



HAL
open science

Boosting Cardiac Color Doppler Frame Rates with Deep Learning

Julia Puig, Denis Friboulet, Hang Jung Ling, François Varray, Jonathan Porée, Jean Provost, Damien Garcia, Fabien Millioz

► **To cite this version:**

Julia Puig, Denis Friboulet, Hang Jung Ling, François Varray, Jonathan Porée, et al.. Boosting Cardiac Color Doppler Frame Rates with Deep Learning. 2024. hal-04495325

HAL Id: hal-04495325

<https://hal.science/hal-04495325v1>

Preprint submitted on 27 Mar 2024

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

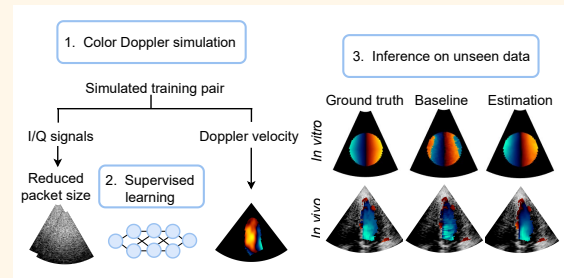
L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Boosting Cardiac Color Doppler Frame Rates with Deep Learning

Julia Puig, Denis Friboulet, Hang Jung Ling, François Varray, Jonathan Porée, Jean Provost, Damien Garcia and Fabien Millioz

Abstract—Color Doppler echocardiography enables visualization of blood flow within the heart. However, the limited frame rate impedes the quantitative assessment of blood velocity throughout the cardiac cycle, thereby compromising a comprehensive analysis of ventricular filling. Concurrently, deep learning is demonstrating promising outcomes in post-processing of echocardiographic data for various applications. This work explores the use of deep learning models for intracardiac Doppler velocity estimation from a reduced number of filtered I/Q signals. We used a supervised learning approach by simulating patient-based cardiac color Doppler acquisitions and proposed data augmentation strategies to enlarge the training dataset. We implemented architectures based on convolutional neural networks. In particular, we focused on comparing the U-Net model and the recent ConvNeXt models, alongside assessing real-valued versus complex-valued representations. We found that both models outperformed the state-of-the-art autocorrelator method, effectively mitigating aliasing and noise. We did not observe significant differences between the use of real and complex data. Finally, we validated the models on *in vitro* and *in vivo* experiments. All models produced quantitatively comparable results to the baseline and were more robust to noise. ConvNeXt emerged as the sole model to achieve high-quality results on *in vivo* aliased samples. These results demonstrate the interest of supervised deep learning methods for Doppler velocity estimation from a reduced number of acquisitions.

Index Terms—Color Doppler, Aliasing mitigation, Echocardiography, Deep learning, ConvNeXt, Ultrasound simulations



I. INTRODUCTION

COLOR Doppler echocardiography facilitates the concurrent visualization of cardiac tissue and intracardiac blood movement. In clinical settings, this imaging modality may serve as a visual aid to diagnose some cardiac pathologies. However, the frame rate currently available is insufficient to quantitatively assess blood velocities during a cardiac cycle. Increasing the temporal resolution of color Doppler could potentially open up new possibilities for diagnosing diastolic impairment through the use of quantitative analysis tools [1].

To generate a color Doppler image, backscattered echoes are acquired in the axial direction (fast-time) at n consecutive instants (slow-time). The radio frequency (RF) signals undergo

I/Q demodulation, followed by clutter filtering and beamforming. The number n of slow-time acquisitions is referred to as the packet size. They enable the estimation of the phase shifts along the slow-time axis, from which Doppler velocities are subsequently derived. Phase shifts are usually estimated in two steps [2]. First, a clutter filter is applied to the I/Q signals to eliminate signals corresponding to tissues and slow-moving artifacts. Then, spatially weighted average autocorrelations are computed to estimate the local phase shifts [3]. In practice, the packet size typically ranges around $n = 8$. The reliability of the autocorrelation-based estimates significantly relies on the packet size; its reduction compromises robustness to noise and other artifacts.

In addition to noise, color Doppler can suffer from aliasing, which occurs when velocities exceed the Nyquist limit set by the pulse repetition frequency (PRF). While limited research has been dedicated to dealiasing in echocardiography, it is an essential step in extracting reliable quantitative information from color Doppler. In the context of ultrafast imaging, Posada *et al.* [4] proposed the use of staggered multi-PRF emissions to provide alias-free Doppler velocities. Dealiasing has also been addressed as a post-processing problem. Muth *et al.* [5] developed the DeAN algorithm for denoising and dealiasing color

Manuscript submitted for review March 7, 2024.

This work was performed within the framework of the LABEX PRIMES (ANR-11-LABX-0063) of Université de Lyon, within the program "Investissements d'Avenir" (ANR-11-IDEX-0007) operated by the French National Research Agency (ANR).

J. Puig, D. Friboulet, H. J. Ling, F. Varray, D. Garcia and F. Millioz are with Univ Lyon, INSA-Lyon, Université Claude Bernard Lyon 1, CNRS, Inserm, CREATIS UMR 5220, U1294 69100 Villeurbanne, France (e-mail: julia.puig@creatis.insa-lyon.fr, fabien.millioz@creatis.insa-lyon.fr).

J. Porée and J. Provost are with the Department of Engineering Physics, Polytechnique Montréal, Montréal, QC H3T 1J4. J. Provost is also with the Montreal Heart Institute, Montréal, QC H1T 1C8, Canada.

Doppler velocity maps. Their method involves unsupervised segmentation of aliased regions and subsequent dealiasing through comparison with neighboring regions. However, the method is prone to failure in cases of deteriorated signal quality.

In recent years, deep learning has led to significant advancements in medical imaging by replacing traditional model-based methods with learning-based approaches. In particular, convolutional neural networks (CNNs) have emerged as pivotal tools in ultrasound imaging, either by introducing entirely learning-based models or by enhancing the performance of existing model-based solutions with a deep learning step [6].

Few works have addressed Doppler velocity estimation with deep learning. For cardiac tissue Doppler estimation, van Sloun *et al.* [7] used an encoder-decoder on the I/Q signals to replace the classical autocorrelator method. Utilizing *in vivo* I/Q data from a porcine model, they derived Doppler estimates that achieved the quality of the autocorrelator method with reduced noise levels. In cardiac color Doppler, Apostolakis *et al.* [8] introduced a serial U-Net designed to estimate Doppler velocities from wall-filtered RF signals with reduced ($n = 4$) or undersampled packet sizes. They used a training dataset of flow phantom acquisitions whose reference velocities were given by an autocorrelator with a packet size of $n = 14$. Lei *et al.* [9] proposed a complex-valued CNN dedicated to carotid blood flow velocity estimation from RF signals, which simultaneously addressed clutter removal and velocity estimation, demonstrating promising results on simulated carotid data and *in vivo* experiments. Regarding clutter filtering, Solomon *et al.* [10] proposed a deep unfolding algorithm for super-resolution ultrasound applications. They formulated the problem as an iterative Robust Principal Component Analysis (RPCA) and solved it using a deep learning algorithm. In the same context, Brown *et al.* [11] performed clutter filtering on both *in vitro* and *in vivo* data using a spatio-temporal 3D CNN.

