Program Digest, 2019 IEEE International Ultrasonics Symposium (IUS)
Glasgow, Scotland, October 6-9, 2019

Clutter Filtering Using a 3D Deep Convolutional Neural Network

Mahdi Tabassian, XingRan Hu, Bidisha Chakraborty, and Jan D'hooge
Department of Cardiovascular Sciences, KU Leuven, Leuven, Belgium
mahdi.tabassian@kuleuven.be

***Abstract*—Echocardiographic imaging is a well-established clinical modality to assess cardiac morphology and function. However, accuracy of visual reading as well as performance of algorithms designed to measure cardiac characteristics, and as a consequence the diagnostic value of the echocardiographic imaging, can significantly be degraded by the presence of image artifacts such as (stationary) reverberations. Although several filtering techniques have been proposed in the literature to tackle this problem, the current study sought to investigate whether deep learning can offer better artifact removal. Given the spatiotemporal nature of the reverberation artifacts, a 3D (i.e. 2D plus time) deep convolutional neural network (3D U-Net) was trained to filter simulated artifacts that were superimposed onto ultra-realistic synthetic 2D echocardiographic sequences. Performance of 3D U-Net was compared with the singular value decomposition (SVD) technique. The obtained results confirmed the utility of deep learning for artifact filtering and showed its advantage over SVD in handling moving artifacts.**

I. Introduction

Transthoracic echocardiography has become the primary noninvasive imaging modality to quantify myocardial morphology and function. Nevertheless, the diagnostic value of this imaging technique can negatively be influenced by acoustic clutter and especially the reverberation artifact. This type of clutter has a spatiotemporal nature as it is mainly generated by some of the slow-moving organs like the ribs and lungs.

Several methods have been proposed in the literature for clutter rejection that mainly work by linear decomposition of the acquired echo data into clutter and signal of interest components using a set of bases. These bases can be defined a priori or be learned from the data. The discrete Fourier transform [1] and the wavelet transform [2] are examples of the clutter filtering methods that use pre-defined bases. Singular value decomposition (SVD) is the most widely used approach to learn bases for clutter filtering [3], [4] but other dictionary learning techniques like K-SVD [5] and morphological component analysis [6] have been also used for this purpose.

Compared to the methods that use pre-defined bases for clutter rejection, the learning strategies have the advantage of adapting their bases to the data characteristics enabling them to better filter clutter artifacts. Nonetheless, the learning strategies that are used in the SVD-based filtering methods have some limitations that hamper their efficient operation. Linear data modeling, lack of hierarchical representation of the data, using a relatively small set of the bases for decomposing the data and regional filtering are some of these limitations.

These limitations can be circumvented by using a deep convolutional neural network (CNN) that provides hierarchical representation of the data based on a non-linear combination of a lot of bases/kernels while taking global characteristics of the data into account.

As such, CNNs have been recently used in several studies as a sophisticated image processing tool in order to enhance the quality of the ultrasound images. In [7], [8] CNNs have been employed in the structure of a generative adversarial network for despeckling of the ultrasound images. Perdios *et al.* [9] used CNNs to learn a mapping between low- and high-quality subspaces of RF images to enhance the images reconstructed from a single plane wave. A 3D CNN architecture was trained in [10] to remove reverberation and thermal noise from ultrasound channel data.

Inspired by the promising performance of CNNs in improving the quality of the ultrasound data, a 3D CNN framework is used in the current study for spatiotemporal clutter filtering of 2D echocardiographic B-mode sequences. The rationale behind using a 3D CNN for clutter filtering is to account for the spatiotemporal nature of this type of artifact. Given that reverberation is mainly generated by some of the slow-moving organs like the ribs and lungs, it affects 2D B-mode images throughout the cardiac cycle resulting in artifact patterns that slowly move in time. Hereto, a 3D CNN can be designed in such a way that learns the behavior of the clutter both in space and time in order to suppress it efficiently.

