Ultrasound Shear Wave Elasticity Imaging With Spatio-Temporal Deep Learning

Maximilian Neidhardt, Marcel Bengs, Sarah Latus, Stefan Gerlach, Christian J. Cyron, Johanna Sprenger, and Alexander Schlaefer

***Abstract*—Ultrasound shear wave elasticity imaging is a valuable tool for quantifying the elastic properties of tissue. Typically, the shear wave velocity is derived and mapped to an elasticity value, which neglects information such as the shape of the propagating shear wave or push sequence characteristics. We present 3D spatio-temporal CNNs for fast local elasticity estimation from ultrasound data. This approach is based on retrieving elastic properties from shear wave propagation within small local regions. A large training data set is acquired with a robot from homogeneous gelatin phantoms ranging from 17.42 kPa to 126.05 kPa with various push locations. The results show that our approach can estimate elastic properties on a pixelwise basis with a mean absolute error of 5.01(437) kPa. Furthermore, we estimate local elasticity independent of the push location and can even perform accurate estimates inside the push region. For phantoms with embedded inclusions, we report a 53.93% lower MAE (7.50 kPa) and on the background of 85.24% (1.64 kPa) compared to a conventional shear wave method. Overall, our method offers fast local estimations of elastic properties with small spatio-temporal window sizes.**

***Index Terms*—Elasticity imaging, 3D deep learning, high-speed ultrasound imaging, spatio-temporal data, soft tissue.**

I. Introduction

QUANTIFYING mechanical properties of soft tissue has many clinical applications ranging from diagnoses [1] to modeling soft tissue response for surgical planning [2]. Ultrasound shear wave elasticity imaging (US-SWEI) is widely used to image the elastic properties of tissue and its clinical applications has been widely demonstrated, e.g., in disease staging of breast lesions [3], thyroid nodules [4] or liver fibrosis [5]. In US-SWEI, an initial high energy acoustic radiation force impulse displaces the tissue. The propagation of the resulting shear wave is then captured with high frequency ultrasound imaging.

Shear wave velocity is commonly used as a surrogate for tissue elasticity, which can be estimated from a sequence of images, considering either the time-domain or the frequency domain. The first approach tracks the peak of the propagating shear wave, often referred to as time-of-flight (ToF). This can be achieved either by applying an autocorrelation of two time-varying signals with a known distance between each other ([6]–[9]) or by performing a linear regression of the wave peaks in a 2D space-time image ([10], [11]). Commonly, ToF-methods assume that shear waves propagate in a fixed direction. To estimate wave velocity independently of the propagating direction, 2D-autocorrelation methods were proposed ([7], [9]). In general, ToF-methods have been evaluated in the clinical setting [12]. However, estimates are dependent on imaging depth [13] and performance has been reported to be limited for stiffer materials [14], which are characterized by faster shear waves. The second approach for US-SWEI estimates the phase velocity of the dominant local wavenumbers in the frequency domain [15]. Similar to a 2D-autocorrelation, this approach is independent of the wave direction but requires intensive tuning of the imaging and filter parameters [16].

Recently, deep learning methods have gained popularity in strain elastography ([17]–[20]) and SWE-imaging ([21]–[23]). These methods allow estimates without intensive preprocessing of the data, manual tuning and do not rely on feature extraction, e.g., the shear wave velocity for elasticity estimation. Previous works have demonstrated that deep learning can be used to estimate distinct tissue parameters from SWEI data. Jin *et al.* [21] predict the shear wave velocity from space-time images including an uncertainty estimate and Vasconcelos *et al.* [22] have shown that the viscoelastic model parameters can be estimated from simulated shear wave motion data. Further, Ahmed *et al.* [23] demonstrated that elasticity maps and segmentation masks can be generated with deep learning from simulated SWEI data. However, the authors also note that the use of simulated data does not seem to be sufficient to represent the noise in real data.

In this study, we present spatio-temporal convolutional neural networks (CNNs) for reconstructing elasticity maps from real ultrasound SWEI data. Our approach is based on the concept of retrieving local elastic properties from shear wave propagation in small regions of several millimeters, which we call

Manuscript received 4 February 2022; revised 6 April 2022; accepted 7 April 2022. Date of publication 19 April 2022; date of current version 20 October 2022. This work was supported in part by the DFG SCHL under Grant 1844/2-2, in part by TUHH i^3 Initiative, and publishing fees supported by Funding Programme Open Access Publishing of Hamburg University of Technology (TUHH). *(Maximilian Neidhardt and Marcel Bengs contributed equally to this work.) (Corresponding author: Maximilian Neidhardt.)*

Marcel Bengs, Sarah Latus, Stefan Gerlach, Johanna Sprenger, and Alexander Schlaefer are with the Institute of Medical Technology and Intelligent Systems, Hamburg University of Technology, Germany.

Christian J. Cyron is with the Department of Continuum and Materials Mechanics, Hamburg University of Technology, Germany.

Maximilian Neidhardt is with the Institute of Medical Technology and Intelligent Systems, Hamburg University of Technology, 21073 Hamburg, Germany (e-mail: maximilian.neidhardt@tuhh.de).