Regarding dealiasing, in 2020 Nahas *et al.* [12] trained a U-Net to identify and segment aliased regions in femoral bifurcation color Doppler images using Doppler frequency, power and bandwidth information. Recently, they proposed an accelerated dealiasing technique for high-frame-rate vector Doppler imaging in femoral imaging [13]. Alternatively, Ling *et al.* [14] compared a deep unfolding method with a data-driven deep learning approach for dealing with aliasing in cardiac color Doppler images. The data-driven approach demonstrated superior results, with both methods outperforming the DeAN method mentioned above. All these dealiasing works can be considered as post-processing methods applied to Doppler velocity maps.

In this study, we propose a deep learning approach as an alternative to the current autocorrelator technique in the context of focused cardiac color Doppler acquisitions. Our goal was to obtain accurate Doppler velocities from clutter-filtered I/Q signals with a reduced packet size of $n = 2$. We contended that convolutional neural networks can identify spatial and channel-wise connections that the autocorrelator overlooks, potentially compensating for the packet size reduction. Our main contributions are as follows:

1) We generated a training dataset for supervised learning

by simulating color Doppler echocardiography with a dedicated pipeline. It included customized data augmentation techniques to increase the dataset size.

- 2) We implemented state-of-the-art deep learning architectures for Doppler velocity estimation from I/Q signals with a reduced packet size. We explored the performance of U-Net and ConvNeXt architectures, and the impact of using real-valued versus complex-valued representations.
- 3) We validated the proposed deep learning models using both *in vitro* and *in vivo* sequences. These experiments showed that the deep learning models achieved comparable results to the autocorrelator while exhibiting robustness to noise and inherent reduction of aliasing.

II. METHODS

A. Doppler velocity estimation in color Doppler

In ultrasound imaging, the received RF signals undergo I/Q-demodulation and beamforming. In the context of Doppler imaging, these resulting complex-valued signals can be organized in arrays of shape $h \times w \times n$, where h and w represent the fast-time and lateral dimensions, respectively, and n denotes the slow-time dimension referred to as the packet size. These processed signals are filtered to remove clutter and retrieve blood-related signals. Blood Doppler velocities are then estimated by calculating the phase shifts of the filtered signals in the slow-time direction, *i.e.* along the third dimension of the array. These phase shifts can be derived by evaluating the argument of the lag-one autocorrelator (R_1) applied to the I/Q signals. We adopted the 2-D autocorrelator introduced by Loupas *et al.* [3] as the baseline method. We smoothed R_1 spatially with a Hamming kernel of size 10×4 pixels to obtain a smoothed autocorrelation R_1^s . Finally, the Doppler velocities were computed by converting the arguments to velocities as follows:

$$v_D = \frac{v_N}{\pi} \arg(R_1^s),$$

where v_N stands for the Nyquist velocity. The reliability of this technique depends on the packet size number, and decreasing it can degrade the estimation quality, especially when dealing with noisy acquisitions.

B. Color Doppler simulations

Training supervised deep learning models requires large annotated datasets. In the context of color Doppler, high-quality *in vivo* ground truth is not available. Therefore, it is necessary to rely on simulations to build a supervised training dataset. In this study, we adapted the patient-based color Doppler simulator developed by Sun *et al.* [15]. A schematic representation of the inputs and outputs of the simulation pipeline is illustrated in Figure 1. Given the patient's B-mode and Doppler information along a cardiac sequence, the simulation pipeline consisted of three main steps.

Highlights

- We explored the use of deep learning for generating color Doppler images from clutter-filtered echocardiographic I/Q signals.
- We proposed a deep learning model that estimates phase shift from I/Q signals and mitigates aliasing from a reduced packet size. The model was trained solely on *in silico* data and validated on both *in vitro* and *in vivo* sequences.
- It was found that deep learning models can inherently reduce aliasing and are robust to noise induced by a reduced packet size.

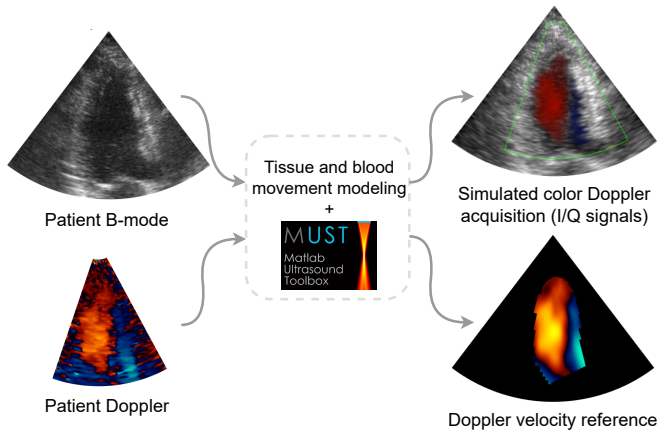


Fig. 1. Inputs and outputs of the color Doppler simulation pipeline.

1) *Modeling cardiac tissue motion*: For each patient’s B-mode image, the inner and outer walls of the left ventricular myocardium were segmented using a deep learning method [16]. A myocardial mesh was generated from the segmentation masks. Tissue scatterers were randomly distributed in the end-systole frame. Their positions were then propagated frame by frame based on the motion of the myocardial mesh. For additional information, refer to [15].

2) *Modeling blood flow motion*: Patient Doppler velocity frames were treated as independent due to the low frame rate of color Doppler. For each color Doppler frame, a 2-D intraventricular vector flow map was computed from the patient’s Doppler velocities using a physics-based method [17]. For each frame, blood scatterers were distributed in the intraventricular region. Series of [number of firings \times n] time steps were defined, during which the blood scatterers were displaced according to the 2-D velocity maps. To enhance realism, fluctuation values were computed from velocity maps and added to velocity values. This process was repeated for all frames of the sequence. From the resulting vector outputs, we used the radial components to build the ground truth Doppler velocity maps $D \in \mathbb{R}^{h \times w}$ in our training dataset.

3) *Simulation of an ultrasound acquisition*: From the tissue and blood scatterer position maps defined at each time step, scanline-based ultrasound acquisitions were simulated using SIMUS [18], [19] of the Matlab UltraSound Toolbox (MUST) [20]. The RF signals were post-processed conventionally with I/Q-demodulation and delay-and-sum beamforming on a $h \times w$ grid. The resulting I/Q signals $S \in \mathbb{C}^{h \times w \times n}$ were the inputs in our training dataset.

C. Doppler velocity estimation with deep learning

We explored the use of convolutional neural networks, arguing that the spatial nature of convolutions could compensate for the loss of temporal information and produce better estimates. We focused on two architectures: U-Net and ConvNeXt.

1) *U-Net*: The U-Net architecture, introduced in 2015 for medical image segmentation, comprises an encoder-decoder structure with skip connections [21]. Given that our input S was complex-valued and the output D was real-valued, we examined two architectures: one entirely real-valued and another that was complex-valued in all layers except the last one.

- *Real U-Net*. The input of the network was in the form $S \in \mathbb{R}^{h \times w \times 2n}$, where the real and imaginary parts of the I/Q signals were concatenated to obtain real representations. We began with the standard U-Net architecture as described in [21] and made two modifications. Firstly, to ensure a lightweight network, we retained two dimension reduction steps, thereby reducing the network to only 6 layers. Secondly, the kernels of the two convolutions at each U-Net stage were of sizes 5×5 and 3×3 . The number of feature maps was (32, 64, 128) for the encoder part and symmetrically for the decoder part. With this configuration, the network had 1.5 million parameters, and the lowest dimension achieved was 45×10 .
- *Complex U-Net*. The input of the network was in the form $S \in \mathbb{C}^{h \times w \times n}$. The network features mirrored those of the Real U-Net, with the following distinctions. We employed complex-valued convolutions using the PyTorch framework. In particular, the network weights were complex-valued in all layers except the last one, where the real and imaginary parts of the feature map were concatenated before applying the real-valued weights. For the complex-valued activation function, we used the Complex ReLU, as advised by Trabelsi *et al.* [22]. Finally, we used the 2-D complex batch normalization available in [23], implemented following the recommendations of [22] in order to obtain equal variance in both the real and imaginary parts. The network contained the same number of parameters as the Real U-Net (1.5 M), but the convolutional operations were four times more computationally intensive.

