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

Deep Learning Based Ultrasound Image Reconstruction Method: A Time Coherence Study

Dimitris Perdios*, Manuel Vonlanthen*, Florian Martinez*, Marcel Arditi*, and Jean-Philippe Thiran*†
*Signal Processing Laboratory 5 (LTS5), École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
†Department of Radiology, University Hospital Center (CHUV) and University of Lausanne (UNIL), Lausanne, Switzerland

***Abstract*—Recently, deep learning entered the ultrasound (US) image reconstruction community, demonstrating unprecedented performances on image reconstruction tasks. The use of deep neural networks to reconstruct, restore or enhance US images has been challenged on its capability to preserve time-coherence, mostly because of their inherent non-linear properties. Most novel image reconstruction methods are typically only evaluated on static images, lacking any demonstration of their potential applicability to other imaging modes, such as vector flow imaging (VFI) and shear-wave elastography, which heavily rely on the time-coherence of consecutive reconstructed images. In this work, we demonstrate that our previously proposed convolutional neural network (CNN)-based image restoration approach, trained exclusively to improve the quality of static images, does not harm the time-coherence of consecutive frames in the context of VFI. Using dynamical numerical phantoms inspired by the synthetic aperture vector flow imaging (SA-VFI) challenge, we quantitatively show that the use of such an image restoration technique does not damage vector flow estimations, computed with a state-of-the-art speckle tracking algorithm, on simple phantoms with constant echogenicity, and that it even has the potential to improve such estimations in more complex scenarios.**

***Index Terms*—Deep learning, speckle tracking, ultrasound imaging, vector flow imaging.**

I. Introduction

Image reconstruction has always been an active field of research in the ultrasound (US) community, as it is an essential part of any US scanner for providing a well-interpretable image to an operator, given raw radio frequency (RF) sensor data. Even though US is a dynamic imaging modality in essence, thanks to its real-time capabilities, the image quality of newly proposed image reconstruction methods is traditionally evaluated on static images, typically using meaningful quantitative metrics such as the ones proposed by the plane-wave imaging challenge in medical ultrasound (PICMUS) [1].

The recent emergence of deep learning techniques in the US image reconstruction community has raised concerns concerning the robustness of such methods to preserve the time-coherence of consecutive images, mostly because of their inherent non-linear properties. Such a concern is of course valid, especially in the ultrafast era that has led to breakthrough new imaging modes, such as shear-wave elastography and vector flow imaging (VFI), which heavily rely on the time-coherence of consecutive frames [2]. Indeed, most of these imaging modes rely on displacement estimations which are only achievable when the time-coherence of the underlying physical phenomenon is preserved in the reconstructed images.

In this work, we propose to assess the time-coherence robustness of a convolutional neural network (CNN)-based image restoration method in the context of VFI, which has recently been the subject of the synthetic aperture vector flow imaging (SA-VFI) challenge [3]. For the image restoration method, we have used the same strategy and network architecture as presented in our previous work [4], which has demonstrated the ability to reconstruct high-quality images from a single plane wave (PW) insonification. For the velocity estimation, we have implemented a method greatly inspired by the winning-team of the SA-VFI challenge [5], i.e. a speckle tracking technique based on normalized cross-correlations of consecutive frames with a coarse-to-fine multi-pass scheme. The resulting vector flow estimations is evaluated both locally and globally using quantitative statistical metrics on numerical dynamic phantoms inspired by the SA-VFI challenge and adapted to the PICMUS imaging configuration. These evaluations are performed both on image sequences reconstructed using conventional delay-and-sum (DAS) beamforming and using the proposed CNN-based approach, allowing us to assess how the latter affects vector flow estimations which are directly related to the time-coherence of consecutive frames.

The remainder of this paper is organized as follows. Section II summarizes the proposed CNN-based image enhancement method proposed in [4] and presents the deployed speckle tracking method. Section III describes the experiments and metrics used to asses the resulting VFI quality, and Section IV discusses the obtained results. Concluding remarks are given in Section V.