Digital Object Identifier 10.1109/TBME.2022.3168566

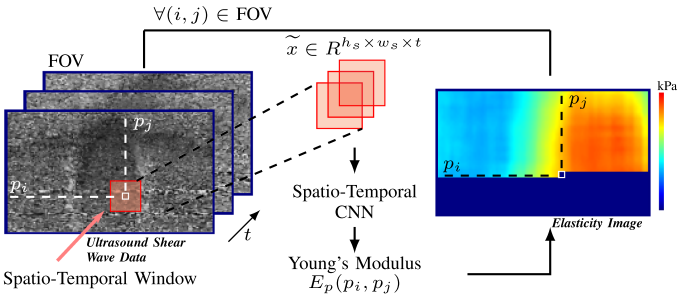

Fig. 1. Our approach for elasticity imaging with spatio-temporal deep learning. A global elasticity map is generated by estimating local elastic properties E_p at each pixel location $p = [p_i, p_j]$ with a spatio-temporal CNN considering a spatio-temporal window $\widetilde{x}$.

spatio-temporal windows (Fig. 1). By performing this local elastic estimation, the network learns the direct relationship between localized shear wave propagation and local elasticity. Our localized approach enables the generation of detailed elasticity maps of inhomogeneities and simplifies the required training data. In particular, this allows us to acquire training data from simple homogeneous phantoms with defined elasticities. Using real ultrasound training data, probe and ultrasound artifacts are directly included in our approach. In this context, we systematically study whether deep learning is able to extract relevant information from these limited areas and whether this approach generalizes to different push locations and elasticities. We evaluate our approach using tissue mimicking gelatin phantoms with Young's moduli ranging from 17.42 kPa to 126.05 kPa. Furthermore, we compare our spatio-temporal CNN approach to ToF shear wave estimation.

In summary, our 3D spatio-temporal CNN approach can estimate local elastic properties from spatio-temporal windows using real ultrasound shear wave data, while being independent of the push location and wave propagation direction. Furthermore, our approach can generate local elasticity maps of non-homogeneous mediums in real-time.

II. Methods

A. Deep Learning Model

We estimate the elasticity locally by applying spatio-temporal CNNs to a small spatio-temporal window as illustrated in Fig. 1. Hence, we perform pixelwise predictions using the neighborhood as context. Formally, given a sequence of images $x \in R^{h \times w \times t}$ which represent shear wave propagation over time with h and w for the spatial dimensions of the FOV and t for the temporal dimension, the elasticity is estimated locally using a spatio-temporal window $\widetilde{x} \in R^{h_s \times w_s \times t}, \widetilde{x} \subset x$ centered at pixel location $p = [p_i, p_j]$. The spatial dimensions of the spatio-temporal window are described by w_s and h_s. Hence, we design and evaluate an approach for learning $f : R^{h_s \times w_s \times t} \rightarrow R$. By using our spatio-temporal CNN to estimate elasticity for each pixel, an entire global elasticity map for x can be estimated, as shown in Fig. 1. The advantage of local elasticity estimation is that the network is trained to learn the relationship between elasticity and shear-wave propagation only for a small spatial patch. In this way, the network can be trained with data from simple homogeneous phantoms, while also being applicable to inhomogeneous data subsequently.

Spatio-temporal CNNs [24] have demonstrated promising results for imaging elastic properties using optical coherence elastography ([25], [26]). The concept is to apply convolutions jointly over space and time, which enables spatio-temporal feature learning from data [24]. In this way, local spatio-temporal dependencies, which are present for the spatio-temporal windows $\widetilde{x}$, are learned and extracted. As a baseline, we consider the concept of Densely Connected Convolutional Networks (DenseNet) [27] due to its parameter and computational efficiency and develop our own custom DenseNet architecture ([25], [26]). Our architecture details are shown in Fig. 2. Our 3D architecture consists of three initial convolutional layers, followed by three DenseNet blocks with four convolutional layers each. Between the DenseNet blocks, we apply average pooling layers for downsampling of the input dimensions. Moreover, we use batch normalization [28] and the rectified linear activation function for our convolutional layers. Using this 3D CNN architecture, we consider spatio-temporal windows $\widetilde{x}$ with a size of $65 \times 65 \times t$, $33 \times 33 \times t$, $17 \times 17 \times t$, $9 \times 9 \times t$ and $5 \times 5 \times t$ with $t = 35$ frames. We set the spatial stride of our architecture in Fig. 2 to one for spatio-temporal window sizes of $9 \times 9 \times 35$ and $5 \times 5 \times 35$.

B. Conventional Shear Wave Velocity Estimation

To compare our spatio-temporal CNN we consider a ToF approach. To reduce speckle noise, we apply a 3D mean filter with a kernel size of 5px along all axis. Furthermore, we process our data with a directional filter in the frequency domain to reduce waves that propagate along the lateral off-axis and to limit high-frequency imaging noise [9]. For a distinct pixel, we estimate the ToF by performing an auto-correlation between the time varying signals measured at an equivalent distance to the distinct pixel along the lateral axis. Using a distance of 65 pixel, we apply a Tukey window on the two time signals, interpolate them by a factor of 10 and subsequently estimate the time delay by auto-correlation. We assign the estimated shear wave velocity to the pixel located at the center between two measurement points. We reject estimates which are not in the range of $0.1\ \mathrm{m\,s^{-1}}$–$10\,\mathrm{m\,s^{-1}}$. Our data processing is similar to Song *et al.* [9]. Please note that we perform data acquisition with a single push sequence and subsequent high frequency plane wave imaging with a single steering angle of 0°. Following [29], the shear wave velocity c_s is mapped to the Young's modulus with the relation

$$E_{ToF} = \alpha \cdot \rho \cdot 2(1 + \nu) \cdot c_s^2 \qquad (1)$$

with the density $\rho = 1000\,\mathrm{kg\,m^{-3}}$ and the Possion's ratio $\nu = 0.5$. For a fair comparison to our deep learning approach, we introduce a scaling factor α, to account for constant errors between our estimates and our ground truth Young's modulus labels estimated from indentation experiments. We estimate $\alpha = 0.75$ by minimizing the offset between the mean of all Young's moduli estimates from a single gelatin concentration and the corresponding indentation experiment. For inclusion