2) *ConvNeXt*: We then considered ConvNeXt [24], a recent convolutional neural network architecture. Since 2020, hierarchical vision Transformer architectures have demonstrated superior performance over convolutional models in various computer vision tasks. However, Transformers models suffer from being data-intensive. As a solution, ConvNeXt was

introduced in 2022. It is a convolutional-based model that integrates strategies from Transformer models, resulting in an architecture that outperforms state-of-the-art Transformers while remaining data efficient. One of its key features is the separate handling of spatial and channel information, akin to the self-attention mechanism in Transformers. This property is achieved through the use of depthwise convolutions. We adapted the ConvNeXt architecture to our problem by utilizing a U-Net architecture with ConvNeXt as its backbone (*i.e.*, ConvNeXt layers served as feature extractors in the encoder part). A schematic representation of the considered architecture is depicted in Figure 2. With this configuration, the network comprised 4.7 million parameters, and the lowest dimension attained was 11×5 .

D. Data augmentation

In computer vision, a common technique for data augmentation involves applying simple transformations to available images to increase the size and variability of the training set, resulting in a more robust model. However, for ultrasound data, the options for realistic data augmentation are limited due to the specific geometry of the acquisition process. Additionally, some augmentation techniques require interpolation, which may significantly alter the signal information. Therefore, we only used vertical flipping and a custom zooming procedure.

1) *Zoomed samples*: Zooming was achieved by simulating color Doppler with $1.5w$ focused firings and then beamforming on a finer grid of $1.5h$ points. The resulting simulated samples were included in the dataset with a probability of 0.2 and randomly cropped to a $h \times w$ shape.

2) *Aliased samples*: Aliasing patterns appear in Doppler velocity maps when the maximum speed exceeds the Nyquist velocity. In such cases, phase unwrapping methods can be applied to the Doppler velocity maps as a post-processing step, as described in the introduction. In contrast to these post-processing approaches, we aimed to make the Doppler estimation inherently robust to aliasing. This was achieved by providing the deep learning models with new training pairs (S, D) , where S represents the I/Q aliased signal and D denotes the corresponding alias-free Doppler ground truth. The aliased samples were obtained during simulations by randomly reducing the original PRF values of each sequence by a factor in the $[0.4, 0.6]$ range. Each simulated aliased sample was considered as an *augmented* sample and added to the dataset with a probability of 0.5.

III. EXPERIMENTS

A. Data acquisition

1) *In silico experiment*: Our simulations, described in Section II-B, aimed to neglect the influence of clutter through the exclusive modeling of tissue-free scenarios. As input, we used patient color Doppler echocardiographic data acquired in a previous study [25]. Examinations adhered standard echocardiographic procedures, resulting in a diverse patient population encompassing individuals with and without cardiac disease. Due to randomization and anonymization, demographic and clinical details were unknown during the

TABLE I

PARAMETERS OF THE SIMULATED ULTRASOUND ACQUISITION

Parameter	Value
simulated probe	Verasonics P4-2v
central frequency (f_c)	2.7×10^6 Hz
pitch	300×10^{-6} m
#elements	64
bandwidth	74%
#cycles B-mode	1
#cycles Doppler	6
PRF	patient-based
sector width	patient-based
depth	patient-based
sampling frequency	$4f_c$
acquisition type	focused waves
#firings B-mode	120
#firings Doppler	40
packet size (n)	8

present work. The sequences, captured from the apical three-chamber view, incorporated both B-mode and Doppler velocity information. We obtained cardiac sequences of 37 patients, each containing at least one complete cardiac cycle (total: 2,576 samples). From these patient echocardiographic data, we simulated 2,576 training pairs. The probe and acquisition parameters are summarized in Table I. During the simulations, we leveraged patient-specific parameters as pulse repetition frequency (PRF), sector angle, and depth. The simulated RF signals were demodulated into I/Q signals and beamformed into a 180×40 grid using delay-and-sum. Each training pair (S, D) consisted of I/Q signals $S \in \mathbb{C}^{180 \times 40 \times n}$ and Doppler velocity references $D \in \mathbb{R}^{180 \times 40}$. Figure 3A shows a simulated sample.

Data augmentation, as described in Section II-D.1, was applied. Samples with reduced packet size were created by selecting two consecutive frames in the slow-time. The augmented dataset of color Doppler simulations contained $\sim 4,500$ pairs and is referred to as Alias-free-Set from here on.

Only 24 of the original samples contained aliased regions, accounting for less than 0.001% of the total dataset pixels. As a consequence, we used the procedure described in Section II-D.2. The subset of aliased samples is referred to as Aliased-Set in the sequel. An example of a sample simulated with both its original and decreased PRF is depicted in Figure 3. The Alias-free-Set dataset augmented with the Aliased-Set contained $\sim 5,300$ pairs, henceforth referred to as Full-Set.

2) *In vitro experiment*: We designed a rotating disk experiment to measure Doppler velocities from a color Doppler acquisition. We constructed a 6 cm diameter agar disk and connected it to a rotating motor. A Verasonics P4-2v cardiac probe was positioned at 8 cm from the center of the disk. To minimize clutter and artifacts, we submerged the entire setup in a water tank. Figure 4A provides a schematic representation of the experiment.

The acquisition parameters were set as described in Table I, with the sector width, depth and PRF configured to 50° , 12 cm and 6,000 Hz, respectively. The packet size was $n = 32$. We ran a total of 12 experiments, varying the maximum outer speed between 0.1 and 1.13 m/s, mirroring intracardiac blood

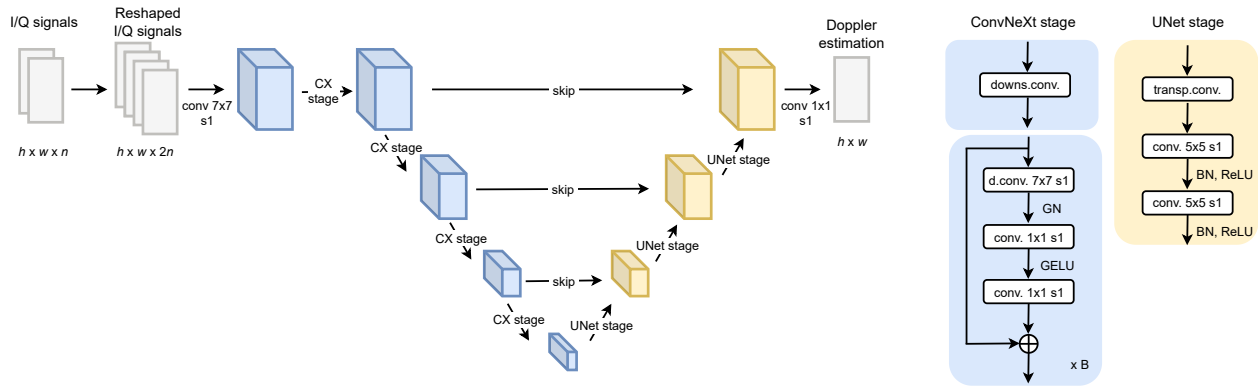


Fig. 2. Architecture of the ConvNeXt-based model. Initially, I/Q signals were reshaped by concatenating their real and imaginary parts. A first set of 7×7 convolutions was applied. Following this, four ConvNeXt stages were implemented, each containing (3, 3, 9, 3) blocks, with a down-sampling step in all stages except the first. The number of feature maps was (32, 64, 128, 256) for the encoder part and symmetrically for the decoder part. "d.conv.", "trans.conv." and "downs.conv." stand for depthwise, transposed and pooling convolution, respectively. "s" indicates the convolution stride.