II. Proposed Approach

In order to assess the effect of a CNN-based image reconstruction approach on the time-coherence of consecutive frames, we have considered speckle tracking, a VFI method that relies strongly on proper time-coherence between frames to provide high-quality velocity estimations. In this section, we first summarize the CNN-based approach proposed in our

This work was supported in part by the Swiss National Science Foundation under Grant 205320_175974 and Grant 206021_170758. *(Corresponding author: Dimitris Perdios.)*

D. Perdios, M. Vonlanthen, F. Martinez, M. Arditi, and J.-Ph. Thiran are with the Signal Processing Laboratory 5 (LTS5), École Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland (email: dimitris.perdios@epfl.ch).

J.-Ph. Thiran is also with the Department of Radiology, University Hospital Center (CHUV) and University of Lausanne (UNIL), 1011 Lausanne, Switzerland.

previous work [4], followed by a detailed description of the deployed speckle tracking algorithm.

A. Ultrasound Image Enhancement Using Deep Learning

As proposed in [4], we apply a carefully trained CNN to low-quality US images, reconstructed using conventional delay-and-sum (DAS) beamforming from the measurements of a single PW acquisition with normal incidence, to significantly reduce the inherent image artifacts such as side and grating lobes. We use a neural network architecture adapted from the popular U-Net. It is a residual CNN and thus performs a mapping $\boldsymbol{F}_{\boldsymbol{\theta}}(\boldsymbol{X}) = \boldsymbol{X} + \boldsymbol{R}_{\boldsymbol{\theta}}(\boldsymbol{X})$, where $\boldsymbol{X} \in \mathbb{R}^{N_x \times N_z}$ and $\boldsymbol{R}_{\boldsymbol{\theta}}$ is a mapping function, with trainable parameters $\boldsymbol{\theta}$, capable of predicting the negative noise to be applied to a low-quality input image to recover a high-quality image.

The training dataset is composed of images reconstructed with a conventional DAS beamformer from simulated data of randomly generated tissue-mimicking phantoms. It contains both low-quality images described above and high-quality images reconstructed from full synthetic aperture measurements. The low-quality and high-quality images serve as input and reference (i.e. labels) during training phase.

B. Speckle Tracking Method

Speckle tracking, more generally called particle image velocimetry (PIV), is a method extensively used in optical flow measurements. It has also been shown to be very successful in the domain of US VFI, as it was the method implemented by the winning-team [5] of the SA-VFI challenge.

At its core, PIV consists of pattern-matching between regions of two consecutive images, $\boldsymbol{X}_0$ and $\boldsymbol{X}_1$, using 2D cross-correlation. A displacement vector field $\hat{\boldsymbol{d}}(x,z)$ can be estimated from the maxima of the obtained correlation coefficients. Given the knowledge of the elapsed time Δt between consecutive frames, the corresponding velocity estimate $\hat{\boldsymbol{v}}(x,z) = \hat{\boldsymbol{d}}(x,z)/\Delta t$ can be determined. We have deployed a state-of-the-art instance of PIV, inspired by the popular PIVlab toolbox [6] and [5]. The method consists of performing multiple passes of displacement estimations using FFT-based correlation. Between passes, the original image $\boldsymbol{X}_1$ is deformed to more closely resemble image $\boldsymbol{X}_0$, using the current displacement estimate. This method allows a progressively higher precision and displacement-range-coverage with each additional pass.

The input to each single pass are two frames, $\boldsymbol{X}_0, \boldsymbol{X}_1 \in \mathbb{R}^{N_x \times N_z}$, as well as both the interrogation window size $N_i \times N_j$ and the window overlap $o \in [0,1]$. First, the two frames are divided into windows according to the chosen interrogation window size and overlap. This way, we obtain interrogation window stacks $\boldsymbol{W}_0, \boldsymbol{W}_1 \in \mathbb{R}^{K \times N_i \times N_j}$, where K is the resulting number of windows. Collectively, the window centers (x_k, z_k), where $k \in \{0,1,\dots,K-1\}$, represent a uniform sampling grid of the physical image region. Each window pair $(\boldsymbol{W}_{0,k}, \boldsymbol{W}_{1,k})$ is used to estimate the displacement at the corresponding window center, by computing the coefficients $\boldsymbol{C}_k \in \mathbb{R}^{N_i \times N_j}$ of the zero-normalized 2D cross-correlation (ZNCC) between them. We use FFT-based correlation for computational efficiency and thus compute

