Shear-Wave Particle-Velocity Estimation and Enhancement Using a Multi-Resolution Convolutional Neural Network

Ultrasound in Medicine & Biology 49 (2023) 1518–1526

Contents lists available at ScienceDirect

Ultrasound in Medicine & Biology

journal homepage: www.elsevier.com/locate/ultrasmedbio

Original Contribution

Shear-Wave Particle-Velocity Estimation and Enhancement Using a Multi-Resolution Convolutional Neural Network

Xufei Chen *, Nishith Chennakeshava, Rogier Wildeboer †, Massimo Mischi, Ruud J.G. van Sloun

Lab of Biomedical Diagnostics, Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands

† *Philips Research Eindhoven, Eindhoven, The Netherlands*

ARTICLE INFO

Keywords:
Particle velocity estimation
Shear wave elastography
Ultrasound
Convolutional neural network
Deep learning

ABSTRACT

Objective: Tissue mechanical properties are valuable markers for tissue characterization, aiding in the detection and staging of pathologies. Shear wave elastography (SWE) offers a quantitative assessment of tissue mechanical characteristics based on the SW propagation profile, which is derived from the SW particle motion. Improving the signal-to-noise ratio (SNR) of the SW particle motion would directly enhance the accuracy of the material property estimates such as elasticity or viscosity.

Methods: In this paper, we present a 3-D multi-resolution convolutional neural network (MRCNN) to perform improved estimation of the SW particle velocity V_z. Additionally, we propose a novel approach to generate training data from real acquisitions, providing high SNR ground truth target data, one-to-one paired to inputs that are corrupted with real-world noise and disturbances.

Discussion: By testing the network on *in vitro* data acquired from a commercial breast elastography phantom, we show that the MRCNN outperforms Loupas' autocorrelation algorithm with an improved SNR of 4.47 dB for the V_z signals, a two-fold decrease in the standard deviation of the downstream elasticity estimates, and a two-fold increase in the contrast-to-noise ratio of the elasticity maps. The generalizability of the network was further demonstrated with a set of *ex vivo* porcine liver data.

Conclusion: The proposed MRCNN outperforms the standard autocorrelation method, in particular in low SNR regimes.

Introduction

Tissue mechanical properties are valuable markers for the discrimination of benign from malignant tissue and other lesions, aiding in the detection and staging of pathologies. Shear wave elastography (SWE) [1] offers a quantitative assessment of tissue mechanical characteristics. Acoustic radiation force (ARF) induced SWE uses an ARF push to locally displace the tissue in the axial direction, creating laterally propagating shear waves. Mechanical properties, such as tissue stiffness or viscosity, can be assessed based on the shear wave propagation profile, derived from the shear wave particle displacement (U_z) or velocity (V_z). The signal-to-noise ratio (SNR) of the shear wave particle motion directly impacts the accuracy of the downstream estimates. The SNR decreases when the ARF push energy is lower [2], the focal depth is deeper, or due to attenuation as the shear wave recedes from the push location [3]. At the same time, minimizing the ARF push energy is of clinical relevance, as lower mechanical and thermal indices are required for clinical safety.

Currently, two main classes of shear wave particle motion estimators are widely used: cross-correlation methods and phase-shift methods. Cross-correlation algorithms operate on the acquired radio-frequency (RF) data and maximize the cross-correlation coefficient between echo signals before and after deformation from subsequent frames. The change in the arrival time of the echoes is used to compute the particle displacement. Although they provide accurate estimates, these methods are computationally expensive. The phase-shift estimators, also known as autocorrelators [4,5], calculate the particle velocity by measuring the phase shift of the demodulated in-phase and quadrature (IQ) signals between subsequent frames. These methods are less computationally heavy and more suitable for real-time applications.

Recently, inspired by the application of convolutional neural networks (CNN) to optical flow problems, deep-learning based tissue displacement estimators have emerged in the field of ultrasound elastography (USE). Kibria *et al.* [6] proposed the GLUENet, a combination of the FlowNet 2.0 [7], an existing effective CNN for optical flow estimation, and the GLUE algorithm [8], a robust optimization-based displacement estimation method. GLUENet takes advantage of the convolutional structure to rapidly calculate coarse time delay estimates that are subsequently fine-tuned by optimizing a global cost function. Peng *et al.* [9] investigated the capability of three existing CNNs to estimate tissue displacement, namely the FlowNet-CSS [7], the PWC-Net

* Corresponding author.: Tel.: +33619475365.
E-mail address: x.chen6@tue.nl (X. Chen).

https://doi.org/10.1016/j.ultrasmedbio.2023.02.004
Received 12 August 2022; Revised 1 February 2023; Accepted 7 February 2023

[10], and the LiteFlowNet [11], before and after network retraining using simulated elastographic datasets. The three networks, originally designed for optical flow estimation, were capable of deriving tissue displacement for simulated, phantom, and *in vivo* breast data, and showed statistically significant improvements after retraining using elastographic data.

To reduce the loss of information in the RF data while downsampling due to the pyramidal structure of the network, Tehrani *et al.* [12] incorporated the analytical signal and B-mode image in addition to the RF data as inputs to the LiteFlowNet. Tehrani and Rivaz [13] also proposed two modified versions of the PWC-Net [10]: MPWC-Net and RFMPWC-Net to better suit ultrasonic RF data. The two models outperformed the original PWC-Net and the FlowNet 2.0 with ten times fewer parameters. The MPWC-Net was further improved into the MPWC-Net++ [14], which is better at estimating large displacements and requires shorter inference time. Østvik *et al.* [15] implemented a myocardial motion estimator for clinical functional use based on a PWC-Net architecture, named EchoPWC-Net. Besides optical flow inspired networks, Chan *et al.* [16] proposed the first micron-level ARFI displacement CNN and showed promising results with *in vivo* human prostate data for cancer diagnostics.

Aside from supervised learning, semi-supervised or unsupervised methods have shown promise in USE while permitting training without the need of ground truth labels. Tehrani *et al.* [17] introduced a semi-supervised method to improve the displacement estimation in quasi-static USE. First, a CNN was trained with abundant labeled datasets with optical flow ground truth in a supervised fashion; then, the latter was fine-tuned with real ultrasound images in an unsupervised manner. Delaunay *et al.* [18] proposed an unsupervised approach for quasi-static USE using CNNs, by optimizing a surrogate cost function, composed of an image similarity (negative normalized cross correlation) and displacement regularization term. Later, Delaunay *et al.* [19] adapted this method for shear wave particle displacement estimation.

Due to the absence of a ground truth target when performing real acquisitions, current supervised methods are limited to training with simulated ultrasound elastography data. Although noise is added to the simulated data, data-driven methods would substantially benefit from training data that exhibits true, real-world noise statistics which are difficult to model. Unsupervised methods are trained without the need of labeled targets. These methods provide similar displacement estimation performance as conventional cross-correlation methods (often requiring extensive grid search), but accelerate the computation. In this work, we aim to improve the particle velocity estimates beyond the conventional methods. To achieve this, we propose a novel approach to generate training data from real acquisitions, and exploit higher quality acquisitions obtained through higher ARF push energy and plane wave compounding (PWC) for the target data preparation. In fact, training towards PWC images has been largely investigated for ultrasound imaging enhancement: to provide high quality images while granting a high temporal resolution [20–24]. The ground-truth high-SNR targets were generated based on an extensive, time-consuming, and higher-radiation-force acquisition that includes coherent spatial and frequency compounding; one-to-one paired to non-compounded, standard acquisition inputs that are corrupted with real-world noise and disturbances.