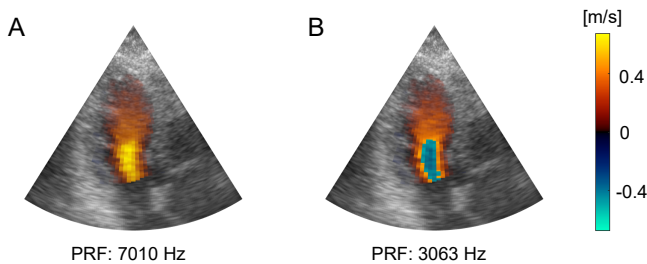


Fig. 3. Example of a simulated echocardiography acquisition with its original (A) and decreased (B) PRF. The Doppler velocities were estimated using an autocorrelator in both cases.

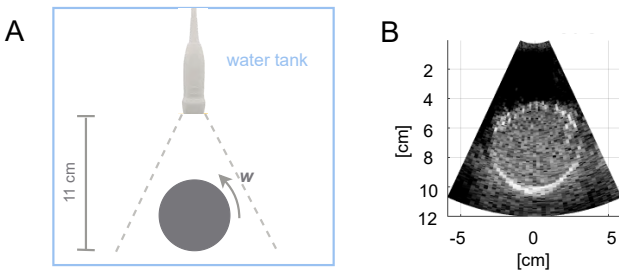


Fig. 4. (A) Schematic representation of the *in vitro* experiment using a rotating disk phantom. (B) B-mode obtained during one of the experiments.

velocities. The Nyquist velocity was $v_N = 0.85$ m/s for all experiments, resulting in four experiments displaying aliased regions. An example of an acquisition result is depicted in Figure 4B.

Samples with reduced packet size were created by selecting two consecutive frames in the slow-time. The ground truth Doppler velocity maps used to compute the models' estimation errors during inference were obtained from the known rotation speed of the phantom.

3) *In vivo* experiment: Cardiac echo-Doppler focused acquisitions were conducted on two healthy volunteers. The study received approval from the ethics and research committee of Polytechnique Montréal (CER-2122-54-D). In accordance

with the simulation and phantom experiments, the P4-2v Verasonics probe was used with identical parameters as those used for the phantom. Data were acquired in the apical three-chamber view, including both B-mode and Doppler information. The packet size was set to $n = 16$. The ultrasound signals were post-processed and beamformed on a 180×40 grid. We obtained 16 color Doppler image samples from each volunteer.

To be consistent with the clutter-free setting of both *in silico* and *in vitro* experiments, we applied an SVD-based clutter filter. We discarded the first four largest singular values to isolate the blood I/Q signals. We segmented the left ventricular endocardium unsupervisedly using a deep learning-based method [16] to mask off I/Q signals not pertaining to intraventricular blood. Samples with low Doppler power (< -20 dB) in over 70% of the ventricular region were discarded, resulting in 20 selected samples. None of the selected samples contained aliased regions. To assess the method's robustness, we deliberately introduced aliasing by considering one out of three frames in the slow-time, thereby artificially decreasing the PRF by a factor of three. We applied this procedure to the 20 selected samples, providing seven new samples with several aliased regions.

Samples with reduced packet size were generated by selecting two consecutive frames in the slow-time. In the *in vivo* experiments, ground truth Doppler velocity maps were not available. Hence, Doppler velocity maps obtained with the autocorrelator using the maximum packet size ($n = 16$) were used as references.

B. Deep learning implementation details

1) Training settings: Because the ultrasound data were simulated with patient-specific PRFs, the Nyquist velocity varied from sample to sample. Therefore, we opted to use phase shifts as the ground truth instead of Doppler velocities to ensure standardized inputs across samples. Each sample was normalized with respect to its maximum complex modulus across the slow-time. For greater robustness, we trained the network using 9-fold cross-validation, where samples from a given patient were assigned to the same fold. Once the test

TABLE II

DOPPLER VELOCITY ESTIMATION RESULTS OF THE BASELINE METHOD AND THE THREE PROPOSED DEEP LEARNING METHODS, TRAINED AND EVALUATED ON ALIAS-FREE-SET. THE RESULTS CORRESPOND TO THE MEAN \pm STANDARD DEVIATION OF THE RMSE [M/S $\times 10^{-2}$].

Method	Alias-free-Set
Autocorrelator	3.7 ± 2.6
Real U-Net	1.8 ± 0.9
Complex U-Net	1.8 ± 0.8
ConvNeXt	1.7 ± 0.9

fold was chosen, the split of samples in training/validation was set to 9/1. The network weights were initialized with the Xavier distribution. The loss function was the masked mean squared error, where the mask corresponded to the left ventricular cavity. The optimizer was AdamW, and the batch size was set to 16. The initial value of the learning rate was set to 10^{-3} and decreased following a plateau scheduler with a patience of 10 epochs. Trainings were performed using the PyTorch library on an NVIDIA V100 GPU with 16 GB of memory.

2) *Inference*: Inference on the *in silico* dataset was executed using the 9-fold cross-validation procedure, thus obtaining a prediction for each sample of the dataset. Ensemble inference was carried out for both the *in vitro* and *in vivo* datasets, by calculating the median value predicted for each pixel by the nine models of the cross-validation.

C. Metrics

To gauge the accuracy of Doppler velocity estimation, we calculated the root mean squared error (RMSE) between the ground truth and predicted maps, and examined its mean and standard deviation. Note that for the *in vivo* experiments, we refer to this metric as the root mean squared difference (RMSD) as we only had access to an imperfect reference, not the ground truth.

IV. RESULTS

A. In silico results

The performance of our methods in Doppler velocity estimation, when trained on Alias-free-Set and using a packet size $n = 2$, is presented in Table II. For all methods, the metrics were computed exclusively within the left ventricle.

All deep learning approaches outperformed the autocorrelator. Real U-Net and Complex U-Net provided comparable results, with ConvNeXt slightly outperforming both. In addition, the variability of the three models was significantly lower than that of the autocorrelator.

We examined the samples with higher errors for both the autocorrelator and the deep learning models. For the former, large estimation errors mainly occurred in two situations: noisy signals and in the presence of aliasing. For the deep learning models, large errors were observed in two specific settings: in a few pixels near the valves and the presence of aliasing across a large region of the ventricle. To mitigate the effects of aliasing, we trained our models on the Full-Set dataset. The results of the estimation are presented in Table III.

TABLE III

DOPPLER VELOCITY ESTIMATION RESULTS OF THE BASELINE METHOD AND THE THREE PROPOSED DEEP LEARNING METHODS TRAINED ON FULL-SET AND EVALUATED ON ALIAS-FREE-SET AND ALIASED-SET. THE RESULTS REPRESENT THE MEAN \pm STANDARD DEVIATION OF THE RMSE [M/S $\times 10^{-2}$].