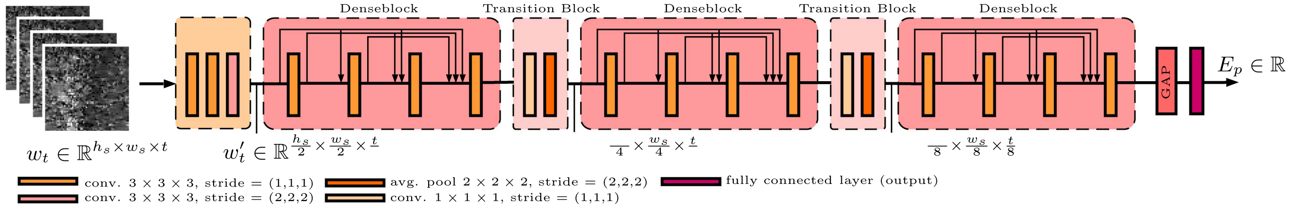
Denseblock
Transition Block
Denseblock
Transition Block
Denseblock
$w_t \in \mathbb{R}^{h_s \times w_s \times t}$
$w'_t \in \mathbb{R}^{\frac{h_s}{2} \times \frac{w_s}{2} \times \frac{t}{2}}$
$\frac{}{4} \times \frac{w_s}{4} \times \frac{t}{}$
$\frac{}{8} \times \frac{w_s}{8} \times \frac{t}{8}$
GAP
$E_p \in \mathbb{R}$
conv. 3 × 3 × 3, stride = (1,1,1)
conv. 3 × 3 × 3, stride = (2,2,2)
avg. pool 2 × 2 × 2, stride = (2,2,2)
conv. 1 × 1 × 1, stride = (1,1,1)
fully connected layer (output)

Fig. 2. Spatio-temporal CNN architecture. We predict the Young's modulus from a 3D spatio-temporal window as input. Our network consists of initial convolutional layers, followed by DenseNet blocks. Between the DenseNet blocks we apply average pooling layers for downsampling of the input dimensions.

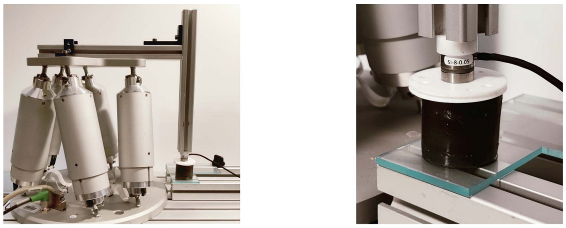

Fig. 3. Experimental setup to estimate the Young's modulus. *Left:* A hexapod robot (PI H-820, Physik Instrumente, Germany) drives a plate into a cylindrical gelatin phantom. *Right:* A high resolution force sensor (Nano17, ATI, USA) is mounted between the robot and the white plate.

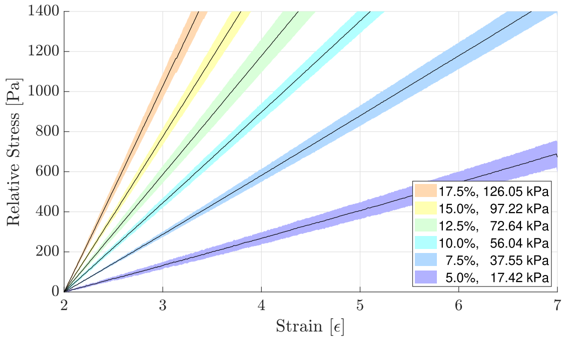

Fig. 4. Stress-strain curve of indentation experiments.

map experiments, we apply a gaussian filter with a kernel size of 2 × 2 mm to account for outliers and close holes with no predicted values due to noise. Also, we increase the FOV and average estimates by combining nine push and imaging sequences distributed evenly across the probe length.

C. Phantom Preparation and Annotation

We use gelatin phantoms as tissue surrogates and prepare batches of gelatin with a weight ratio of gelatin to water ranging from 5% to 17.5% in increments of 2.5%. For precise and reproducible manufacturing, we thoroughly follow this procedure: mix ballistic gelatin (250 Bloom Type A Ordenance Gelatin, Gelita) and water, let the mixture mature for 2 hours, heat the mixture automatically controlled to 50 °C and add 1 g of graphite per 800 g weight for ultrasound speckle. Experiments are performed after approximately 24 hours of cooling. Three types of phantoms are manufactured in-house: (1) for ground truth annotation we prepare eight cylindrical phantoms of each concentration with a radius $\mathrm{r} = 10\,\mathrm{mm}$ and a height $\mathrm{l}_0 = 40\,\mathrm{mm}$ as shown in Fig. 3, right, (2) for training and testing of our network we prepare block phantoms (∼100 × 100 × 100 mm) of each concentration and (3) inclusion phantoms with a gelatin concentration of 7.5% for the background and a gelatin concentration of 15% for embedded circular inclusions with a radius of approximately 5 mm and 10 mm, as well as embedded chicken heart tissue. To avoid gelatine layers, the casted cylindrical inclusions were fixed on both ends to the phantom wall before gelatin was added.