In addition, we propose a 3-D multi-resolution convolutional neural network (MRCNN) that estimates and enhances the shear wave particle velocity V_z from the acquired IQ data. The specific network architecture operates at multiple scales and combines the different scale results to form enhanced V_z estimates, providing better quality downstream elasticity measurements. Despite its 3-D structure, the proposed network remains shallow and relatively light, with 1.5 times fewer parameters than the LiteFlowNet, the lightest well-known optical flow network applied to ultrasound elastography.

Methods

Data preparation

As said, we propose to train neural networks for shear wave particle velocity estimation by generating high-quality targets with an extended transmit-receive sequence following high-pressure push pulses. To this end, we perform full PWC [25–27] and compound a total of 45 successive high-frame-rate plane-wave acquisitions comprising varying plane wave angles and frequencies to yield a single high-quality target. The data acquisition was performed on 15 different regions of a commercial elasticity phantom (Model 049, CIRS, Inc., Norfolk, VA, USA) with different stiffnesses and inclusions for variability, using the Verasonics L11-4v transducer on a Vantage 128^{TM} system (Verasonics, Inc., Kirkland, WA, USA). This strategy enables training on real *in vitro* data that includes realistic noise statistics.

Each of the sequences in the full PWC setup consists of the same push sequence (f_c = 4.5 MHz, 1500 cycles, 50 V, i.e. high-pressure) focused at 1.97 cm, followed by a diverse set of imaging sequences, comprising transmissions at 9 different angles between -7^o and 7^o (at equally spaced intervals of 2^o and an additional angle at 0^o), and 5 different frequencies between 5.25 MHz and 7.25 MHz (at intervals of 0.5 MHz). The IQ data of the 45 acquisitions were each processed using the Loupas 2-D autocorrelator to generate the shear wave particle velocity data with an axial window of 5 wavelengths (1.2 mm) and a temporal window of 6 frames (600 μs). The 45 sets of V_z data were then compounded into one high-quality target as described in Fig. 1, denoted as $V_{z,\ 50V-PWC}$. Note that a high frame rate of 10 kHz was preserved within the full PWC method. This sequence is time-consuming, data and memory/storage intensive, and requires high radiation force, but is only used to generate target data. The network input data was the IQ data of the single PW acquisition at 0^o and 6.25 MHz, following the same push sequence with a lower excitation voltage of 30 V, denoted as IQ_{30V-NC}. Each input dataset was acquired right after the acquisition of the target dataset, with the exact same setup configuration; this way the inputs and targets were perfectly aligned. The real and imaginary parts of the input IQ_{30V-NC}, and the target $V_{z,\ 50V-PWC}$, were normalized between -1 and 1 according to their respective minimum and maximum of the overall dataset for training.

Shear wave propagation data was acquired for 5 ms at 10 kHz, resulting in 50 frames per acquisition. We performed 15 such acquisitions on different regions of the phantom, of which 14 were used for training and 1 for validation (network hyper-parameter tuning).

The hold-out test data was acquired in the same way on a different phantom: a commercial breast elastography phantom (Model 059, CIRS, Inc., Norfolk, VA, USA). The $V_{z,\ 30V-NC}$ was derived from the IQ_{30V-NC} using the Loupas 2-D autocorrelator with an axial window of 5 wavelengths (1.2 mm) and a temporal window of 6 frames (600 μs). The $V_{z,\ 50V-PWC}$ was derived the same way as described in **Methods** *Data preparation*. Here, the focal depth was set to 1.23 cm due to the physical dimensions of this particular phantom.

In addition, an *ex vivo* tissue dataset was acquired from a piece of fresh porcine liver to further demonstrate the generalizability of the proposed network. The acquisition and processing were identical to previous.

Phantom and tissue details

The commercially-available elasticity phantom (Model 049A, CIRS, Inc., Norfolk, VA, USA) contained 4 stepped cylindrical inclusions of different Young's modulus: 6, 9, 36, 70 kPa, in a homogeneous background of 18 kPa. By varying the imaging planes, inclusion shapes as circle, semicircle and quarter circle were obtained.

The breast elastography phantom (Model 059, CIRS, Inc., Norfolk, VA, USA) had a background of approximately 10 to 15 kPa, and spherical inclusions of at least two times the background elasticity.

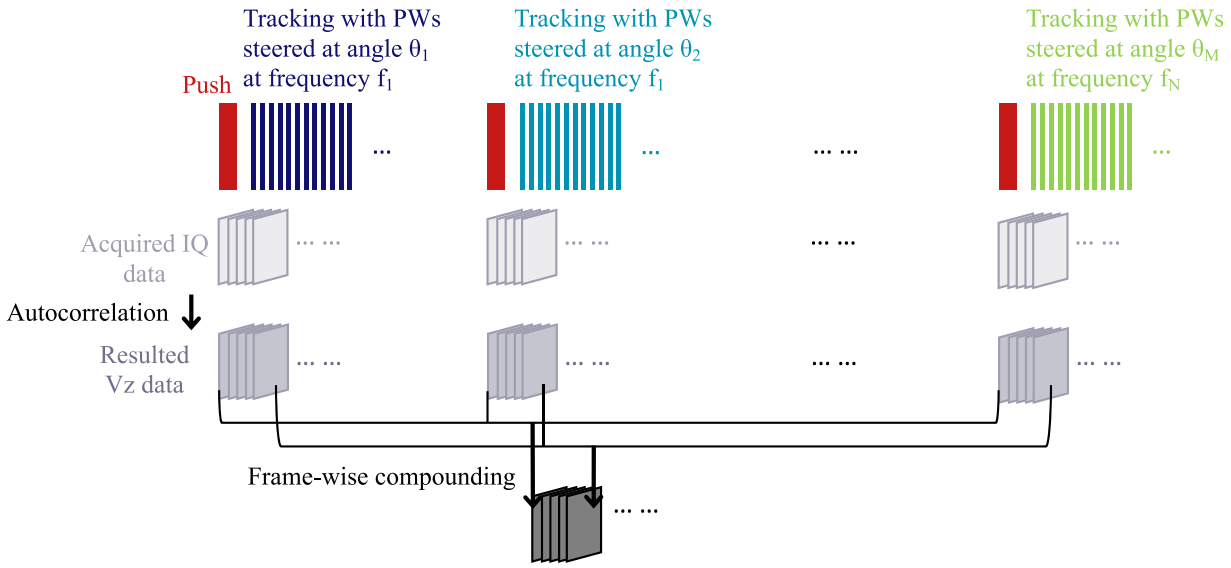

Fig. 1. Diagram of the full plane wave compounding (full PWC) method. The region of interest undergoes *MN* push and track sequences at different angles and frequencies. Frame-wise compounding is subsequently performed on the axial velocity data after autocorrelation to obtain high SNR V_z without compromising frame rate, which we use for training.

The piece of fresh porcine liver was purchased at a local butcher. It was subsequently put into a tank of degassed water for half an hour to degas. The imaging was done with the porcine liver in the degassed water tank, at 19^oC.

3-D Multi-Resolution Convolutional Neural Network

Inspired by the Loupas autocorrelator that exploits the slow time data in addition to the fast time data, we take 6 IQ_{30V-NC} frames to map to one $V_{z,\ 50V-PWC}$ frame. Furthermore, based on prior experimentation, training towards successive frames rather than individual frames provided additional temporal information, helping to improve the V_z estimation in the temporal dimension. Thus, the output data was arranged into blocks of five successive $V_{z,\ 50V-PWC}$ frames.

Therefore, the input data was arranged into blocks of ten successive IQ_{30V-NC} frames of size 200×160, with two channels for the real and imaginary parts. The output data was arranged into blocks of five successive $V_{z,\ 50V-PWC}$ frames of size 200×160. The axial size of the $V_{z,\ 50V-PWC}$ was 180, and zero-padded to 200 to ease the design and the training. The final results were truncated back to 180 axial pixels.