Method	Alias-free-Set	Aliased-Set
Autocorrelator	3.7 ± 2.6	16 ± 9.8
Real U-Net	1.9 ± 0.8	2.5 ± 1.4
Complex U-Net	1.8 ± 0.8	2.4 ± 1.6
ConvNeXt	1.7 ± 0.7	2.3 ± 1.5

We noted that when testing on Alias-free-Set, we maintained a consistent performance compared to the models trained on Alias-free-Set. Moreover, when testing on Aliased-Set, we observed a substantial reduction in error compared to the autocorrelator method, indicating that the models are capable to alleviate aliasing to some extent.

Illustrative examples are shown in Figure 5. We only report predictions obtained with ConvNeXt (the best-performing model) for the sake of brevity. As explained above, ConvNeXt trained in both ways achieved accurate estimates in the absence of aliasing (samples A and B). In such cases, we observed that the autocorrelator was more susceptible to noise. When encountering small aliased regions (sample C, left region), the model trained with Alias-free-Set effectively removed aliasing. However, when aliasing extended over a large region, it either smoothed the region, leading to velocity underestimation (sample C, right part, and sample D), or produced a pattern similar to that of the autocorrelator, albeit smoothed at the edges (sample E). In contrast, ConvNeXt trained on Full-Set yielded results closely aligned with the ground truth.

B. In vitro validation

Figure 6 shows the RMSE obtained with the baseline and deep learning-based methods for all 12 *in vitro* experiments, using the same packet size $n = 2$. The RMSE is depicted as a function of the phantom outermost speeds. The Nyquist velocity v_N was kept unchanged across all experiments. We observed that our methods gave state-of-the-art results for $v < v_N$ and showed improved performance when $v > v_N$. Real U-Net and Complex U-Net gave comparable results, while ConvNeXt slightly underperformed them in terms of aliasing reduction for high rotation speeds.

Estimation maps for two of the experiments are presented in Figure 7. In the first experiment, where the maximum velocity was below the Nyquist limit, we observed estimations closely aligned with the ground truth for all our networks. Moreover, our estimates were less noisy compared to those of the autocorrelator. In the second experiment, where the maximum velocity exceeded the Nyquist limit, the autocorrelator estimation clearly revealed the aliased regions. In contrast, our methods significantly reduced the aliased patterns. However, ConvNeXt displayed noisy artifacts in the aliased regions.

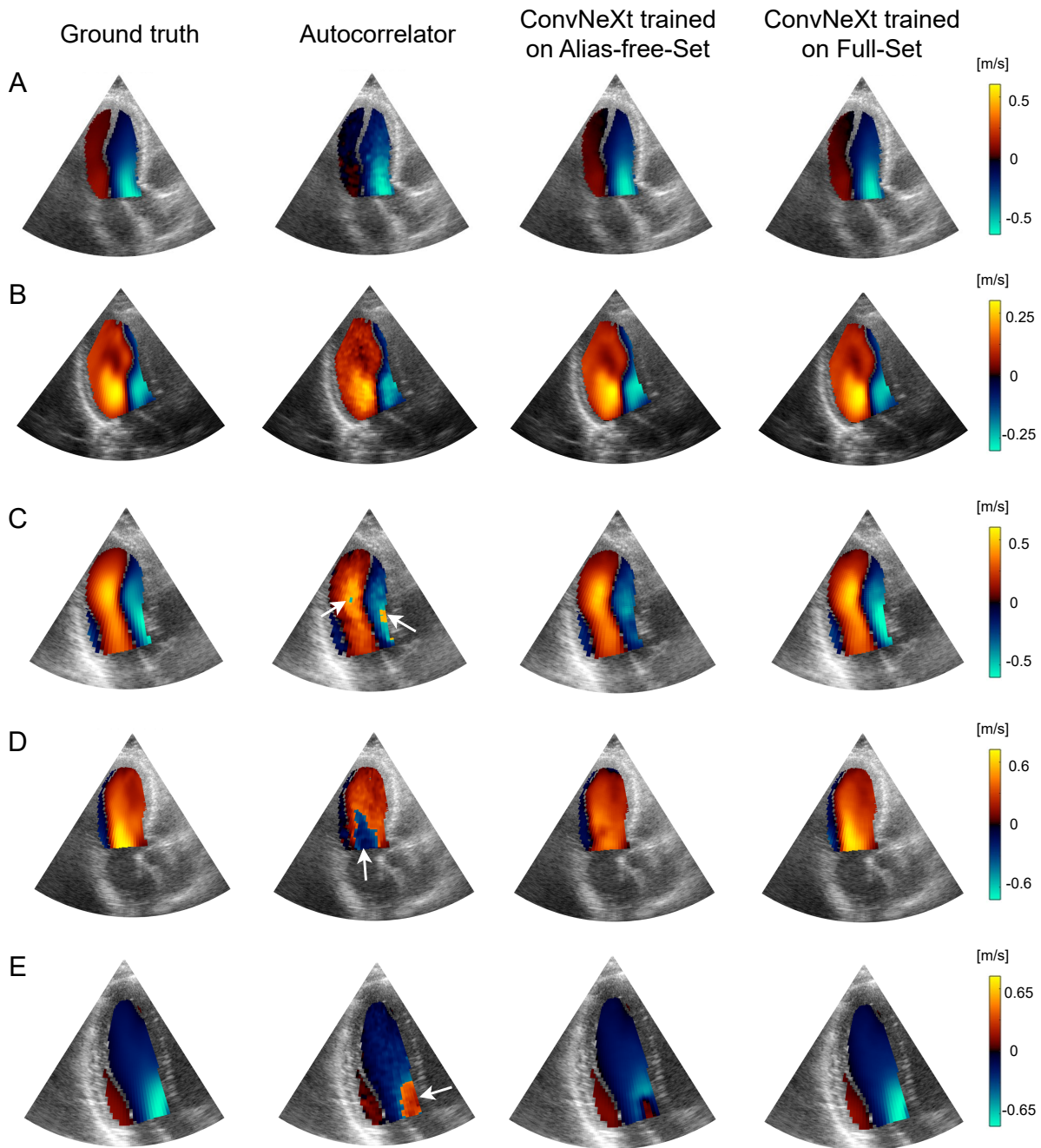


Fig. 5. Doppler velocity estimations for five *in silico* samples. B-mode and color Doppler images were simulated from the data of five different patients. Samples A and B had velocities below the Nyquist limit. Samples C, D and E had velocities above the Nyquist limit. Aliased regions are indicated with white arrows.

C. *In vivo* validation

Estimation results on the *in vivo* sequences are reported in Table IV. We summarized results on 20 *standard* volunteer samples, and on 7 aliased samples, and the reference used for all RMSD calculations was the autocorrelation result with $n = 16$.

For *standard* samples, the autocorrelator with $n = 2$ produced the results closest to the reference. All three deep learning methods led to comparable results, albeit slightly further away from the reference. Figure 8 shows illustrative

examples for Complex U-Net and ConvNeXt. In sample A, our methods estimated a Doppler velocity map comparable to the autocorrelator estimation with $n = 2$. In this case, all three estimates computed with $n = 2$ produced smaller velocity estimates than those provided by the autocorrelator with $n = 16$. In sample B, the three methods also yielded lower velocity estimates, but the two deep learning methods behaved as denoisers, unlike the autocorrelator with $n = 2$. The denoising capability of our methods is evident in sample C.