We estimate the ground truth elasticity using the cylindrical phantoms. We perform unconfined compression tests to estimate the Young's modulus as the ratio of stress to strain ([37], [38]). The experimental setup is shown in Fig. 3. During indentation, the sensor records forces F with a frequency of 200 Hz and a resolution of 3.1 mN. We restrict the applied forces to a maximum of 2 N, drive the plate with a constant velocity of $0.01\ \mathrm{mm\,s^{-1}}$ and apply lubricant between contact surfaces to reduce bulking of the phantom. Viscous effects are reduced due to our slow indentation speed. To estimate the elasticity of our gelatin phantoms, we perform a single indentation experiment per phantom, to avoid material defects due to indentation. We perform 8 indentation experiments per gelatin concentration and consider the Young's modulus as

$$E = \frac{\sigma}{\epsilon} = \frac{F}{\pi r^2}\frac{l_0}{\Delta l} \tag{2}$$

with the stress σ and strain ϵ. We estimate E with a linear regression applied to all indentation experiments performed on a gelatin concentration. The strain range is limited between 2% and 7% ([37]–[39]). Results for the stress-strain curves and the corresponding Young's moduli are shown in Fig. 4. For comparison, elasticities of real tissue reported in the literature are presented in Table I.

D. Data Acquisition

Our experimental setup for US-SWEI data acquisition is shown in Fig. 5. For pushing and imaging, we use a linear array probe (128 elements, 0.29 mm pitch, center frequency 7.5 MHz) and a 128-channel ultrasound system (Cicada, Cephasonics Inc, USA). The ultrasound probe is positioned by a serial robot (UR3, Universal Robots, Denmark) for automatic data acquisition. A force sensor (Nano43, ATI, USA) is mounted to the end-effector of the robot to ensure a repeatable contact pressure of 0.2 N between phantom and probe. Ultrasound gel is applied to the surface to reduce imaging artifacts. Once the transducer is positioned, an unfocused push sequence

TABLE I
REPORTED YOUNG'S MODULI OF SOFT TISSUES IN THE LITERATURE, ESTIMATED EITHER BY INDENTATION [I] OR SHEAR WAVE ELASTICITY IMAGING [SWEI]

Tissue	E [kPa]	Method	Reference
Adipose tissue	5.60	[I]	[30]
Thyroid	10.97 ± 3.1	[SWEI]	[31]
Porcine Liver	10-15	[I]	[32]
Muscle	12-13	[SWEI]	[33]
Spleen	14-35	[I]	[2]
Renal Cortex	15 ± 7.2	[I]	[2]
Prostate (Healthy)	17.0 ± 9.0	[I]	[34]
Normal Fat	18-24	[I]	[35]
Renal pelvis	23.60 ± 5.4	[SWEI]	[31]
Prostate (Cancerous)	24.1 ± 14.5	[I]	[34]
High Grade Breast Tissue	43 ± 12	[SWEI]	[36]
Achiles Tendon	51.50 ± 25.1	[SWEI]	[31]
Tendon	70-72	[SWEI]	[33]
Fibrous Breast Tissue	96-244	[I]	[35]

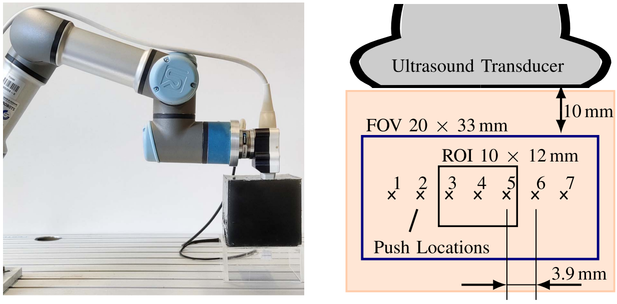

Fig. 5. Our setup for ultrasound shear wave data acquisition. (Left) A linear ultrasound probe is positioned by the robot on the black gelatin block. (Right) The different push locations relative to the ROI.

(120 V, 2000 push cycles, 10 mm depth) excites a shear wave inside the phantom. A continuous segment of 11 elements with the center element defined as the push location was used to transmit the unfocused push. Subsequently, we perform plane wave imaging with an imaging frequency of 7000 Hz and a FOV of 20 × 33 mm along the depth and lateral image axis. We record 35 subsequent images after each push sequence with a resolution of 250 × 600 pixels along the depth and lateral axis, respectively, after beamforming. Image recording is engaged 0.13 ms after the push sequence to reduce imaging artifacts. Loupas's algorithm [40] is applied on the IQ demodulated data to estimate axial displacement relative to a reference frame prior to excitation. We use the resulting displacement data as input to our network. In total, the robot positions the ultrasound transducer randomly at 80 positions on the surface of each gelatin block for data variation. At each position, we perform seven push and imaging sequences with individual pushes applied at the locations depicted in Fig. 5, right. Using a robot allows us to efficiently acquire real ultrasound training data that includes probe and system characteristics.

E. Training and Evaluation

We train our approach with spatio-temporal windows $\widetilde{x}$ spatially located within a defined ROI with a size of 121 × 181 pixel (10 × 12 mm), see Fig. 5. To study the flexibility of our method with respect to push locations, we consider seven different push positions relative to the ROI. We train our networks using homogeneous phantoms with defined ground truth elasticity E_{gt}, determined by indentation experiments. Hence, we assign the corresponding ground truth elasticity E_{gt} to a spatio-temporal window $\widetilde{x}$ and the learning task is to perform a regression of $\widetilde{x}$ to the corresponding elasticity E_{gt}. Therefore, a network is trained to learn the relationship between elasticity and shear-wave propagation for a small local region. For training, we minimize the mean squared error (MSE) loss function between the defined target ground truth elasticity E_{gt} and our predicted elasticity E_p defined as

$$\mathcal{L}(E_{gt}, E_p) = \frac{1}{N}\sum_{k=1}^{N}\left\| E_{gt}^{\{k\}} - E_{p}^{\{k\}} \right\|^2 \tag{3}$$

with N for the number of samples. During one training epoch, we take one spatio-temporal window $\widetilde{x}$ with random location within the ROI from every image sequence x in our training data set. Each network is trained for 250 epochs with a batch size of 250 using Adam for optimization with a learning rate of $l_r = 1e^{-4}$. After 150 epochs, we divide the learning rate by a factor of two every 50 epochs. We normalize the pixel intensities of each input $\widetilde{x}$ to have a zero mean and standard deviation of one. To augment our training data, we randomly apply horizontal and vertical flipping, multiple 90° rotations, Gaussian blur and randomized input erasing of the input data during training.

