of Oslo, in 1996 and 1998, respectively. From 1998 to 2000, he was with the Equinor/Statoil Research Centre, Trondheim, with problems in seismic processing and inversion. Since 2000, he has been with GE Vingmed Ultrasound AS, where he today is architect responsible for image quality development in cardiovascular ultrasound.

BASTIEN DÉNARIÉ received the M.Sc. degree in engineering cybernetics and the Ph.D. degree in high frame-rate echocardiography from the Norwegian University of Science and Technology (NTNU), Trondheim, Norway, in 2010 and 2014, respectively. He is currently the Manager of External Imaging Research with the GE HealthCare Cardiovascular Ultrasound Business (GE Vingmed Ultrasound). During the last decade, he was instrumental in developing a software beamforming platform that is now at the core of GE Vivid ultrasound scanners, revolutionizing high frame-rate echocardiography and enabling clinicians to detect heart disease with unprecedented accuracy. His research interests include developing novel algorithms to optimize frame rate and enhance image resolution in echocardiography.

ERIK STEEN received the Ph.D. degree in medical image processing and visualization from the Norwegian University of Science and Technology (NTNU), Trondheim, Norway, in 1996. He joined GE Vingmed Ultrasound, Horten, Norway, in 1996. He has his daily work with GE Vingmed Ultrasound. He is currently a Chief Engineer with the GEHC Cardiovascular Ultrasound Business (GE Vingmed Ultrasound). He became the leader of a group specializing in image processing and visualization, in 1998. Then, he worked on the design of a software-based image processing pipeline that became an integral part of the Vivid 7 scanner launched, in 2001, and was later migrated to a large number of GEHC ultrasound scanners. He became a principal engineer, in 2006, and a chief engineer, in 2018. In later years, he has particularly focused on utilizing modern graphics processors to accelerate image reconstruction, processing, AI, and visualization algorithms. The results of this work are integrated into the premium cardiovascular 3-D ultrasound system Vivid E95. For the last three years, he has been leading a research and development program focusing on deep learning in cardiac ultrasound. The program resulted in increased efficiency and diagnostic accuracy in everyday use of the ultrasound scanners.

TORE BJÅSTAD was born in Ålesund, Norway, in 1977. He received the M.Sc. degree in electronics and telecommunication and the Ph.D. degree in high frame rate imaging in cardiac ultrasound from the Norwegian University of Science and Technology, Trondheim, Norway, in 2008. He has since mainly worked with GE Vingmed Ultrasound AS on methods for improved image quality in cardiac and fetal imaging. In 2017 and 2022, he was with consultancy firm InPhase Solutions AS, Trondheim, Norway, where he also was involved in industrial application of ultrasound like f.ex. NDT. In this period, he also held a position with NTNU facilitating testing and implementation of algorithms for faster transition from research to product feature.

ANNE H. SCHISTAD SOLBERG received the M.Sc. degree in computer science and the Ph.D. degree in image analysis from the University of Oslo, Norway, in 1989 and 1995, respectively. She is currently a Professor with the Digital Signal Processing and Image Analysis Group, Department of Informatics, University of Oslo. She has worked on a broad range of machine learning applications within image analysis. Her research interests include medical imaging, remote sensing, sonar imaging, ultrasound imaging, seismic imaging, feature extraction, and machine learning.