$$\boldsymbol{C}_k = \frac{1}{N_i N_j} \mathcal{F}^{-1}\left(\tilde{\boldsymbol{W}}_{0,k} \tilde{\boldsymbol{W}}_{1,k}\right), \tag{1}$$

where $\mathcal{F}$ denotes the 2D discrete Fourier transform (DFT) and the 2D DFT of the normalized windows is defined as

$$\tilde{\boldsymbol{W}}_{n,k} = \mathcal{F}\left(\frac{\boldsymbol{W}_{n,k} - \mu_{n,k}}{\sigma_{n,k}}\right) \qquad \text{for } n \in \{0,1\}, \tag{2}$$

with $\mu_{n,k}$ and $\sigma_{n,k}$ representing the mean value and standard deviation of the respective window $\boldsymbol{W}_{n,k}$. From the ZNCC coefficients $\boldsymbol{C}_k$ we determine a first estimate of the displacement $\hat{\boldsymbol{d}}(x_k, z_k)$, by extracting the positions of the maximal correlation value for each window pair k. Using the correlation coefficients $\boldsymbol{C}_k$ around each peak, we apply 2D Gaussian regression to increase the displacement estimation accuracy to a sub-pixel level, as suggested in [7]. At the end of each pass, the resulting displacement vector field is added to the running total of all previous passes.

Between passes, the current estimate $\hat{\boldsymbol{d}}(x_k, z_k)$ is processed by first removing statistical outliers and then smoothing the resulting displacement vector field, using the robust unsupervised smoothing algorithm proposed in [8]. Finally, the resulting displacement estimates are used to deform the original frame $\boldsymbol{X}_1$, using bivariate spline interpolation, which reduces the difference between the two input frames $\boldsymbol{X}_0$ and $\boldsymbol{X}_1$ at each pass.

III. Experiments

In this section, we provide a detailed description of the pipeline used to evaluate the preservation of the time-coherence between consecutive low-quality frames, that were restored using specifically trained CNNs [4]. We provide in-depth descriptions of the numerical phantoms, the imaging configuration and the exact speckle tracking parameters used to estimate the phantom velocities, as well as the metrics used to evaluate the quality of the vector flow estimations.

A. Numerical Phantoms

We exclusively use numerical dynamic phantoms to have precise control and knowledge of the exact displacement between frames. Three phantoms of identical geometric properties are considered, differing only in their echogenicity properties. For a fair evaluation of the time-coherence preservation to be possible, the phantoms have been chosen to only generate a limited amount of imaging artifacts, to still allow high-quality velocity estimations even for low-quality images.

All three phantoms, denoted A, B and C, are counter-clockwise spinning disks, embedded in a fully anechoic background, with a maximum velocity $v_{\max}$. The disks have a radius r and are centered at $(x_c, z_c)^\mathsf{T}$ with respect to the transmitter array. The relevant values are listed in Table I. The reflection coefficients inside the respective phantoms are chosen to result in the dB-compressed echogenicity distributions specified in Table I. Inside the phantom, a density of 20 points per resolution cell has been used to ensure a fully diffusive speckle pattern.

TABLE I
SPINNING DISK PHANTOM PARAMETERS

Phantom	x_c [mm]	z_c [mm]	r [mm]	$v_{\max}$ [m/s]	dB-compressed echogenicity
A					uniform
B	0	27.5	10	3	angular, linear gradient of 40 dB
C					radial, linear gradient of 40 dB

B. Imaging Configuration

All RF channel data are generated using an in-house spatial impulse response model simulator, which accounts for 3D effects and element directivity. The used imaging setup is identical to the one used in the PICMUS challenge[1]. Single PW acquisitions with normal incidence of the phantoms are simulated at a 5 kHz pulse repetition frequency (PRF).

From each PW pulse-echo measurement, a low-quality RF image is reconstructed, using conventional DAS beamforming with spline interpolation and a receive apodization based on the element directivity. The image grid limits are chosen according to the PICMUS settings and the grid is spaced with $(\Delta x, \Delta z) = (\lambda/4, \lambda/8)$. The low-quality images are then either fed into U-Net16 or U-Net32, as described in [4], or are left unaltered for comparison purposes.