The remainder of this paper is organized as follows. Section II describes the data, architecture of the employed 3D CNN and its training details. Results are presented in Section III and the obtained results are discussed in Section IV. Finally, Section V draws conclusions and summarizes the paper.

II. Materials and Methods

A. Data

In order to learn how to remove clutter from an input sequence of 2D echocardiographic images, a 3D CNN should be provided with a corresponding clutter-free sequence as its output. It is thus of paramount importance to use output sequences that are completely free of artifacts in order to make sure that the network learns well how to differentiate clutter from signals of interest. With this in mind, we used a database of ultra-realistic synthetic 2D echocardiographic sequences [11] in our experiments. The database consisted of 30 apical 4-chamber sequences of six vendors each with five

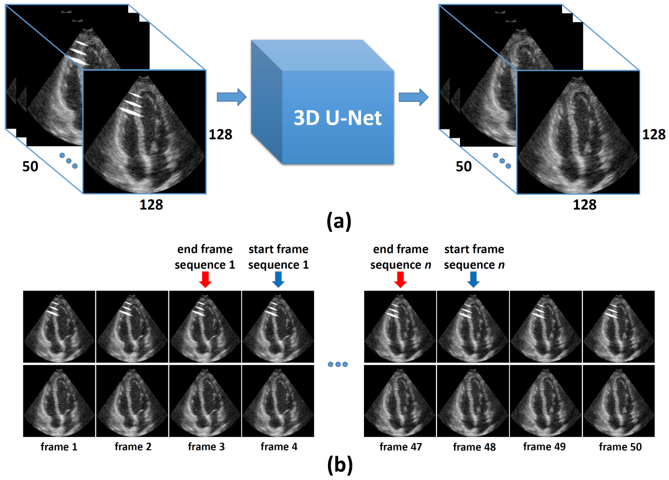

Fig. 1. (a) Schematic illustration of training a 3D U-Net with cluttered 2D echocardiographic sequences as input to the network and their corresponding clutter-free sequences as its output. All 50 frames of a sequence were fed into the input and the temporal information of this sequence was preserved during the encoding phase. (b) In order to make the network robust to the starting point of the sequence in the cardiac cycle, multiple input-output sequences were generated from a given sequence by shifting the start (showed with blue arrows) and end frames (showed with red arrows) in time.

electro-mechanical models of a heart including one normal and four ischemic motion patterns. Each subject had one simulated heart beat with 50 2D images resized to 128×128. In order to generate artifactual 2D sequences, clutter artifacts were heuristically superimposed as three bright ellipsoidal regions with small, medium and large sizes onto the left and right edges of the sector images. To simulate slow-moving clutter, the ellipsoids were moved by one pixel every four frames in the radial direction.

B. 3D CNN Architecture & Training

Motivated by the successful application of the deep auto-encoders in handling different image processing tasks, an efficient convolutional auto-encoder network called 3D U-Net [12] was used for clutter filtering in our study. The architecture of the network was similar to that of the original 3D U-Net except for the stride and max pooling of the time dimension that was set to one. As a result, the temporal information was preserved in the encoding phase enabling the network to learn the behavior of the clutter patterns throughout the whole cardiac cycle (Fig. 1(a)).

The network was trained with the artifactual input sequences, each of size $128 \times 128 \times 50$, and their corresponding artifact-free sequences. In order to train a network that is robust to the starting point of the sequence in the cardiac cycle, multiple input-output sequences were created by shifting each sequence pair in time such that the starting frames were taken from different time points during the cardiac cycle (Fig. 1(b)). As a result of this process, the size of the database was also increased by a factor of 10.

TABLE I
MEAN AND STANDARD DEVIATIONS OF THE RMADS COMPUTED FROM THE CLUTTER-FREE AND CLUTTERED FRAMES AS WELL AS CLUTTER-FREE AND CLUTTER- FILTERED FRAMES FOR ALL FRAMES OF THE TESTING SUBJECTS.