We evaluate the performance of our method with all elasticities present during training and perform four-fold cross-validation on the 80 different positions of each concentration. For each fold we use 60 positions of each concentration for training, and 10 positions each for validation and testing. Second, we evaluate the regression performance of our method on unseen elasticities and perform a cross-validation approach, where we leave out the entire data of one gelatin concentration. We do not perform cross-validation on boundary elasticities, e.g., 5% and 17.5% as this leads to out of distribution predictions for the regression task. Hence, we perform four-fold cross-validation using the gelatin concentrations starting from 7.5% up to 15%. In each fold, we randomly split the data into 50% of the fold's data for testing and 50% for validation. Moreover, for all our trainings we remove push one and seven completely from our training data, to evaluate unseen push locations further away from the ROI. For elasticity estimation on inclusion phantoms, we refine the network trained on homogenoues phantoms, by fine-tuning the network for additional 10 epochs with inhomogeneous phantom data. Thereby, the network learns wave reflections at boundaries which are not present in homogenous phantoms.

III. Results

A. Homogeneous Phantoms

We study our spatio-temporal CNN approach qualitatively to ToF in Fig. 6 and evaluate the prediction maps of both approaches with respect to varying push locations and phantom elasticities. For this and the following evaluations, if not indicated otherwise, we consider the more challenging case for

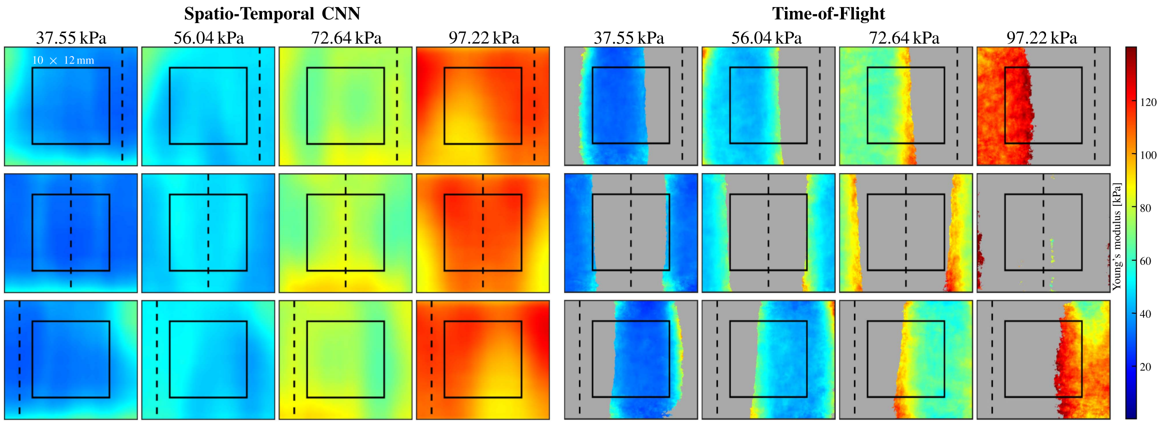

Fig. 6. Prediction of gelatin elasticity. Predictions of the Young's modulus with a spatio-temporal CNN (left) and with the conventional ToF method (right) for different push locations and gelatin concentrations. The push location is indicated by the black dashed line for push 2 (top row), 4 (middle row) and 6 (bottom row). For each pixel we show the mean Young's modulus from 40 individual push and imaging sequences. Failed estimates, i.e, estimates which are not in the range of 0.024 kPa–237 kPa are indicated in gray; the black square indicates the ROI. We use a spatio-temporal window size of $65 \times 65 \times 35$.

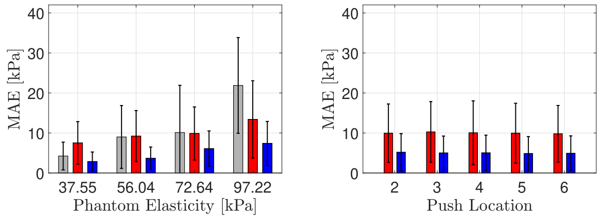

Fig. 7. Elasticity estimation performance. Left: Concentration / phantom stiffness vs. MAE considering the results of all push locations. Right: Push location vs. MAE across all elasticities. (Red) Spatio-temporal CNN where evaluated elasticity is left out during training; (Blue) Spatio-temporal CNN where evaluated elasticity is also present during training; (Grey) Time-of-Flight method, note that we exclude failed predictions.

TABLE II
MAE and Pearson Correlation Coefficient (pCC) for Different Window Sizes. Throughput Refers to the Number of Positions for Which Elasticity Can Be Estimated Within One Second. We Measure the Throughput of Our Methods on a NVIDIA Tesla V100-SXM2-32GB Using a Batch Size of 500