A 3-D multi-resolution CNN comprising three parallel branches with three different sizes of convolutional kernels was designed to map the IQ_{30V-NC} (of the low SNR $V_{z,\ 30V-NC}$ data) to the corresponding high SNR $V_{z,\ 50V-PWC}$ data, as described in Fig. 2. The individual branches were composed of four 3-D convolutional layers with [axial, lateral, temporal] kernel sizes of [5, 1, 2], [9, 1, 3] (alternated with [9, 3, 3]), and [15, 3, 6] for the so-called small, medium, and large branches, respectively. The three first convolutional layers had Exponential Linear Unit (ELU)

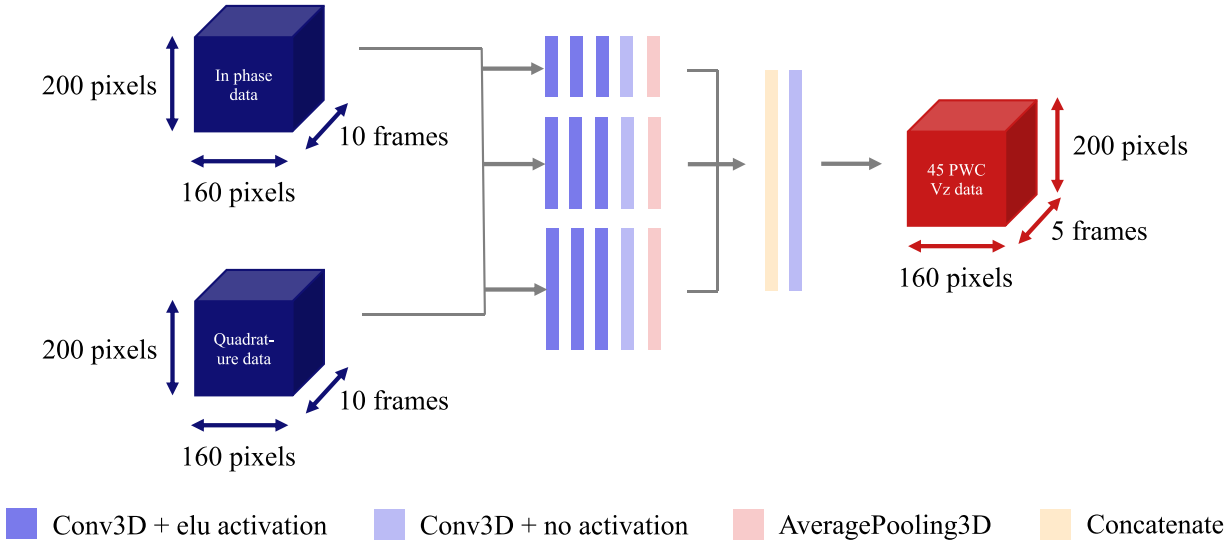

Fig. 2. Diagram of the 3-D multi-resolution convolutional neural network. The network comprised three branches of three different sizes of convolutional kernels. The input data was arranged into blocks of ten successive low SNR IQ_{30V-NC} frames of size 200×160, separately injected into the network through two channels for the real and imaginary parts. The output data was arranged into blocks of five successive high SNR $V_{z,\ 50V-PWC}$ frames of size 200×160. The axial size of the $V_{z,\ 50V-PWC}$ was 180, and zero-padded to 200 to ease the design and the training. The outputs were truncated back to 180 axial pixels, and underwent frame-wise averaging as described in **Methods** *3-D Multi-Resolution Convolutional Neural Network* and Fig. 3 to yield the final results $V_{z,\ CNN}$.

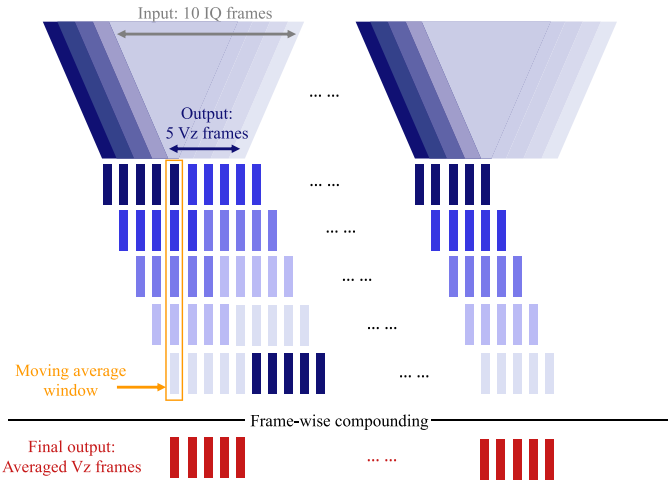

Fig. 3. Diagram of the moving average method for the final output generation. The outputs of the network were blocks of five successive V_z frames, thus, a moving average method was used to average the successive outputs frame-wise with a moving window to yield the final output $V_{z,\,\mathrm{CNN}}$.

activations. The 3-D convolutional layers operated on the two spatial and the temporal dimensions. Each branch was completed by an average pooling layer that operated on the temporal axis. The outputs of the three branches were concatenated and convolved to form one final output at the end of the network. Details of the architecture are shown in Fig. 2 and Appendix A. In early experimentation, we observed that the contrast of the velocity images was enhanced with the small- and medium-sized kernels, while larger-kernels offered improved denoising. This motivated the parallel multi-resolution architecture.

The final frames were formed from the output blocks (containing five frames) by averaging the overlapping frames resulted from successive estimations using a moving window (Fig. 3).

Training strategy

The network was optimized on the training and validation datasets. The parameters were tuned based on the training dataset; while the validation dataset was used to monitor the training process to avoid the network from overfitting and provide insight for the adjustment of the hyper-parameters. The hold-out test dataset remained unseen from the network during the entire training process, and was only used to test the trained network.

The network was trained towards each data point of the target dataset. Thus, 18,144,000 points (14 acquisitions of 45 V_z frames of size 180×160) were used for the training, and 1,296,000 data points (1 acquisition of 45 V_z frames of size 180×160) were used for the validation (network hyper-parameter tuning). No data augmentation technique was employed.

The network was trained by minimizing the mean squared error (MSE) loss between the target (x) and the network estimations ($\hat{x}$), given by,

$$\mathscr{L}_{MSE} = \frac{1}{N}\sum_{i=1}^{N}(x_i - f_\theta(y_i))^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - x_i)^2, \tag{1}$$

where $f_\theta(\cdot)$ is the neural network, parameterized by θ, N is the number of samples, and y is the input data.

The Adam optimizer was used with a learning rate of 10^{-5}, $\beta_1 = 0.5$ and $\beta_2 = 0.999$, performing stochastic gradient descent on batches of 4 randomly selected blocks of data. The model was implemented in Python, with the TensorFlow framework [28]. An NVIDIA GPU was used as an accelerator during training and inference [28].

The model was trained until convergence on the training dataset (based on the loss function (Eq. 1)), which was reached at 3450 epochs.

Validation methodology

Quality metrics used to evaluate the performance of the proposed network compared to the Loupas autocorrelator are presented below.

1. Root-mean-square error (RMSE)

The RMSE was calculated between the estimated and the reference V_z signals to quantify the axial velocity errors, with $V_{z,\,50\mathrm{V-PWC}}$ being the reference signal. Note that for the MRCNN results (starting from **Results** *MRCNN breast phantom results*), all V_z signals were normalized for fair comparison, thus the RMSE is unitless. The RMSE was calculated as,

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(X_i - \widehat{X}_i)^2}{N}}, \tag{2}$$

with N being the number of samples, X_i the reference V_z signal ($V_{z,\,50\mathrm{V-PWC}}$), and $\widehat{X}_i$ the estimated V_z signal.