Clutt.-free vs Clutt.	Clutt.-free vs Clutt.-filt. 3D U-Net	Clutt.-free vs Clutt.-filt. SVD
18.64 ± 4.88	7.93 ± 1.43	25.77 ± 4.76

The leave-one-out cross-validation technique was used to train and validate the clutter filtering network. The training dataset was augmented by applying rotation, horizontal/vertical flip and height/width shift transformations to the sequences where all 2D B-mode images belonging to an input-output sequence pair were transformed similarly. The network was trained using the Adam optimizer with a learning rate of 10^{-4}, mean-squared-error as loss function and 30 epochs using the Keras library with the TensorFlow backend and one GPU (NVIDIA Titan V, 11GB VRAM).

C. SVD filtering

In order to compare the filtering performance of 3D U-Net with the classical learning-based filtering methods, the SVD filter was implemented based on the multi-ensemble approach using a region of interest (ROI) of size 10×10 pixels [3]. An eigenvector was included in the set of clutter eigencomponents if the ratio of its eigenvalue to the most dominant eigenvalue was larger than a threshold value of 7.5%.

III. RESULTS

Examples of the clutter-free and cluttered frames of the employed vendors and their corresponding clutter-filtered frames using 3D U-Net and SVD are shown in Fig. 2. The zoomed-in clutter-filtered regions show the promising performance of 3D U-Net in rejecting the clutter patterns and reconstructing the pixels behind them. Although the SVD filter could, to some extend, suppress the cluttered regions, it failed to reconstruct the clutter-free pixels and led to extended artifactual regions.

The filtering performance of 3D U-Net and SVD were quantitatively assessed by computing a regional-mean-absolute-difference (RMAD) metric as follows:

$$RMAD = \frac{\sum_{i=1}^{N} |x_i - \hat{x}_i|}{N} \tag{1}$$

In (1), $\hat{x}_i$ is the grayscale value of the ith pixel inside a cluttered or clutter-filtered region with N pixels (shown by a red rectangular in Fig. 2) and x_i is the corresponding grayscale value of that pixel in a clutter-free image. This metric was computed for all frames of a given sequence and for all testing subjects and the obtained mean and standard deviation values are listed in Table I.

As can be seen from Table I, utilizing 3D U-Net yielded a mean RMAD value of 7.93 between the clutter-free and