$h_s \times w_s$	MAE [kPa]	pCC [%]	Throughput [px/s]
5×5	14.40±10.82	72.81	15745
9×9	12.63±9.01	86.53	5651
17×17	11.83±8.25	87.54	2348
33×33	10.64±7.48	88.55	1688
65×65	9.99±7.49	93.39	508

our deep learning approach where we left out entire elasticities during training. Our findings in Fig. 6 demonstrate that our spatio-temporal CNN approach leads to more consistent estimations for all experiments. Notably, our results show that our spatio-temporal CNN provides estimations inside the push location in contrast to ToF that fails in general closer to the push region. Moreover, our results in Fig. 7 confirm quantitatively that the performance of the spatio-temporal CNN is independent of the push location considering all phantom elasticities. Further, we study the performance of spatio-temporal CNNs and the ToF approach and evaluate the performance quantitatively with respect to phantom elasticity, see Fig. 7. Our results show that the pixelwise mean absolute error (MAE) increases with increasing elasticity for both ToF and spatio-temporal CNN. Our spatio-temporal CNN approach leads to an overall MAE of 5.01(437) kPa when all elasticities are present during training, compared to an overall MAE of 9.99(749) kPa where evaluated elasticities are left out during training. The ToF approach leads to an overall MAE of 11.61(876) kPa. Second, performance with respect to the spatio-temporal window size is given in Table II. Our results demonstrate that larger spatial input sizes work better at the expense of reduced model throughput, e.g., using the largest spatial input sizes of 65 × 65 pixels (∼4 × 5 mm) improves performance by 30% while reducing the throughput by a factor of 31 compared to the smallest input size of 5 × 5 pixels (∼0.32 × 0.4 mm).

Third, we further study the robustness of our methods and show the standard deviation of predictions at each pixel for the complete FOV for the spatio-temporal CNN and the ToF approach using push one, four and seven, see Fig. 8. Note that push one and seven are completely removed from the training data and that we only consider the ROI during training. Our results demonstrate that the spatio-temporal CNN provides consistent estimates with a low standard deviation also at the previously unseen push locations and at a larger FOV than the ROI. Moreover, Fig. 8 demonstrates that for lower phantom elasticity (37.55 kPa) the predictions of the spatio-temporal CNN show a high standard deviation far away from the push location, similar to ToF.

B. Inclusion Phantoms

Results for estimates using our spatio-temporal CNNs on phantoms with embedded cylindrical inclusions are shown in Fig. 9 for spatio-temporal window sizes of 17 × 17, 33 × 33 and 65 × 65 pixels. Depicted is the mean of nine push and imaging sequences. We report the MAE for the phantoms backgrounds and inclusions separately by calculating the pixel-wise errors between the prediction of the network and the corresponding ground truth elasticity. The results for a spatio-temporal window

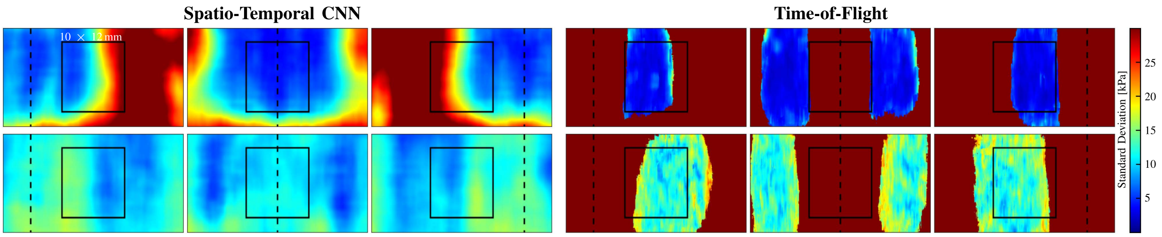

Fig. 8. Full image elasticity predictions. Shown are the standard deviations of the pixelwise estimated Young's moduli of 40 push and image sequences. Results for ground truth elasticity 37.55 kPa and 72.64kPa are shown in the top and bottom row, respectively. The black square indicates the ROI used for training, the push location is indicated by the black dashed line. We use a spatio-temporal window size of $65 \times 65 \times 35$.