Eventually, before speckle tracking is applied, the envelope of the each RF image is extracted using the Hilbert transform along the z-axis, and the z-dimension is then down-sampled by a factor of two, resulting in a square, uniform sampling grid.

C. Speckle Tracking Settings

Our phantoms were designed to have geometric properties and velocity-to-effective-PRF ratios, that are very similar to the spinning disk phantom setup provided by the SA-VFI challenge [3]. For this reason, and because of the similarity of the deployed speckle tracking algorithm, we use the same four-pass coarse-to-fine PIV configuration with square window sizes of 4 mm, 2.5 mm, 2 mm and 1 mm, and a window overlap of 65 % as it was proposed in [5]. For our purpose, however, neither ensemble-cross-correlation nor any other averaging method is deployed. Thus, for each velocity estimation, a single pair of frames is used, reconstructed as previously described from two consecutive PW acquisitions.

D. Performance Evaluation

For each combination of one of the phantom types, i.e. A, B or C, with one of the reconstructed image, i.e. low-quality (LQ), U-Net16 or U-Net32, $M = 400$ pairs of consecutive frames are evaluated resulting in M independent velocity vector field estimations $\hat{\boldsymbol{v}}_m(x,z)$, where $m \in \{0, 1, \ldots, M-1\}$. The estimations $\hat{\boldsymbol{v}}_m(x,z)$ are compared to the theoretic velocity $\boldsymbol{v}(x,z) = v_{\max}/r\,(x - x_c, z - z_c)^\mathsf{T}$ inside a spinning disk.

Mean local bias $\overline{e_\mu}$ and mean local standard deviation $\overline{e_\sigma}$ are used as performance metrics. Given M independent scalar field estimations $\hat{p}_m(x,z)$, the corresponding theoretical reference values $p(x,z)$, as well as a discrete sampling grid composed of K points (x_k, z_k), these metrics are defined as

$$\overline{e_\mu}(\hat{p}) = \frac{1}{K}\sum_{k=0}^{K-1}\left|e_\mu(\hat{p}(x_k,z_k))\right|, \tag{3}$$

$$\overline{e_\sigma}(\hat{p}) = \sqrt{\frac{1}{K}\sum_{k=0}^{K-1}e_\sigma^2(\hat{p}(x_k,z_k))}, \tag{4}$$

where the local error bias and standard deviation are defined as

$$e_\mu(\hat{p}(x_k,z_k)) = \frac{1}{M}\sum_{m=0}^{M-1}\hat{p}_m(x_k,z_k) - p(x_k,z_k), \tag{5}$$

$$e_\sigma(\hat{p}(x_k,z_k)) = \sqrt{\frac{1}{M}\sum_{m=0}^{M-1}\left(\hat{p}_m(x_k,z_k) - \overline{\hat{p}}(x_k,z_k)\right)^2}, \tag{6}$$

and $\overline{\hat{p}}(x_k,z_k)$ denotes the scalar field average of the M independent estimations.

The metrics are calculated on both the normalized estimated velocity magnitudes $\|\hat{\boldsymbol{v}}_m(x,z)\|/v_{\max}$ and the estimated velocity angles $\varphi(\hat{\boldsymbol{v}}_m(x,z))$, for the natural sampling grid of the last speckle tracking pass, with parameters as previously defined. In the case of the velocity angle, circular statistics are used.

TABLE II
VELOCITY ESTIMATION QUALITY METRICS

Phantom	Image Quality	$\overline{e_\mu}(\|\hat{\boldsymbol{v}}\|)$ [%]	$\overline{e_\sigma}(\|\hat{\boldsymbol{v}}\|)$ [%]	$\overline{e_\mu}(\varphi(\hat{\boldsymbol{v}}))$ [°]	$\overline{e_\sigma}(\varphi(\hat{\boldsymbol{v}}))$ [°]
A	LQ	**2.27**	3.69	**1.98**	3.94
	U-Net16	2.44	3.45	2.13	3.97
	U-Net32	2.39	**3.34**	2.11	**3.83**
B	LQ	**2.72**	5.61	**2.25**	10.14
	U-Net16	2.98	5.24	2.48	9.97
	U-Net32	2.92	**5.13**	2.39	**9.65**
C	LQ	2.76	6.88	**1.95**	13.01
	U-Net16	2.40	3.85	2.22	**5.85**
	U-Net32	**2.35**	**3.76**	2.17	6.07