2. Improved signal-to-noise ratio

To assess the denoising ability of the proposed network, the $\mathrm{SNR_{imp}}$ was calculated to quantify the mean SNR enhancement over the entire acquisition of the V_z signals in dB. The $\mathrm{SNR_{imp}}$ was defined as,

$$\mathrm{SNR_{imp}} = \frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N} 10\log_{10}\frac{|X_{\mathrm{noisy,m,n}} - X_{\mathrm{clean,m,n}}|^2}{|X_{\mathrm{denoised,m,n}} - X_{\mathrm{clean,m,n}}|^2}, \tag{3}$$

with M being the number of axial samples, N the number of lateral samples, $V_{z,\,30\mathrm{V-NC}}$ as X_{noisy}, $V_{z,\,50\mathrm{V-PWC}}$ as X_{clean}, and $V_{z,\,\mathrm{CNN}}$ as X_{denoised}.

3. Contrast-to-noise ratio (CNR)

The downstream elasticity maps were evaluated in terms of CNR between the lesion and the background. The higher the CNR, the better the image quality. The CNR was defined as,

$$\mathrm{CNR} = \frac{|S_{\mathrm{Inclusion}} - S_{\mathrm{Background}}|}{\sqrt{\sigma^2_{\mathrm{Inclusion}} + \sigma^2_{\mathrm{Background}}}}, \tag{4}$$

with S being the average value of the elasticity estimates and σ being the standard deviation of the elasticity estimates.

4. Normalized mean absolute error (NMAE)

For the ablation study that compares the performance of different network architectures, V_z and the downstream elasticity E were derived. The NMAE was computed between the elasticity estimates and the reference elasticity $E_{50\mathrm{V-PWC}}$ derived using $V_{z,\,50\mathrm{V-PWC}}$, to quantify and compare the elasticity errors, identifying the best network architecture. The NMAE was calculated as,

$$\mathrm{NMAE} = \frac{1}{N}\sum_{i=1}^{N}\frac{|x_i - \widehat{x}_i|}{x_i}, \tag{5}$$

with N being the number of samples, x_i the reference elasticity, and $\widehat{x}_i$ the estimated elasticity.

Results

Target data optimization

To optimize the target data for the network training, several tracking schemes were tested. The V_z signals at the farthest point from the push focus, i.e., the bottom right corner of the region of interest (ROI), are shown in Fig. 4.

The V_z signal resulting from a non-compounded (NC) single plane-wave tracking acquisition was highly noisy, to the extent that the peak axial velocity was masked by the noise. Although improved, the V_z signals following only frequency (1 angle × 5 frequencies) or spatial (9 angles × 1 frequency) compounding remained noisy. The RMSE at the farthest point from the push focus between the 45 PWC V_z and the V_z

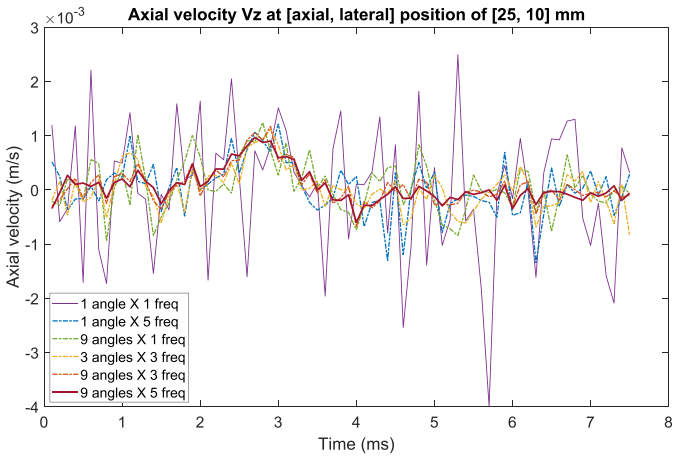

Fig. 4. Benchmark of different plane-wave compounding tracking schemes. The V_z signal following non-compounded single plane wave tracking was highly noisy, the peak axial velocity was masked by the noise. Although improved, V_z following only frequency (1 angle × 5 frequencies) or only spatial (9 angles × 1 frequency) compounding remained noisy. Combining the frequency and spatial compounding (3 angles × 3 frequencies) reduced the noise level further. For increasing number of employed angles and frequencies, the resulting V_z signal improves and the noise level decreases.

following frequency or spatial compounding only were 0.42 mm/s and 0.36 mm/s, respectively. Combining the frequency and spatial compounding strategies (3 angles × 3 frequencies), the RMSE decreased to 0.31 mm/s, while the same number of PWs was transmitted as in the spatial compounding only scheme. Thus, exploiting both frequency and spatial compounding is advantageous.

The SNR (calculated as in [26] and recapitulated in Appendix B) of the V_z signal improves with the number of transmitted PWs exploiting different frequencies and angles, although with a decreasing rate (Fig. 5). The 45 PWC V_z showed the highest SNR with 34.33 dB. Considering the decrease in the signal enhancement rate, the long acquisition time, the extensive data memory and storage requirements, the 45 PWC V_z was considered a good compromise between the signal quality and the abovementioned effects, and was thus chosen as the target data to train the network.

MRCNN breast phantom results

To demonstrate the proposed network's ability to derive and denoise V_z from IQ, the IQ_{30V-NC} of the breast elastography phantom (hold-out test data) was input to the trained network to estimate the axial velocity,

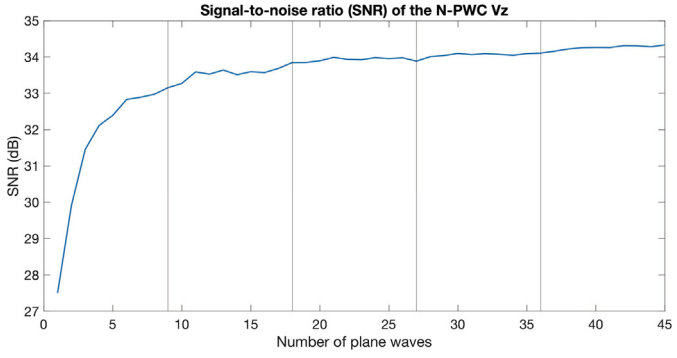

Fig. 5. Signal-to-noise ratio (SNR) of the N-PWC V_z. The SNR of the N-PWC V_z increased drastically for the first few compounded PWs. Despite the enhancing trend of the SNR with the increasing number of compounded PWs, the rate of this improvement went down.

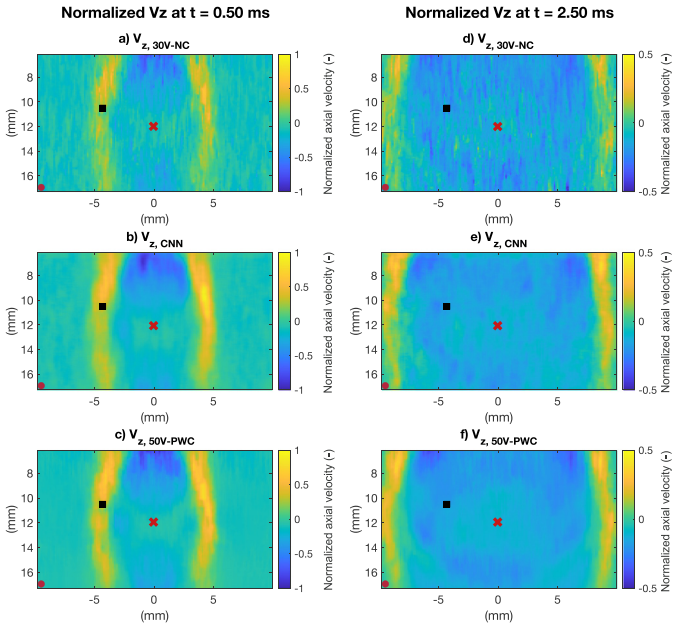

Fig. 6. Two frames from the SW propagation video. The push focus at the center of the ROI is shown by the red cross. The V_z signals at the black square and red dot markers are shown in Fig. 7.