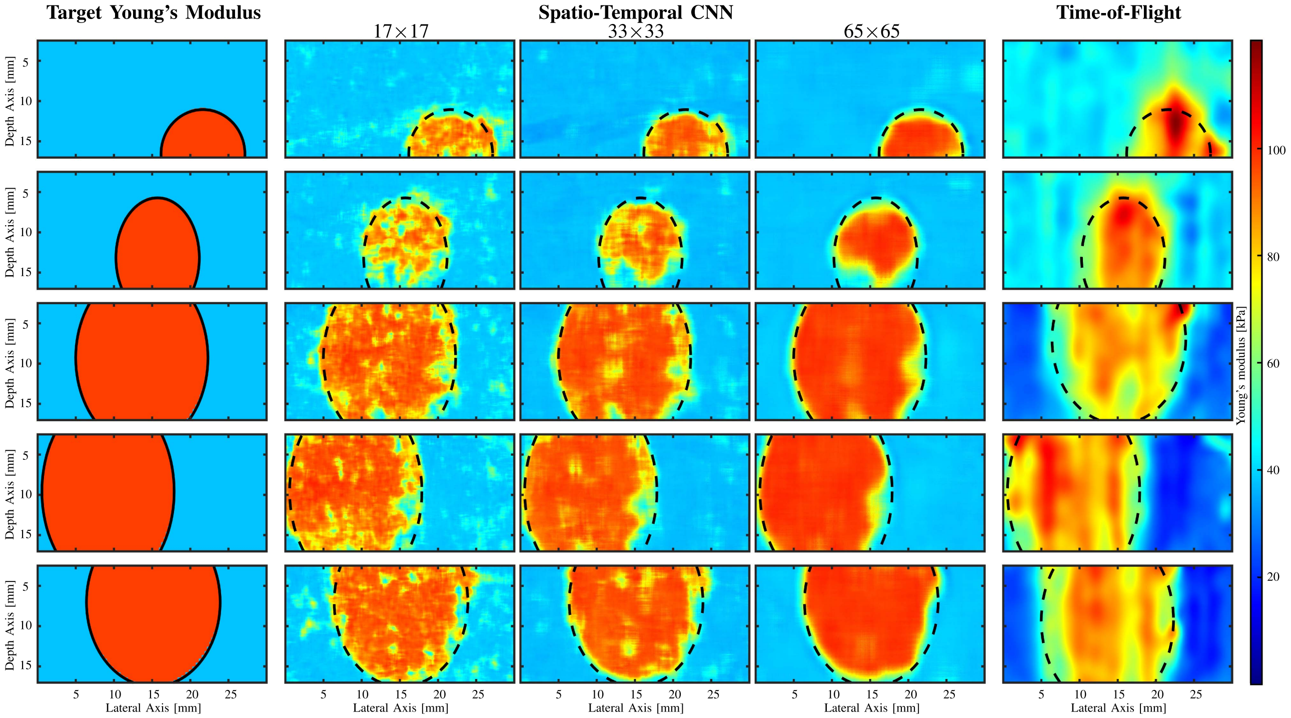

Fig. 9. Elasticity maps of five inclusion shapes. Column 1 shows the target Young's modulus, column 2-4 spatio-temporal CNN predictions with spatio-temporal window sizes of 17 × 17 pixels, 33 × 33 pixels and 65 × 65 pixels. Time-of-flight estimates are depicted in column 5.

size of 65 × 65 pixels are given in Table III. The combined MAE across all phantoms with deep learning is 7.5(1287) kPa for all inclusions and 1.64(0432) kPa for the background. The combined MAE with ToF is 16.28(1005) kPa for all inclusions and 11.11(1008) kPa for the background. The threshold for the Dice coefficient is set to 67.38 kPa and is estimated by the mean target Young's modulus of inclusion and background. The mean Dice coefficient for the inclusion shapes depicted in Fig. 9 is 0.93 for our deep learning approach and 0.86 for ToF. Table IV shows that the Dice coefficient and the MAE decrease for smaller spatio-temporal windows sizes. This is consistent for all binarization thresholds as shown in Fig. 10. The elasticity map of chicken heart tissue, B-Mode ultrasoundx image and cross-section of the phantom is given in Fig. 11.

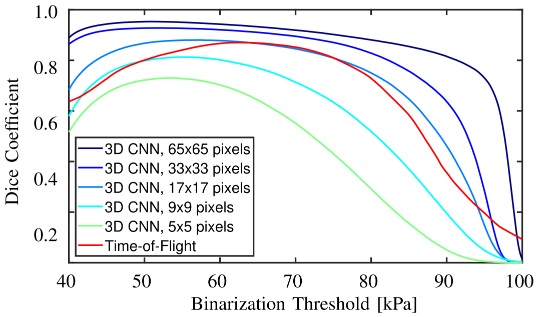

Fig. 10. Mean Dice coefficients vs. binarization threshold for all inclusion shapes. Given are all analyzed spatio-temporal window sizes and time-of-flight.

IV. Discussion

We present a deep learning approach for local elasticity estimation from real 3D ultrasound data. This task has been addressed with conventional methods by extracting the shear wave velocity as an explicit feature ([10], [41]). In contrast, deep learning methods allow estimates without explicit feature extraction, intensive pre-processing and manual tuning. We

TABLE III
MAE for Inclusion (in) and Background (bg) and the Dice Coefficient With the Corresponding Mean μ for All Five Inclusion Shapes. Metrics for the Spatio-Temporal CNN Are Derived With Spatio-Temporal Windows With a Size of 65 × 65 Pixels. The Inclusion Shapes Are Displayed in Fig. 9 With, e.g., Row One Referring to Phantom #1

Method	#	MAE$_{in}$ [kPa]	MAE$_{bg}$ [kPa]	Dice
3D CNN	1	10.16±13.13	0.97±2.30	0.93
	2	19.52±19.06	1.06±3.25	0.81
	3	4.70±8.32	2.42±5.01	0.98
	4	5.50±10.98	2.13±5.25	0.96
	5	6.11±11.51	2.27±5.84	0.95
	μ	**7.50±12.87**	**1.64±4.32**	**0.93**
ToF	1	15.39±10.14	9.19±10.09	0.79
	2	13.28±10.38	8.37±0.89	0.85
	3	17.14±9.66	14.49±9.24	0.89
	4	13.87±9.05	13.60±9.77	0.93
	5	19.91±10.15	11.99±10.89	0.86
	μ	**16.28±10.05**	**11.11±10.08**	**0.86**

TABLE IV
Mean MAE and Dice Score for All Inclusion Shapes. Given Is the Inclusion (in) and Background (bg) With All Studied Spatio-Temporal Window Sizes. The Binarization Threshold Is Set to 67.38kPa

$h_s \times w_s$	MAE$_{in}$ [kPa]	MAE$_{bg}$ [kPa]	Dice
5 × 5	29.68±13.84	6.37±7.47	0.60
9 × 9	22.74±14.43	5.30±7.54	0.75
17 × 17	14.29±13.59	4.21±7.81	0.86
33 × 33	11.45±13.55	1.91±5.05	0.90
65 × 65	7.50±12.87	1.64±4.32	0.93

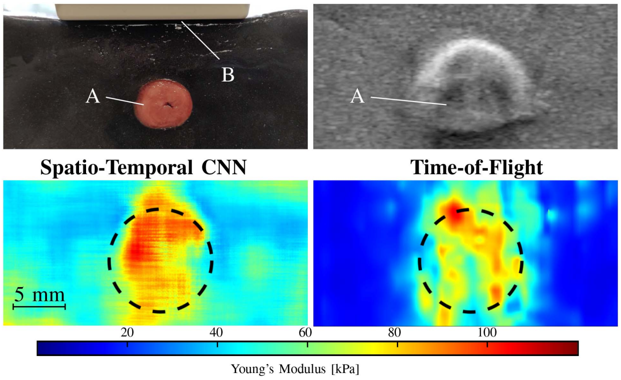

Fig. 11. Elasticity estimation in soft tissue. Top Left: Phantom cross-section with chicken heart tissue (A) embedded in gelatin and positioned ultrasound transducer (B). Top Right: B-Mode image of inclusion. Bottom: Elasticity maps estimated with our spatio-temporal CNN and time-of-flight approach.

present a local elasticity estimation approach with real ultrasound data, where a 3D spatio-temporal CNN is trained to predict tissue elasticity from spatio-temporal windows.