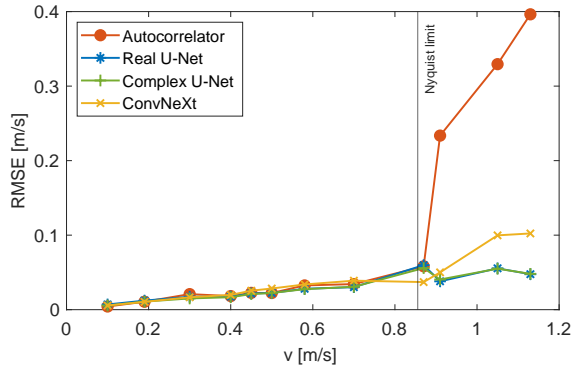


Fig. 6. Evolution of the RMSE with respect to the outermost speed of the phantom, obtained by the autocorrelator and deep learning methods. The Nyquist limit is indicated with a gray line.

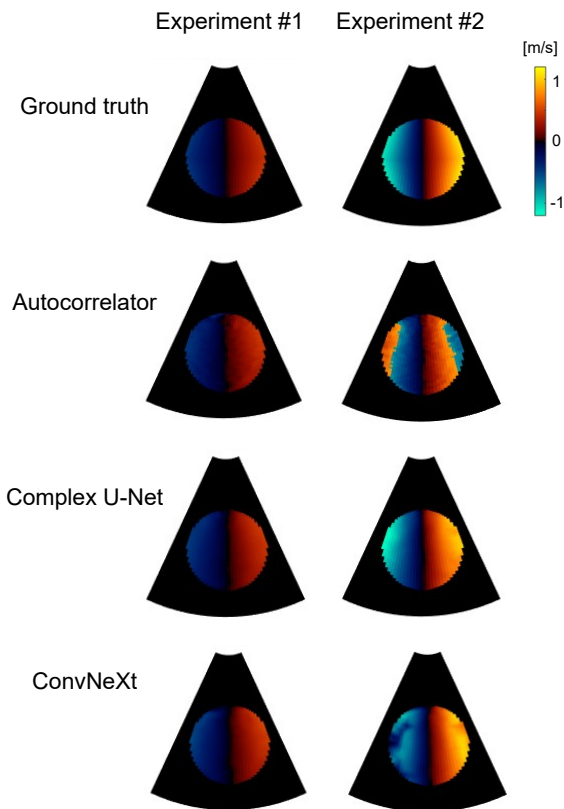


Fig. 7. Doppler velocity estimation results for two *in vitro* experiments. Results are masked to show the phantom region only. Experiment #1 had $v_{\max} = 0.45$ m/s. Experiment #2 had $v_{\max} = 1.15$ m/s.

For aliased samples, as anticipated, the autocorrelator exhibited a substantial deviation from the reference due to aliasing, as reported in Table IV. Among the deep learning methods, ConvNeXt gave results significantly closer to the reference compared to Real U-Net and Complex U-Net. Sample D exemplifies a scenario with induced aliasing, where both Complex U-Net and ConvNeXt successfully reduced aliasing. However, Complex U-Net yielded an incorrect prediction in the lower right corner. Similarly, sample E illustrates another example of induced aliasing where ConvNeXt reduced aliasing, while Complex U-Net yielded relatively large errors in the apical

TABLE IV

DOPPLER VELOCITY ESTIMATION RESULTS OF THE BASELINE METHOD AND THE THREE PROPOSED DEEP LEARNING METHODS, EVALUATED ON STANDARD VOLUNTEER AND ALIASED VOLUNTEER SAMPLES. THE RESULTS REPRESENT THE MEAN \pm STANDARD DEVIATION OF THE RMSD [$\text{m/s} \times 10^{-2}$].

Method	Standard samples	Aliased samples
Autocorrelator	5.8 ± 6.0	11 ± 2.1
Real U-Net	6.4 ± 6.3	8.1 ± 2.0
Complex U-Net	6.6 ± 6.3	8.1 ± 1.3
ConvNeXt	6.5 ± 5.9	4.0 ± 2.3

TABLE V

COMPARISON OF THE COMPLEXITY OF THE AUTOCORRELATOR AND THE THREE DEEP LEARNING MODELS. THE TRAINING TIMES ARE REPORTED FOR THE TRAINING OF ONE FOLD, ASSUMING THAT ALL FOLDS CAN BE TRAINED IN PARALLEL. THE INFERENCE TIMES ARE REPORTED FOR THE INFERENCE OF IN VIVO FRAMES. TRAINING AND INFERENCE WERE BOTH CONDUCTED ON GPU FOR DEEP LEARNING MODELS AND ON CPU FOR THE AUTOCORRELATOR.

Model	# Parameters	Training time	Inference time
Autocorrelator	x	x	~ 1.6 ms
Real U-Net	1.5 M	~ 0.5 h	~ 2.5 ms
Complex U-Net	1.5 M	~ 5 h	~ 18 ms
ConvNeXt	4.7 M	~ 1 h	~ 5.3 ms

region.

D. Computational time

Table V reports the sizes of the models and their training and inference times. With the same parameter configuration, Complex U-Net required significantly more training and inference time than Real U-Net. In contrast, ConvNeXt offered the most advantageous trade-off between the number of parameters and the training and inference times. The inference times of the three models were sufficiently low to support real-time imaging.

V. DISCUSSION

A. Doppler velocity estimation with deep learning

We achieved reliable Doppler velocity estimates from a reduced packet size $n = 2$ with all proposed deep learning models. This held true for alias-free samples across *in silico*, *in vitro* and *in vivo* data (Table II, Figure 6 and Table IV). When working with aliased sequences, we noticed that the models trained on alias-free data (Alias-free-Set) tended to smooth the aliased regions (sample E of Figure 5), rendering post-processing dealiasing methods less suitable. Augmenting the training dataset with aliased simulations notably improved the model's ability to mitigate aliasing in both *in silico* and *in vitro* data (sample E of Figure 5 and experiment 2 of Figure 8). For *in vivo* data, only ConvNeXt reliably reduced aliasing in all samples (Table 6 and sample E of Figure 8).

Regarding the autocorrelation method, our analysis revealed that our deep learning models yielded comparable quantitative results on *in vitro* and *in vivo* validations (Figure 6 and Table IV). We also observed that this method remained stable