$V_{z,\,CNN}$. The resulting $V_{z,\,CNN}$ was compared to $V_{z,\,30V-NC}$ and $V_{z,\,50V-PWC}$, derived using the conventional autocorrelation method as described in **Methods** *Data preparation*. Elasticity maps were then computed using Eq. 6 and the cross-correlation-derived shear wave speeds c_{SW} using $V_{z,\,30V-NC}$, $V_{z,\,CNN}$, and $V_{z,\,50V-PWC}$, denoted as E_{30V-NC}, E_{CNN}, and $E_{50V-PWC}$, respectively:

$$E = 3\rho c_{SW}^2, \tag{6}$$

with ρ being the material density, c_{SW} the shear wave speed derived from V_z.

In addition, each $V_{z,\,30V-NC}$ frame was smoothed using a 2-D median filter (kernel dimensions: 0.49 mm × 0.25 mm), denoted as $V_{z,\,30V-NC,\,smoothed}$, to calculate $E_{30V-NC,\,smoothed}$. This allowed us to verify the network denoised the input data or performed a simple smoothing.

Two frames of the shear wave propagation video at 0.50 ms (left) and 2.50 ms (right) after the push are shown in Fig. 6 for the three methods. For the frame at 0.50 ms, the dynamic range is between -1 and 1 (normalized values), and between -0.5 and 0.5 for the frame at 2.50 ms, for a better visualization of the propagating shear waves. The V_z signals at a point close to the push (black square in Fig. 6) and a point far from the push (red dot in Fig. 6) are shown in Fig. 7.

Shortly after the push (t = 0.50 ms), the shear waves are clearly distinct from the background for the three methods. Qualitatively, $V_{z,\,CNN}$ appears less noisy and shows more clearly pronounced wavefronts compared to $V_{z,\,30V-NC}$; and is similar to $V_{z,\,50V-PWC}$. After 2.50 ms of propagation, the wavefronts of $V_{z,\,CNN}$ remain clear and continuous, similar to $V_{z,\,50V-PWC}$, in contrast with $V_{z,\,30V-NC}$, which appears more noisy and presents discontinuous wavefronts.

Figure 6 shows the V_z temporal signals at a close and a far point from the push focus. At the close point (black square in Fig. 6), the peak axial velocity is clear for the three V_z signals, with a lower SNR for $V_{z,\,30V-NC}$. At the far point (red dot in Fig. 6), $V_{z,\,30V-NC}$ becomes very noisy and the peak axial velocity is masked by the noise; while $V_{z,\,CNN}$ presents low noise, and the peak axial velocity is clearly visible. At both locations, the $V_{z,\,CNN}$ is similar and comparable to $V_{z,\,50V-PWC}$.

These observations were validated with the RMSE (Eq. 2) and the improved SNR (Eq. 3). In fact, the RMSE of $V_{z,\,NC-30V}$ at the black square and the red dot, was 0.060 and 0.065, respectively; the RMSE of $V_{z,\,CNN}$ was 0.035 and 0.020, respectively. The mean RMSE over the entire

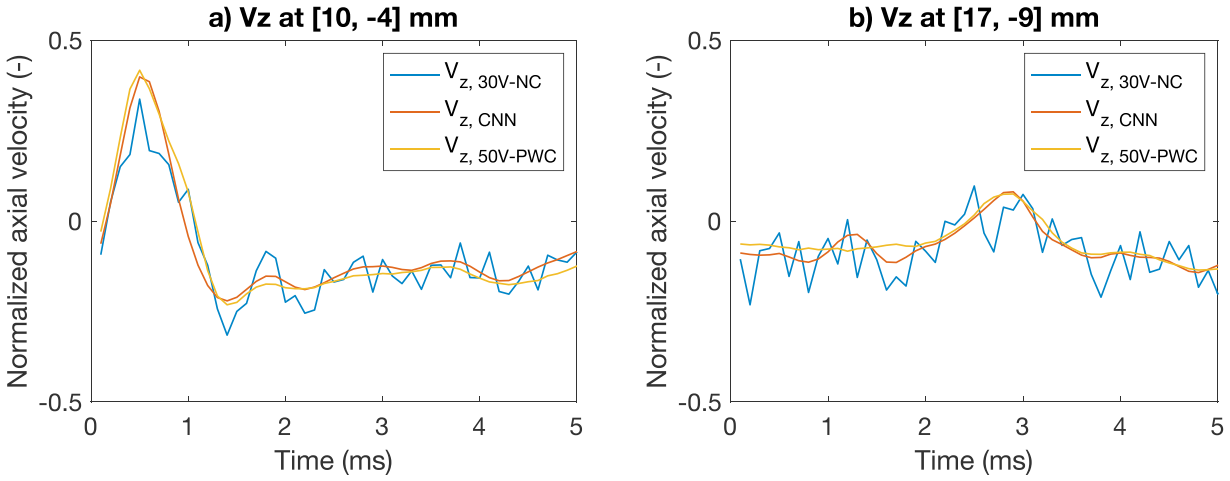

Fig. 7. V_z signals at the black square (a) and red dot (b) in Fig. 6. The RMSE at a point close to the push focus (a) is 0.060 for $V_{z,\ \mathrm{NC-30V}}$, and 0.035 for $V_{z,\ \mathrm{CNN}}$. The RMSE at a point far from the push focus (b) is 0.065 for $V_{z,\ \mathrm{NC-30V}}$, and 0.020 for $V_{z,\ \mathrm{CNN}}$.

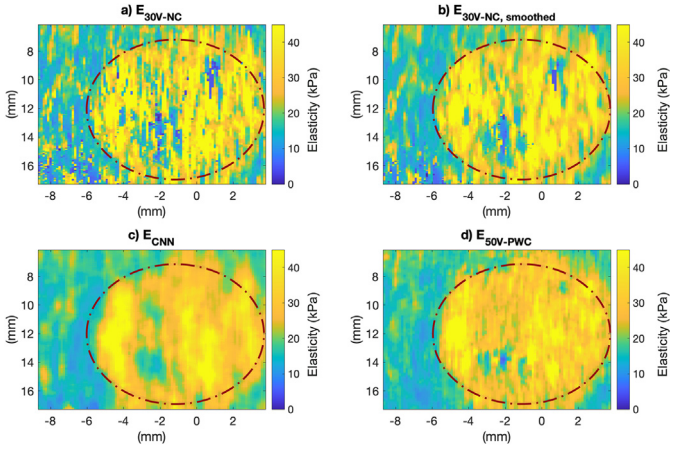

Fig. 8. Elasticity maps of the breast phantom CIRS 059. The CNR for $E_{30\mathrm{V-NC}}$, $E_{30\mathrm{V-NC,\ smoothed}}$, E_{CNN} and $E_{50\mathrm{V-PWC}}$ are 0.61, 0.75, 1.17 and 1.22, respectively.

dataset was 0.057 for $V_{z,\ \mathrm{NC-30V}}$, and 0.036 for $V_{z,\ \mathrm{CNN}}$. The SNR improvement was 4.47 dB.

Figure 8 shows the elasticity maps of the breast phantom. The distribution of elasticity within the inclusion and the background of the breast phantom are summarized in box plots (Fig. 9). The standard deviation within the inclusion is 23.01, 17.71, 8.67, 9.23, and 12.03, 8.92, 5.01, 5.14 in the background for $E_{30\mathrm{V-NC}}$, $E_{30\mathrm{V-NC,\ smoothed}}$, E_{CNN}, and $E_{50\mathrm{V-PWC}}$, respectively. The contrast-to-noise ratio (CNR) calculated using Eq. 4 of the breast phantom was 0.61, 0.75, 1.17, 1.22 for the elasticity maps derived using $V_{z,\ 30\mathrm{V-NC}}$, $V_{z,\ 30\mathrm{V-NC,\ smoothed}}$, $V_{z,\ \mathrm{CNN}}$, and $V_{z,\ 50\mathrm{V-PWC}}$, respectively.