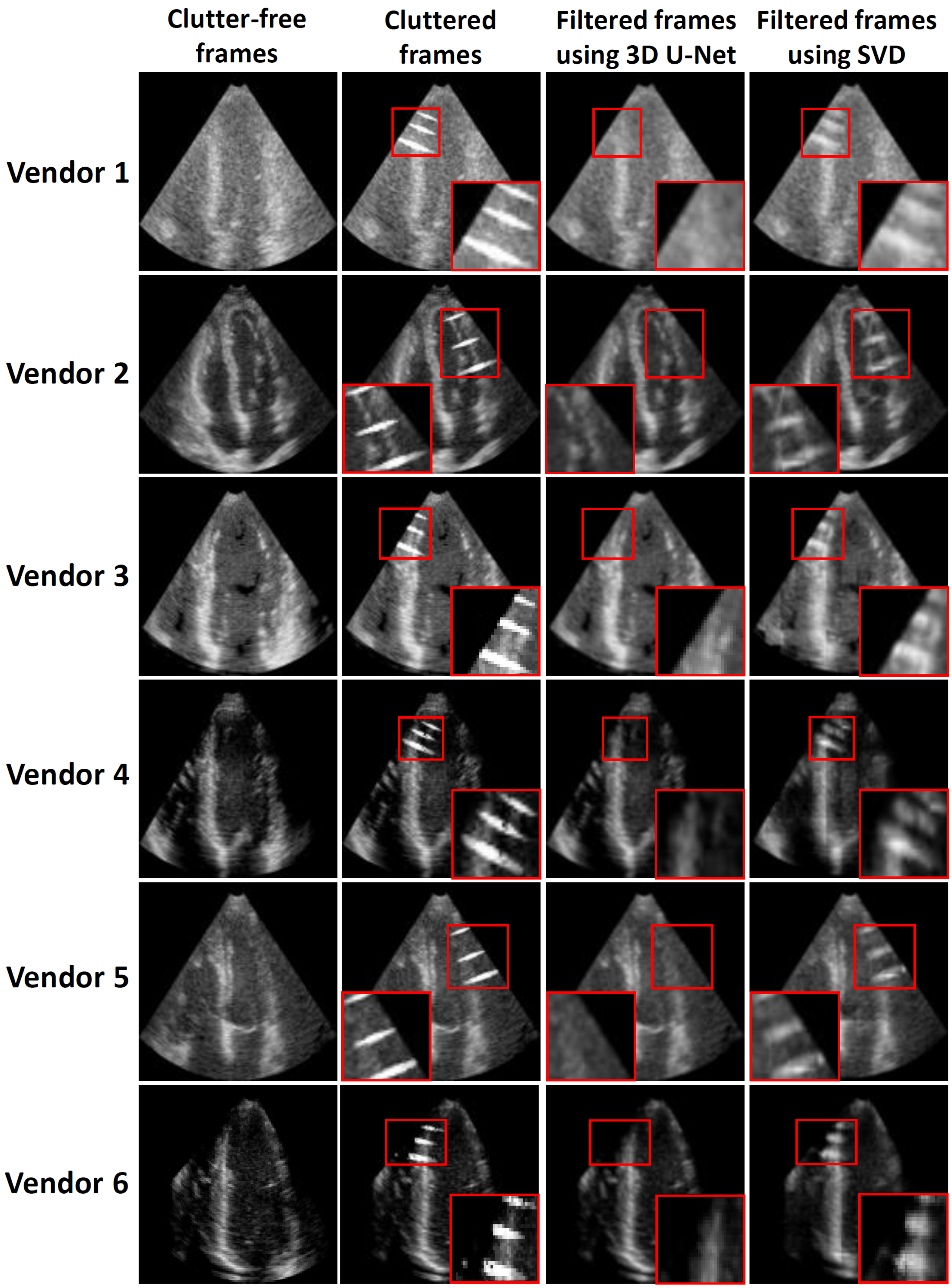

Fig. 2. Examples of the clutter-free and cluttered frames from the six vendors that were used in our experiments and their corresponding clutter-filtered images generated by 3D U-Net and SVD. Zoomed-in versions of the cluttered/clutter-filtered regions are shown for an easy evaluation of the filtering performance.

clutter-filtered regions which was significantly smaller ($p < 0.001$) than the mean RMAD value of 18.64 between the clutter-free and cluttered regions. The SVD filter, on the other hand, could not reduce the difference between the clutter-free and cluttered regions.

IV. Discussion

The obtained results confirm the suitability of 3D U-Net for spatiotemporal clutter rejection and show its advantage over the SVD filtering approach. The efficient performance of 3D U-Net can be explained by considering the following characteristics of this deep neural network.

First, unlike SVD that filters a small ROI, 3D U-Net takes both the local and global characteristics of the 2D image sequences into account to reject the clutter patterns by taking advantage of hierarchical representation of the images through using convolutional and pooling layers. As such, the clutter location and its relation to other regions in the myocardium can help the network to reconstruct the artifactual regions and to predict the patterns of the clutter-free pixels that should be substituted in place of the cluttered ones. This can be seen by considering the clutter-filtered regions in Fig. 2 generated by 3D U-net where the reconstructed pixels properly represent the clutter-free regions.