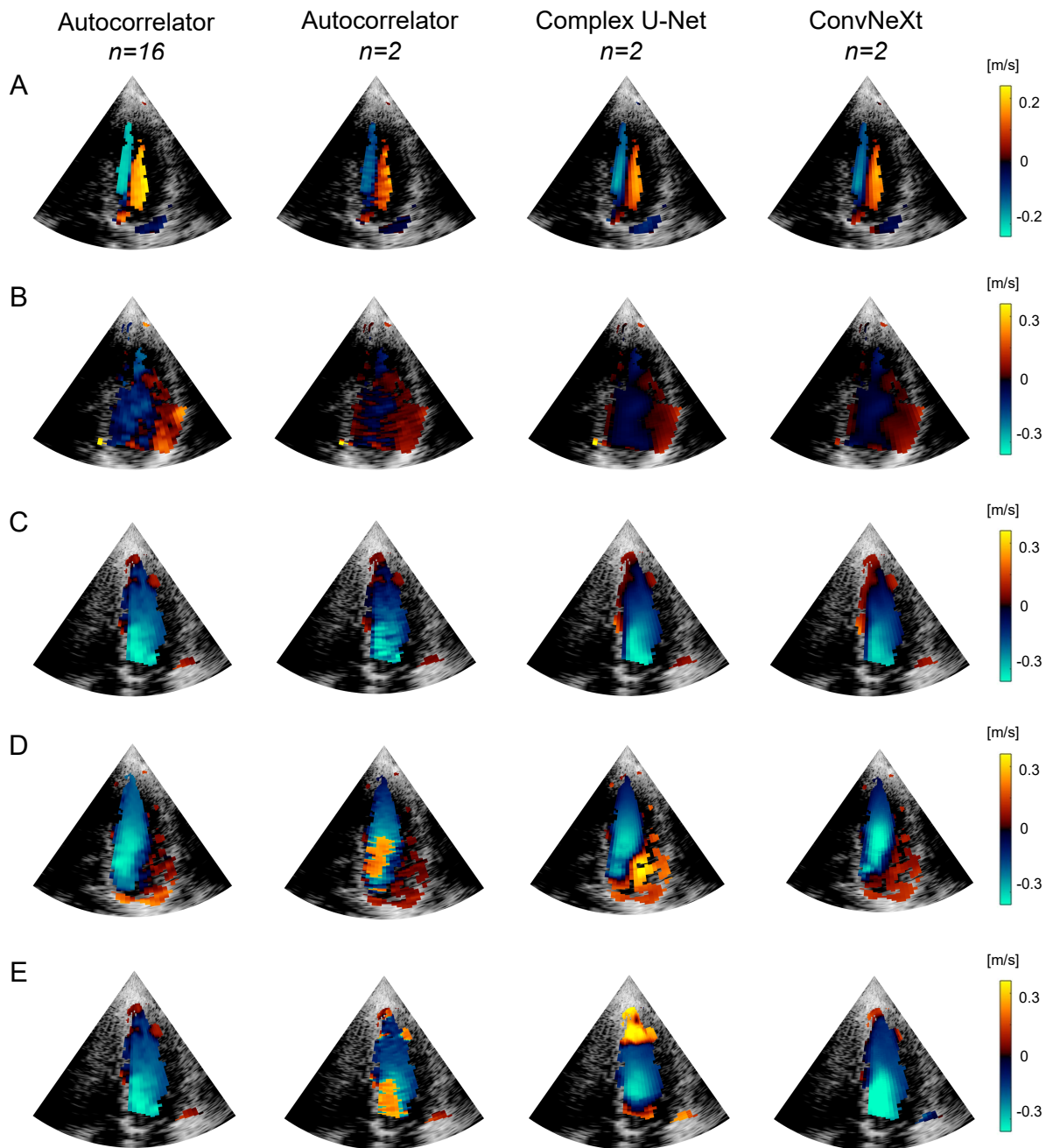


Fig. 8. Doppler velocity estimation results on five *in vivo* samples using the autocorrelator with packet sizes 16 and 2, the Complex U-Net and the ConvNeXt. Samples A, B and D are from volunteer #1. Samples C and E are from volunteer #2. Samples D and E induced aliasing.

when the packet size was reduced to $n = 2$, albeit producing some level of noise on the velocity maps (samples A and C in Figure 8). However, in some cases, the estimation displayed a noticeable level of noise (sample B in Figure 8). In contrast, our method appeared to be more robust to noise (as evident in sample B in Figure 8).

A critical concern with deep learning methods is their capacity to generalize. Models frequently fail to extrapolate when the distribution shift between training and validation datasets is too large. This issue is particularly pronounced in medical imaging, where distribution shifts can arise from

various factors, such as differences in data acquisition or preprocessing, and training with unrealistic simulations. In our study, we obtained encouraging results regarding the generalization capability of our models, despite being trained exclusively on *in silico* samples. Notably, our models demonstrated high-quality results on unseen *in vitro* and *in vivo* data. We only observed large errors in a limited number of cases (as exemplified by sample E in Figure 8). This highlights the effectiveness of our simulation pipeline in generating color Doppler samples to build supervised training datasets. Our pipeline enables users to expand the range of clinical

scenarios, including different echocardiographic views, transmission types (e.g. focused, diverging), and transducer types. It uses patient-derived Doppler echocardiographic data to model cardiac tissue and blood backscatterers. Although our focus was on 3-chamber views, other views could thus be explored. Additionally, the SIMUS simulator provides the flexibility to specify various probe and transmission parameters, allowing for the simulation of ultrasound signals with non-focused waves. This capability facilitates the generation of a vast pool of training data, enabling the creation of more generalized models. It is also worth noting that the favorable results obtained in non-flow data during the *in vitro* disk experiments suggests the potential for directly applying our model to other modalities involving Doppler velocity estimation, such as tissue Doppler.

Finally, this work could offer practical advantages for focused color Doppler echocardiography. For instance, assuming a typical packet size of around $n = 8$, reducing the packet size to $n = 2$ could result in a fourfold increase in frame rate, yielding approximately 60 frames/cardiac cycle. Such an enhancement could ease quantitative assessment of blood velocity throughout the cardiac cycle, potentially simplifying the application of techniques like intracardiac Vector Flow Mapping.

B. Model comparison

Two data representation methods were evaluated: the real-valued and complex-valued. Additionally, two types of architectures were benchmarked: U-Net and ConvNeXt.

1) *Real-valued vs. complex-valued representations*: We anticipated that Complex U-Net would outperform Real U-Net, in line with findings from our previous studies [26]. However, Real U-Net and Complex U-Net yielded comparable results. One hypothesis is that while the Doppler velocity estimation problem involves complex-valued inputs, the output data are real-valued. Therefore, a transition from complex-valued to real-valued representations may be necessary, and a complex-valued architecture may not fully leverage its capabilities in this specific scenario. We also noted that the use of complex-valued representations significantly increased computational time during both training and inference due to the higher number of operations associated with complex-valued convolutions.

2) *U-Net vs ConvNeXt*: In our study, we analyzed ConvNeXt, a state-of-the-art deep learning architecture built on enhanced convolutional layers mirroring Transformer blocks for improved performance. We found similar performance from U-Net and ConvNeXt on *in silico* and *in vitro* data. Although ConvNeXt exhibited imperfect results on highly aliased *in vitro* samples, it demonstrated greater robustness when applied to *in vivo* aliased samples. Moreover, despite including more parameters, ConvNeXt provided fast training and inference times.

C. Limitations and perspectives

Despite the positive findings presented in this paper, we acknowledge two main limitations of this work. First, the *in*

vivo results must be regarded as preliminary as they were based on only two subjects, and the absence of ground truth made it difficult to draw firm conclusions. Second, we did not account for clutter influence and assumed that clutter filtering could be performed effectively on a packet size of two. A challenging perspective is to develop a clutter filtering method specifically designed for reduced packet sizes. We believe that exploring deep learning techniques could be a promising avenue for addressing this task.

Another perspective is to incorporate temporal information [16]. At this stage, our simulations were based on patient data from an echocardiographic laboratory, and thus lacked sufficient temporal resolution to assume continuity between consecutive frames. Realistic color flow imaging simulations based on a numerical heart, such as those employed previously in our research [15], [27], would provide the necessary frame rate to leverage temporal information effectively and improve the robustness of our models.

VI. CONCLUSION

We explored deep learning models for Doppler velocity estimation from clutter-filtered IQ signals with a reduced packet size of $n = 2$. Using a color Doppler simulation pipeline, we built a supervised training dataset and applied custom data augmentation, including artificially adding aliased samples. Our method exhibited superior performance on *in silico* data compared to the autocorrelation method, and achieved proficient results on *in vitro* and *in vivo* data. We demonstrated the effectiveness of our approach in reducing aliasing and handling noise. Three deep learning strategies were evaluated, with ConvNeXt showing the best generalization performance on *in vivo* samples. Overall, our study highlights the potential of deep learning for quantitative color Doppler analysis with minimal data.