IV. RESULTS AND DISCUSSION

As mentioned before, the simplicity of the chosen phantoms enables a rather accurate vector flow estimation even using low-quality images, which is confirmed by the achieved results shown in Table II. Overall, we can see that while using the CNNs leads to a slightly worse velocity estimation bias, it increases the robustness of the estimation slightly or even quite significantly, as in the case of phantom C. Generally, the differences are below one percent and thus the influence of the CNNs on the time-coherence between consecutive US frames can be considered insignificant. In the case of phantom C, for which qualitative representations of the estimated vector flow using LQ and U-Net32 enhanced images are shown in Fig. 1, the velocity estimations are even improved using the CNNs.

Analyzing the local errors (e_μ, e_σ) depicted in Fig. 2 and 3, we can conclude that the estimation improvement is particularly located in the low-intensity areas (towards disk center) of

[1] https://www.creatis.insa-lyon.fr/EvaluationPlatform/picmus/about_settings.html

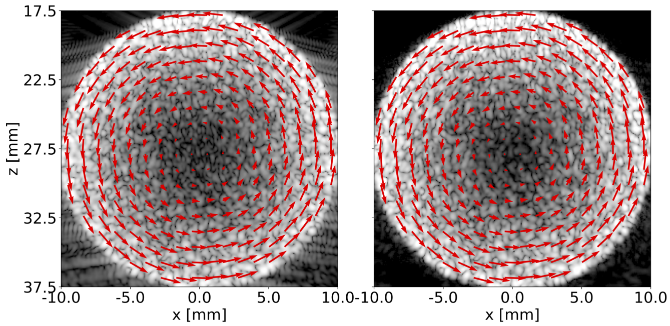

Fig. 1. Vector flow estimation on phantom C using LQ (left) and U-Net32 enhanced (right) images. The estimated vector flow is represented as red arrows overlayed over one of the two images used for the respective estimations. The images are displayed in B-Mode with a dynamic range of 50 dB.

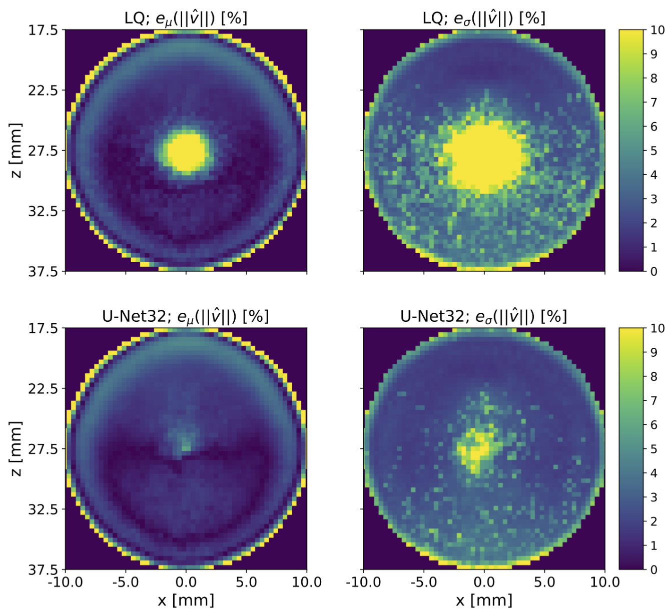

Fig. 2. Local velocity magnitude estimation error bias and standard deviation for both LQ and U-Net32 enhanced images of phantom C.

phantom C. This behavior seems to come from side-lobe artifacts originating from the high-intensity phantom regions (towards disk border), interfering with the speckle pattern inside the low-intensity regions of phantom C. This would imply that the CNNs are not only successful in reducing said artifacts, but also in restoring the underlying speckle patterns, including the contained physical information.