An ablation study was conducted to compare the single-resolution networks (with different convolutional kernel sizes) with the multi-resolution networks. The MSE losses were computed between each $V_{z,\ \mathrm{CNN}}$ and $V_{z,\ 50\mathrm{V-PWC}}$ frames using Eq. 1. The RMSE was calculated between each $V_{z,\ \mathrm{CNN}}$ and $V_{z,\ 50\mathrm{V-PWC}}$ temporal signals at every spatial pixel using Eq. 2. The normalized mean absolute error (NMAE) was estimated between the elasticity maps E_{CNN} and $E_{50\mathrm{V-PWC}}$ using Eq. 5. The results are presented in Table 1.

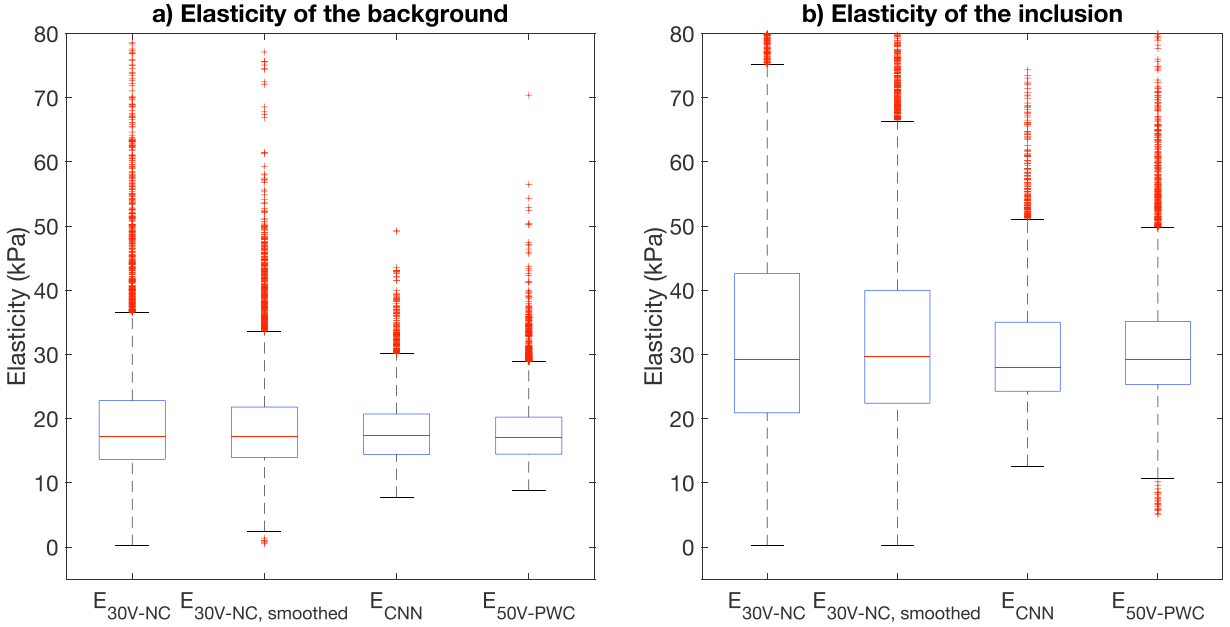

Fig. 9. Box plots of the downstream elasticity estimates computed using $V_{z,\ 30\mathrm{V-NC}}$, $V_{z,\ 30\mathrm{V-NC,\ smoothed}}$, $V_{z,\ \mathrm{CNN}}$, $V_{z,\ 50\mathrm{V-PWC}}$. The standard deviation is 12.03, 8.92, 5.01, 5.14 in the background, and 23.01, 17.71, 8.67, 9.23 in the inclusion, for $E_{30\mathrm{V-NC}}$, $E_{30\mathrm{V-NC,\ smoothed}}$, E_{CNN}, $E_{50\mathrm{V-PWC}}$, respectively.

Table 1
The following table summarizes the ablation study results.

Branch scale (size of the 3-D kernels)			Train towards	MSE	RMSE	NMAE
5×1×2	9×3×3	15×3×6	successive frames	**(mean ± std)**	**(mean ± std)**	**(mean ± std)**
✓	-	-	-	8.5×10^{-3} ± 3.4×10^{-3}	$8.5\times10^{-2}\pm3.6\times10^{-2}$	25.2 ± 25.6%
-	✓	-	-	6.2×10^{-3} ± 1.1×10^{-3}	$7.6\times10^{-2}\pm2.3\times10^{-2}$	15.8 ± 18.5%
-	-	✓	-	7.2×10^{-3} ± 5.5×10^{-3}	$7.9\times10^{-2}\pm3.2\times10^{-2}$	17.2 ± 19.2%
✓	✓	✓	-	4.2×10^{-3} ± 3.1×10^{-3}	$6.1\times10^{-2}\pm2.3\times10^{-2}$	14.7 ± 18.1%
✓	✓	✓	✓	1.5×10^{-3} ± 7.7×10^{-4}	$3.6\times10^{-2}\pm1.5\times10^{-2}$	13.5 ± 16.9%

Finally, the computational time was assessed for the three methods. For one IQ dataset of shape 200×160×50 (axial × lateral × temporal dimensions), the V_z was derived in 12.28 ± 0.18 s using the proposed network on an NVIDIA 2080 Ti GPU. The Loupas autocorrelator took 6.58 ± 0.35 s (or 35.80 ± 0.32 s) to derive a single PW acquisition, and 297.86 ± 7.33 s (or 26.85 ± 0.02 min) for the 45 PWC acquisition on a 1.7 GHz Quad-Core Intel Core i7 (Coffee Lake) CPU (or an NVIDIA 2080 Ti GPU, respectively).

MRCNN porcine liver results

The generalizability of the proposed network was further investigated using a set of *ex vivo* porcine liver data. In a similar way to the processing of the breast phantom data, the acquired IQ_{30V-NC} was input to the network to derive $V_{z,\,CNN}$. $V_{z,\,30V-NC}$ and $V_{z,\,50V-PWC}$ were calculated using the conventional autocorrelation method as described in **Methods** *Data preparation*. Elasticity maps (of size 8 mm × 3 mm) were subsequently computed using Eq. 6 and the cross-correlation-derived shear wave speeds c_{SW} using $V_{z,\,30V-NC}$, $V_{z,\,CNN}$, and $V_{z,\,50V-PWC}$, denoted as E_{30V-NC}, E_{CNN}, and $E_{50V-PWC}$, respectively.

The mean RMSE calculated using Eq. 2 over the entire dataset was 0.065 and 0.046 for $V_{z,\,30V-NC}$ and $V_{z,\,CNN}$, respectively. The SNR improvement was 1.96 dB. The standard deviation was 6.20, 2.25, 2.87 for E_{30V-NC}, E_{CNN}, and $E_{50V-PWC}$, respectively. The distribution of elasticity is summarized in box plots (Fig. 10).

Discussion

In this paper, a novel method is presented to generate high quality target data, one-to-one paired to input data that is corrupted with real-world noise and disturbances. The training data was prepared exclusively from real acquisitions, in which the ground truth target data was based on an extensive, time-consuming, and higher-radiation-force acquisition that includes coherent spatial and frequency compounding. The incorporation of realistic noise statistics in the training dataset is key to optimize the network's denoising ability. In addition, we propose a novel neural network capable of deriving and denoising V_z from IQ data. We demonstrated the generalizability of the proposed MRCNN with data from a phantom unknown to the trained network, and a piece of real tissue. In fact, the model was trained with data from the elasticity phantom CIRS 049A, whereas the presented results were derived from data acquired on a breast phantom (CIRS 059) and an *ex vivo* porcine liver sample with different mechanical properties and at a different depth (Figures 6 to 10). To the best of our knowledge, this is the first work in ultrasound elastography that exploits solely data from real acquisitions; using spatial and frequency compounding to generate ground truth labels, to train a network that estimates particle velocity from IQ data.