Second, 3D U-Net learns a large set of kernels for encoding spatiotemporal features of the 2D sequences and uses a non-linear combination of them in order to filter the cluttered regions. It is, thus, more efficient than SVD in decomposing the signal of interest and clutter components in space and time given that SVD linearly combines a limited set of bases for this purpose. In Fig. 2, the bright regions in the images filtered by SVD are bigger than the superimposed clutter patterns. This resulted in a mean RMAD value for SVD that is larger than that of the clutter-free and cluttered frames (Table I) which can be explained by the failure of SVD in properly decomposing the slow-moving clutter components in the time domain.

Although 3D U-Net yielded promising clutter rejection results, its generalization performance is heavily dependent on the diversity of the clutter patterns that it learned during the training phase. Hereto, a major future step is to simulate different types of clutter artifacts that represent real-world scenarios in order to train a capable deep clutter filtering tool. Training the 3D network with sequences of different echocardiographic views and evaluating the performance of the network on in-vivo data are other important future works that could be accomplished.

V. Conclusions

In this study, we have proposed to use a 3D deep CNN called 3D U-Net for the spatiotemporal clutter filtering of 2D echocardiographic B-mode sequences. The network was trained with simulated slow-moving clutter artifacts and was designed such that it was able to learn the temporal behavior of the clutter patterns throughout the whole cardiac cycle. The performance of the deep network in rejecting clutter patterns both in space and time was promising and highlights its advantage over the classical learning-based filtering methods like SVD. The clinical implication of these results is significant given that efficient filtering of the echocardiographic clutter artifacts could dramatically increase the diagnostic value of this imaging modality.

VI. Acknowledgment

We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan V GPU used for this research and Prof. Piet Claus for providing us with the workstation of his Lab.

References

[1] 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, 2002.

[2] P. C. Tay, S. T. Acton, and J. A. Hossack, "A wavelet thresholding method to reduce ultrasound artifacts," *Computerized Medical Imaging and Graphics*, vol. 35, no. 1, pp. 42–50, 2011.

[3] C. Alfred 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, 2010.

[4] 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, 2011.

[5] J. S. Turek, M. Elad, and I. Yavneh, "Sparse signal separation with an off-line learned dictionary for clutter reduction in echocardiography," in *IEEE Convention of Electrical & Electronics Engineers*, 2014, pp. 1–5.

[6] ——, "Clutter mitigation in echocardiography using sparse signal separation," *Journal of Biomedical Imaging*, pp. 1–18, 2015.

[7] D. Mishra, S. Chaudhury, M. Sarkar *et al.*, "Ultrasound image enhancement using structure oriented adversarial network," *IEEE Signal Processing Letters*, vol. 25, no. 9, pp. 1349–1353, 2018.

[8] F. Dietrichson, E. Smistad, A. Ostvik *et al.*, "Ultrasound speckle reduction using generative adversial networks," in *IEEE Int. Ultrason. Symp. (IUS)*, 2018, pp. 1–4.

[9] D. Perdios, M. Vonlanthen, A. Besson *et al.*, "Deep convolutional neural network for ultrasound image enhancement," in *IEEE Int. Ultrason. Symp. (IUS)*, 2018, pp. 1–4.

[10] L. L. Brickson, D. Hyun, and J. J. Dahl, "Reverberation noise suppression in the aperture domain using 3d fully convolutional neural networks," in *IEEE Int. Ultrason. Symp. (IUS)*, 2018, pp. 1–4.

[11] M. Alessandrini, B. Chakraborty, B. Heyde *et al.*, "Realistic vendor-specific synthetic ultrasound data for quality assurance of 2-d speckle tracking echocardiography: Simulation pipeline and open access database," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 65, no. 3, pp. 411–422, 2017.

[12] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp *et al.*, "3d u-net: learning dense volumetric segmentation from sparse annotation," in *Int. Conf. on Medical Image Comp. and Computer-assisted Interv.* Springer, 2016, pp. 424–432.