ACKNOWLEDGMENTS

This work has benefited from access to IDRIS computing resources through the allocation of resources 2022-AD011013679 granted by GENCI. The *in vitro* material is based upon work done on the PILoT facility (PILoT, INSA-Lyon). The RF Verasonics generator was cofounded by the FEDER program, Saint-Etienne Metropole (SME), and Conseil General de la Loire (CG42) within the framework of the SonoCardio-Protection Project led by Pr Pierre Croisille.

REFERENCES

- [1] J. Faurie, M. Baudet, J. Poree, G. Cloutier, F. Tournoux, and D. Garcia, "Coupling myocardium and vortex dynamics in diverging-wave echocardiography," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 66, no. 3, pp. 425–432, 2019.
- [2] A. C. H. Yu, K. W. Johnston, and R. S. C. Cobbold, "Frequency-based signal processing for ultrasound color flow imaging," *Canadian Acoustics*, vol. 35, no. 2, pp. 11–23, 2007.
- [3] T. Loupas, J. Powers, and R. Gill, "An axial velocity estimator for ultrasound blood flow imaging, based on a full evaluation of the doppler equation by means of a two-dimensional autocorrelation approach," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 42, no. 4, pp. 672–688, 1995.
- [4] D. Posada, J. Poree, A. Pellissier, B. Chayer, F. Tournoux, G. Cloutier, and D. Garcia, "Staggered multiple-prf ultrafast color doppler," *IEEE Trans Med Imaging*, vol. 35, no. 6, pp. 1510–21, 2016.

- [5] S. Muth, S. Dort, I. Sebag, M. Blais, and D. Garcia, "Unsupervised dealiasing and denoising of color-doppler data," *Med Image Anal*, vol. 15, no. 4, pp. 577–88, 2011.
- [6] B. Luijten, N. Chennakeshava, Y. C. Eldar, M. Mischi, and R. J. G. van Sloun, "Ultrasound Signal Processing: From Models to Deep Learning," *Ultrasound in Medicine & Biology*, vol. 49, no. 3, pp. 677–698, 2023.
- [7] R. J. Van Sloun, H. Belt, K. Janse, and M. Mischi, "Learning doppler with deep neural networks and its application to intra-cardiac echography," *IEEE International Ultrasonics Symposium (IUS)*, pp. 1–4, 2018.
- [8] I. Z. Apostolakis, F. C. Meral, J. S. Shin, F. G. G. M. Vignon, S. Wang, and J. F. Robert, "Systems and methods for generating color doppler images from short and undersampled ensembles," *U.S. Patent WO20228873A1*, Jul. 20, 2023.
- [9] B. He, J. Lei, X. Lang, Z. Li, W. Cui, and Y. Zhang, "Ultra-fast ultrasound blood flow velocimetry for carotid artery with deep learning," *Artificial Intelligence in Medicine*, vol. 144, 2023.
- [10] O. Solomon, R. Cohen, Y. Zhang, Y. Yang, Q. He, J. Luo, R. J. van Sloun, and Y. C. Eldar, "Deep unfolded robust PCA with application to clutter suppression in ultrasound," *IEEE Transactions on Medical Imaging*, vol. 39, no. 4, pp. 1051–1063, 2020.
- [11] K. G. Brown, D. Ghosh, and K. Hoyt, "Deep Learning of Spatiotemporal Filtering for Fast Super-Resolution Ultrasound Imaging," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 67, no. 9, pp. 1820–1829, 2020.
- [12] H. Nahas, J. S. Au, T. Ishii, B. Y. S. Yiu, A. J. Y. Chee, and A. C. H. Yu, "A deep learning approach to resolve aliasing artifacts in ultrasound color flow imaging," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 67, no. 12, pp. 2615–2628, 2020.
- [13] H. Nahas, B. Y. Yiu, A. J. Chee, J. S. Au, and A. C. Yu, "Deep-learning-assisted and gpu-accelerated vector doppler imaging with aliasing-resistant velocity estimation," *Ultrasonics*, vol. 134, p. 107050, 2023.
- [14] H. J. Ling, O. Bernard, N. Ducros, and D. Garcia, "Phase unwrapping of color doppler echocardiography using deep learning," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 70, no. 8, pp. 810–820, 2023.
- [15] Y. Sun, F. Vixège, K. Faraz, S. Mendez, F. Nicoud, D. Garcia, and O. Bernard, "A Pipeline for the Generation of Synthetic Cardiac Color Doppler," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 69, no. 3, pp. 932–941, 2022.
- [16] H. J. Ling, N. Painchaud, P.-Y. Courand, P.-M. Jodoin, D. Garcia, and O. Bernard, "Extraction of volumetric indices from echocardiography: Which deep learning solution for clinical use?" *Proc. FIMH*, pp. 245–54, 2023.
- [17] F. Vixège, A. Berod, Y. Sun, S. Mendez, O. Bernard, N. Ducros, P.-Y. Courand, F. Nicoud, and D. Garcia, "Physics-constrained intraventricular vector flow mapping by color doppler," *Physics in Medicine & Biology*, vol. 66, no. 24, p. 245019, Dec. 2021.
- [18] D. Garcia, "SIMUS: An open-source simulator for medical ultrasound imaging. Part I: Theory & examples," *Computer Methods and Programs in Biomedicine*, vol. 218, p. 106726, 2022.
- [19] A. Cigier, F. Varray, and D. Garcia, "SIMUS: An open-source simulator for medical ultrasound imaging. Part II: Comparison with four simulators," *Computer Methods and Programs in Biomedicine*, vol. 220, p. 106774, 2022.
- [20] D. Garcia, "Make the most of MUST, an open-source Matlab UltraSound Toolbox," in *IEEE International Ultrasonics Symposium (IUS)*. Xi'an, China: IEEE, 2021, pp. 1–4.
- [21] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pp. 234–241, 2015.
- [22] C. Trabelsi, O. Bilaniuk, Y. Zhang, D. Serdyuk, S. Subramanian, J. F. Santos, S. Mehri, N. Rostamzadeh, Y. Bengio, and C. J. Pal, "Deep Complex Networks," *ICLR*, 2018.
- [23] M. W. Matthès, Y. Bromberg, J. de Rosny, and S. M. Popoff, "Learning and avoiding disorder in multimode fibers," *Phys. Rev. X*, vol. 11, 2021.
- [24] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A convnet for the 2020s," *CVPR*, 2022.
- [25] F. Mehregan, F. Tournoux, S. Muth, P. Pibarot, R. Rieu, G. Cloutier, and D. Garcia, "Doppler vortography: a color Doppler approach for quantification of the intraventricular blood flow vortices," *Ultrasound in medicine & biology*, vol. 40, no. 1, pp. 210–221, 2014.
- [26] J. Lu, F. Millioz, D. Garcia, S. Salles, D. Ye, and D. Friboulet, "Complex convolutional neural networks for ultrafast ultrasound imaging reconstruction from in-phase/quadrature signal," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 69, no. 2, pp. 592–603, 2022.
- [27] C. Chnafa, S. Mendez, R. Moreno, and F. Nicoud, "Using Image-based CFD to Investigate the Intracardiac Turbulence," in *Modeling the Heart and the Circulatory System*, A. Quarteroni, Ed. Cham: Springer International Publishing, 2015, vol. 14, pp. 97–117, series Title: MS&A.