V. Conclusion

We have proposed to assess the preservation of time-coherence of a specific CNN-based image reconstruction approach, by comparing vector flow estimations of consecutive US images reconstructed using the CNN-based approach to the estimations obtained from images reconstructed using conventional DAS beamforming. The vector flow estimation was performed using a state-of-the-art speckle tracking algorithm, whose performance relies heavily on an intact time-coherence of the analyzed images. Using this approach, we were able to

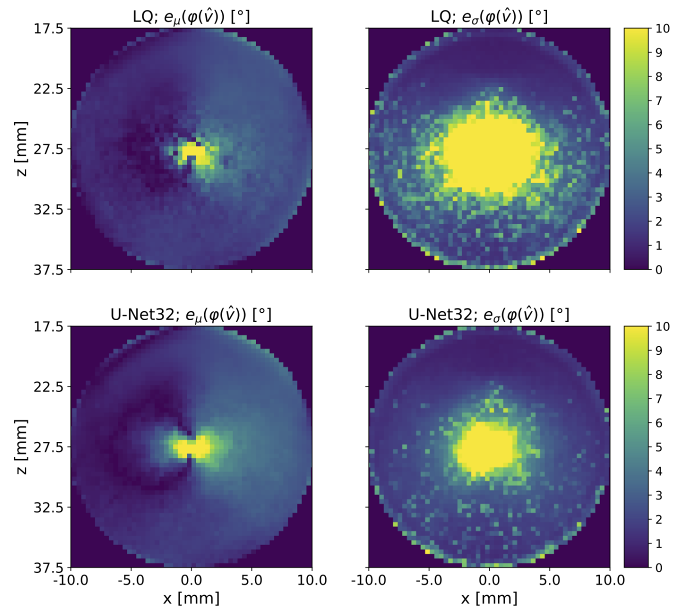

Fig. 3. Local velocity angle estimation error bias and standard deviation for both LQ and U-Net32 enhanced images of phantom C.

demonstrate that the vector flow estimation is not altered by the CNN in simple configurations where the phantom echogenicity is constant, hence guaranteeing the time-coherence preservation of the proposed CNN-based image restoration method.

Furthermore, we have shown that the use of the CNN-based restoration method proposed in [4] seems to have the potential to improve the resulting vector flow estimations in more challenging environments where the analyzed displacement spans a large dynamic range. This observation, which will be investigated in further studies, could be highly beneficial to applications in which only a very small amount of acquisitions per frame are allowed, such as in cardiac elastography.

References

[1] H. Liebgott, A. Rodriguez-Molares, F. Cervenansky, J. Jensen, and O. Bernard, "Plane-Wave Imaging Challenge in Medical Ultrasound," in *2016 IEEE Int. Ultrason. Symp.*, 2016, pp. 1–4.

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

[3] J. A. Jensen, H. Liebgott, F. Cervenansky, and C. A. Villagomez Hoyos, "SA-VFI: the IEEE IUS Challenge on Synthetic Aperture Vector Flow Imaging," in *2018 IEEE Int. Ultrason. Symp.*, 2018, pp. 1–5.

[4] D. Perdios, M. Vonlanthen, A. Besson, F. Martinez, M. Arditi, and J.-P. Thiran, "Deep Convolutional Neural Network for Ultrasound Image Enhancement," in *2018 IEEE Int. Ultrason. Symp.*, 2018, pp. 1–4.

[5] V. Perrot and D. Garcia, "Back to Basics in Ultrasound Velocimetry: Tracking Speckles by Using a Standard PIV Algorithm," in *2018 IEEE Int. Ultrason. Symp.*, 2018, pp. 206–212.

[6] W. Thielicke and E. J. Stamhuis, "PIVlab – Towards User-friendly, Affordable and Accurate Digital Particle Image Velocimetry in MATLAB," *J. Open Res. Softw.*, vol. 2, no. 1, 2014.

[7] H. Nobach and M. Honkanen, "Two-dimensional Gaussian regression for sub-pixel displacement estimation in particle image velocimetry or particle position estimation in particle tracking velocimetry," *Experiments in Fluids*, vol. 38, no. 4, pp. 511–515, 2005.

[8] D. Garcia, "Robust smoothing of gridded data in one and higher dimensions with missing values," *Computational Statistics & Data Analysis*, vol. 54, no. 4, pp. 1167–1178, 2010.