The proposed MRCNN extends beyond the conventional autocorrelator in terms of the quality of the derived V_z and the downstream elasticity estimates. While using the same IQ data, the MRCNN estimates axial velocities with an SNR 4.47 dB higher than that obtained with the Loupas autocorrelator for the breast phantom, and 1.96 dB for the *ex vivo* porcine liver sample. The standard deviation of the downstream elasticity measures is reduced by more than half: from 23.01 kPa to 8.67 kPa in the inclusion, from 12.03 kPa to 5.01 kPa in the background of the breast phantom (Fig. 9); and from 6.20 kPa to 2.25 kPa for the *ex vivo* porcine liver sample (Fig. 10). Furthermore, the CNR of the elasticity maps of the heterogeneous breast phantom shows almost a two-fold increase (Fig. 8), from 0.61 to 1.17. The MRCNN yielded comparable results to those derived by the Loupas autocorrelator, using IQ data from a higher quality acquisition.

Moreover, the comparison of E_{CNN} with $E_{30V-NC,\,smoothed}$ demonstrated that the network did not perform a simple smoothing, but denoised the input data. In fact, although smoothed, $E_{30V-NC,\,smoothed}$ remained noisy with high standard deviation for the elasticity estimates: 17.71 kPa in the inclusion and 8.92 kPa in the background. The CNR was slightly improved to 0.75, but remained low.

The authors would like to point out that an overestimation of the elasticity is observed for all the 4 datasets of the breast phantom (E_{30V-NC}, $E_{30V-NC,\,smoothed}$, E_{CNN}, and $E_{50V-PWC}$). In fact, all measurements (summarized in box plots in Fig. 9) were slightly higher than the ground truth values recapitulated in **Methods** *Phantom and tissue details*. Thus, we believe this overestimation to be associated with the shear wave speed estimator that we used in the post-processing (which is not the focus of this work), and not due to our network estimation; otherwise, the overestimation would occur only for E_{CNN}. Also for the porcine liver elasticity, the median values match across the 3 datasets (E_{30V-NC}, E_{CNN},

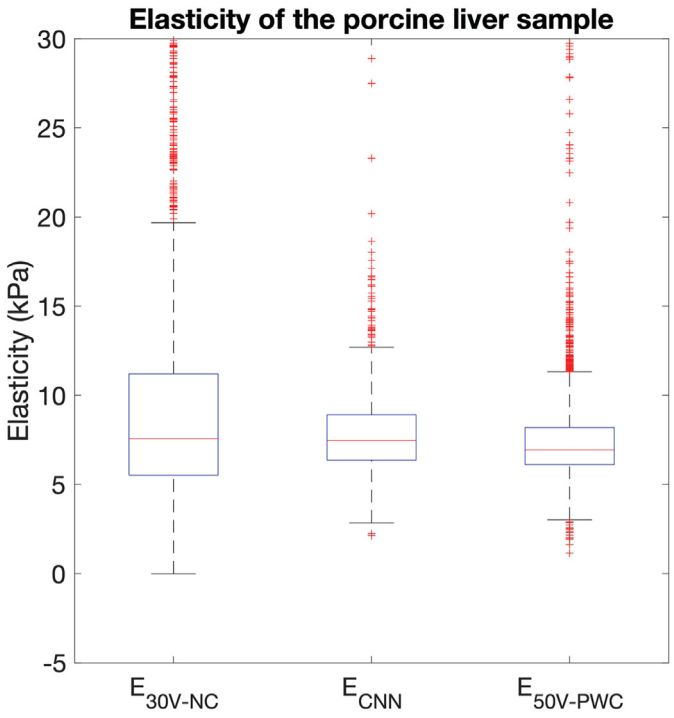

Fig. 10. Box plots of the downstream elasticity estimates computed using $V_{z,\,30V-NC}$, $V_{z,\,CNN}$, $V_{z,\,50V-PWC}$. The standard deviation is 6.20, 2.25, 2.87 for E_{30V-NC}, E_{CNN}, $E_{50V-PWC}$, respectively.

and $E_{50V-PWC}$), and are in accordance with the shear moduli μ (with $E \sim 3\mu$) reported in the literature [29,30]. Yet, our values are slightly higher, which we think again to be ascribed to the post-processing step. Here, a cross-correlation-based time-of-flight method was applied to derive the shear wave speed. We chose this method for its straightforwardness and accessibility. This approach does not take into account the medium's viscosity, which may cause such an overestimation. Future work should consider more sophisticated post-processing techniques including viscoelasticity estimators.

The ablation study (Table 1) shows that the multi-resolution CNN outperforms all single-resolution CNNs; and training towards successive frames rather than individual frames aids in the amelioration of the MRCNN's performance.

In this study, the computational time to derive $V_{z,\,CNN}$ is twice the time to derive $V_{z,\,30V-NC}$. However, $V_{z,\,CNN}$ is both qualitatively and quantitatively comparable to a 45 PW acquisition, exploiting both spatial and frequency compounding ($V_{z,\,50V-PWC}$). The computational time for such a signal is ~24 times the duration required to compute $V_{z,\,CNN}$. Besides the computational time, the full PWC acquisition is difficult to implement in the clinic. This method requires the region of interest to be perfectly identical for each PW acquisition for subsequent compounding, and unfortunately, tissue motion is inevitable. Furthermore, full PWC necessitates multiple pushes, while only one push is needed for $V_{z,\,CNN}$, minimizing acoustic exposure, acquisition time, and data memory/storage requirements.

In spite of its 3-D structure, the proposed model remains shallow and relatively light with only 3.36 million parameters. The lightest optical flow network used for displacement estimation in ultrasound elastography, the LiteFlowNet, has 5 million parameters; the heavier model, FlowNet 2.0, is composed of over 160 million parameters.

The model can be further improved by adding more training data of greater diversity, to increase its robustness. In this study, the network was trained with data from five different media, of varying stiffnesses from one commercial elasticity phantom only. Next to data from phantoms, it would be of interest to include clinical training data, to further generalize the model and hopefully translate this novel approach into clinical practice. The challenge, as mentioned above, would be to acquire target data *in vivo*, as performing full PWC in a clinical situation is difficult and hardly realizable due to tissue motion, high acoustic exposure, and intensive data storage that causes longer acquisition time. Two solutions can be considered: 1) *ex vivo* static tissues can be used in a first place; 2) phantoms of similar mechanical properties to biological tissues can be developed and used for full PWC acquisitions in a static setting.

Another potential enhancement can be the diversification of the push sequences. It is shown that the proposed model was able to derive results from an acquisition at a different focal depth from the one of the training data. It would be worthwhile to include different cycle numbers, push pressures, center frequencies, and/or focal depths to make an extensive network that is able to evaluate data acquired from different push regimes.

Alternative projects will consider unsupervised methods.

Conclusion

In this work, we propose a 3-D multi-resolution convolutional neural network trained with realistic noise statistics, capable of deriving V_z from the acquired IQ data while denoising and enhancing the estimates at the same time. Using the IQ data from a low quality acquisition (i.e. 30-V push excitation followed by single plane-wave tracking), the MRCNN outperforms the Loupas autocorrelator with an increase of 4.47 dB for the axial velocity signals, and a two-fold increase in the CNR of the elasticity maps. The estimations given by the MRCNN based on IQ data from low quality acquisition are qualitatively and quantitatively comparable to the autocorrelator-derived estimates using IQ data from high quality acquisition (i.e. 50-V push excitation followed by 45 PW tracking and compounding).

Further work would be expanding the training data with *ex vivo* acquisitions, and eventually validate the model with *in vivo* data, with the aim of extending the model for clinical practice.

Data availability statement

Data supporting this study are available from the corresponding author on request.

Conflict of interest statement

The authors have no conflict of interest to declare.

Acknowledgements

This research was conducted in the framework of the e/MTIC program, in collaboration with Philips Research, and supported by the Dutch Research Council (NWO) through the research program "High Tech Systems and Materials (HTSM)" under Project 17144.

The authors would like to express their gratitude to the reviewers for their time and constructive feedback that polished the manuscript. The authors would like to thank Ir. de Bruijn, Dr. Kuenen for their insightful inputs, and address their special thanks to Ir. Huang for his continuous help and bright suggestions.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.ultrasmedbio.2023.02.004

References

[1] Sarvazyan AP, Rudenko OV, Swanson SD, Fowlkes JB, Emelianov SY. Shear wave elasticity imaging: a new ultrasonic technology of medical diagnostics. Ultrasound in medicine & biology 1998;24:1419–35.
[2] Deng Y, Palmeri ML, Rouze NC, Rosenzweig SJ, Abdelmalek MF, Nightingale KR. Analyzing the impact of increasing mechanical index and energy deposition on shear wave speed reconstruction in human liver. Ultrasound in medicine & biology 2015;41:1948–57.
[3] Szabo TL. Diagnostic ultrasound imaging: inside out. Academic press; 2004.
[4] Kasai C, Namekawa K, Koyano A, Omoto R. Real-time two-dimensional blood flow imaging using an autocorrelation technique. IEEE Transactions on sonics and ultrasonics 1985;32:458–64.
[5] Loupas T, Powers JT, Gill RW. 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 1995;42:672–88.
[6] Kibria M, Rivaz H, et al. Gluenet: Ultrasound elastography using convolutional neural network. Simulation, Image Processing, and Ultrasound Systems for Assisted Diagnosis and Navigation. Springer; 2018. p. 21–8.
[7] Ilg E, Mayer N, Saikia T, Keuper M, Dosovitskiy A, Brox T. Flownet 2.0: Evolution of optical flow estimation with deep networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2017. p. 2462–70.
[8] Hashemi HS, Rivaz H. Global time-delay estimation in ultrasound elastography. IEEE transactions on ultrasonics, ferroelectrics, and frequency control 2017;64:1625–36.
[9] Peng B, Xian Y, Zhang Q, Jiang J. Neural-network-based motion tracking for breast ultrasound strain elastography: An initial assessment of performance and feasibility. Ultrasonic imaging 2020;42:74–91.
[10] Sun D, Yang X, Liu MY, Kautz J. Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2018. p. 8934–43.
[11] Hui TW, Tang X, Loy CC. Liteflownet: A lightweight convolutional neural network for optical flow estimation. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2018. p. 8981–9.
[12] Tehrani AKZ, Amiri M, Rivaz H. Real-time and high quality ultrasound elastography using convolutional neural network by incorporating analytic signal. 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE; 2020. p. 2075–8.
[13] Tehrani AKZ, Rivaz H. Displacement estimation in ultrasound elastography using pyramidal convolutional neural network. IEEE transactions on ultrasonics, ferroelectrics, and frequency control 2020;67:2629–39.
[14] Tehrani AKZ, Rivaz H. Mpwc-net++: evolution of optical flow pyramidal convolutional neural network for ultrasound elastography. Medical Imaging 2021: Ultrasonic

Imaging and Tomography, 11602. International Society for Optics and Photonics; 2021:1160206.
[15] Østvik A, Salte IM, Smistad E, Nguyen TM, Melichova D, Brunvand H, Haugaa K, Edvardsen T, Grenne B, Lovstakken L. Myocardial function imaging in echocardiography using deep learning. IEEE Transactions on Medical Imaging 2021;40:1340–51.
[16] Chan DY, Morris DC, Polascik TJ, Palmeri ML, Nightingale KR. Deep convolutional neural networks for displacement estimation in ARFI imaging. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control 2021;68:2472–81.
[17] Ali KZT, Mirzaei M, Rivaz H. Semi-supervised training of optical flow convolutional neural networks in ultrasound elastography. International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer; 2020. p. 504–13.
[18] Delaunay R, Hu Y, Vercauteren T. An unsupervised approach to ultrasound elastography with end-to-end strain regularisation. International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer; 2020. p. 573–82.
[19] Delaunay R, Hu Y, Vercauteren T. An unsupervised learning-based shear wave tracking method for ultrasound elastography. Medical Imaging 2022: Ultrasonic Imaging and Tomography, 12038. SPIE; 2022. p. 149–55.
[20] Gasse M, Millioz F, Roux E, Garcia D, Liebgott H, Friboulet D. High-quality plane wave compounding using convolutional neural networks. IEEE transactions on ultrasonics, ferroelectrics, and frequency control 2017;64:1637–9.
[21] Zhang X, Li J, He Q, Zhang H, Luo J. High-quality reconstruction of plane-wave imaging using generative adversarial network. 2018 IEEE International Ultrasonics Symposium (IUS). IEEE; 2018. p. 1–4.
[22] Khan S, Huh J, Ye JC. Universal plane-wave compounding for high quality us imaging using deep learning. 2019 IEEE International Ultrasonics Symposium (IUS). IEEE; 2019. p. 2345–7.
[23] Guo B, Zhang B, Ma Z, Li N, Bao Y, Yu D. High-quality plane wave compounding using deep learning for hand-held ultrasound devices. International Conference on Advanced Data Mining and Applications. Springer; 2020. p. 547–59.
[24] Chennakeshava N., Luijten B., Mischi M., Eldar Y.C., van Sloun R.J.G. Deep proximal learning for high-resolution plane wave compounding. arXiv preprint arXiv:2112.12410, 2021.
[25] Bae S, Kim P, Kim K, Jeong J, Song Tk. Enhanced shear wave elastography for HIFU monitoring of stiff uterine fibroids. In: Abstract presented at IEEE International Ultrasonics Symposium; 2018.
[26] Capriotti M, Greenleaf JF, Urban MW. Time-aligned plane wave compounding methods for high-frame-rate shear wave elastography: Experimental validation and performance assessment on tissue phantoms. Ultrasound in Medicine & Biology 2021;47:1931–48.
[27] Montaldo G, Tanter M, Bercoff J, Benech N, Fink M. Coherent plane-wave compounding for very high frame rate ultrasonography and transient elastography. IEEE transactions on ultrasonics, ferroelectrics, and frequency control 2009;56:489–506.
[28] Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, Corrado GS, Davis A, Dean J, Devin M, Ghemawat S, Goodfellow I, Harp A, Irving G, Isard M, Jia Y, Jozefowicz R, Kaiser L, Kudlur M, Levenberg J, Mané D, Monga R, Moore S, Murray D, Olah C, Schuster M, Shlens J, Steiner B, Sutskever I, Talwar K, Tucker P, Vanhoucke V, Vasudevan V, Viégas F, Vinyals O, Warden P, Wattenberg M, Wicke M, Yu Y, Zheng X. TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from tensorflow.org.
[29] Orescanin M, Qayyum MA, Toohey KS, Insana MF. Dispersion and shear modulus measurements of porcine liver. Ultrasonic imaging 2010;32:255–66.
[30] Chen S, Urban MW, Pislaru C, Kinnick R, Greenleaf JF. Liver elasticity and viscosity quantification using shearwave dispersion ultrasound vibrometry (SDUV). 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE; 2009. p. 2252–5.