Elasticity estimation: We study the performance of our methods on homogeneous phantoms considering various push locations and elasticities. We demonstrate that predictions can be performed for an elasticity range of 38 kPa–98 kPa which reflects reported tissue elasticities in the literature as shown in Table I. Our findings highlight that elasticity estimation on stiffer tissue with spatio-temporal CNNs is also consistent, while in the literature typically estimated gelatin elasticities in the range of up to 10% are reported ([8], [11], [15]). Naturally, faster shear waves reduce the amount of shear wave information which leads to an increased error for conventional methods. In contrast, stiffer elasticities only lead to slightly reduced performance for our deep learning approach, see Fig. 7. Hence, our findings show that our spatio-temporal CNN approach leads to robust and consistent results across a wide range of elasticities. In general, the performance could be further enhanced by the use of image compounding during data acquisition. The state-of-art is using three angled plane waves for image acquisition which increases the SNR for more robust estimates with ToF [9]. However, in this case the image acquisition frequency decreases by a factor of three, which results in fewer images containing shear wave information which makes it difficult to estimate shear wave velocity for stiff elasticities. Note that we already observe that performance decreases for stiffer elasticities. Hence, we consider the maximum available imaging frame rate without image compounding.

In this work, we estimate the Young's modulus as a surrogate for tissue elasticity with ground truth annotation performed by indentation experiments. In general, our approach only requires selected material parameters as training targets and subsequently we are able to generalize to unseen data. It is noticeable that predicting known elasticities improves the performance of our spatio-temporal CNN approach. This could be considered as a relevant scenario as further elasticities can be included in the training data. Also, we demonstrate that even in the challenging case of predicting elasticities that are not present during training our spatio-temporal CNN approach leads to competitive performance compared to our ToF method. This demonstrates that our spatio-temporal CNN approach generalizes well between different elasticities even with few ground truth elasticities. Fig. 7 suggests a better MAE for ToF for elasticities under 72 kPa, when evaluated elasticities are left out during training. However, ToF-results do not include failed estimates, i.e., we exclude all outlier ToF estimations which are not in the range of 0.1–10 m/s. This leads to several missing ToF estimates as shown in Fig. 6 in gray. Notably, previous work on estimating the elasticity from SWEI data build on simulated data ([22], [23]). However, this only solves the inverse problem of the model underlying the simulation. In addition, performance is limited on real data as, e.g., noise and image artifacts are not sufficient represented in simulated data [23]. In contrast, our approach is trained with real data, which includes, e.g., imaging noise or probe artifacts. Also, our local approach can be adapted to real tissue samples, e.g., obtained from tumor resections, by using pathological tumor properties as training targets. Subsequently, pathological properties of soft tissue can be predicted and imaged with our local estimation approach.

Push dependency: Second, we study the robustness of our approach concerning the spatio-temporal window position relative to the push location. Our results, depicted in Fig. 6 and Fig. 7, show qualitatively and quantitatively that our deep learning approach provides more consistent estimates than ToF and provides accurate estimations independent of the push location relative to the ROI. In the case that all elasticities are present during training our deep learning approach outperforms ToF.

We studied similar result in soft tissue phantoms with consistent predictions for the individual push locations. This is in contrast to other work in the field that require, e.g., two fine-tuned pushes to the left and right of the ROI [15]. Furthermore, it stands out that with a spatio-temporal CNN we can perform predictions within the push location, where shear wave propagation is complex and diffuse and the imaging is dominated by relaxation dynamics. Predictions can even be performed inside the push region for small spatial window sizes of 5 × 5 pixels. It can be assumed that the network learns the relaxation dynamics of the gelatin which changes for different elasticities. This has not yet been shown in any previous work that uses deep learning methods in combination with simulated data ([22], [23]) or conventional methods [9]. Still, further investigation on the push depth as well as the material viscosity are necessary. We also study our approach on the complete FOV, while only training with image crops from the ROI. In this way, we evaluate our approach for completely unseen locations relative to the push location and are able to study the performance far away from the push location. For push locations (push one and seven) not included in the training data, we can still perform accurate estimates, see Fig. 8. This demonstrates that our approach also leads to robust results for unknown push locations. Moreover, our results in Fig. 8 further confirm that our spatio-temporal CNN approach outperforms ToF and provides accurate estimates for a much larger FOV than the ROI. Similar to ToF, it stands out that our spatio-temporal CNN approach does not provide consistent estimates far away from the push location. This shows that our approach does not over-fit on specific phantom features such as speckle characteristics and fails when no wave information is present in the data.

Inference and Performance: Third, we study the performance of our network concerning inference time and spatio-temporal window sizes, see Table II. Increasing the spatial window size leads to more accurate results compared to using smaller spatial window sizes. This is most likely related to the fact that larger window sizes cover a larger spatial area, hence providing more information about wave propagation. However, using smaller window sizes allows for notably increased model throughput, which is important to provide real-time estimates for larger FOVs and higher resolution. Considering our results in Table II, using a smaller spatial window size, e.g., 33 × 33 pixels might be a good starting point for further work as there is similar performance compared to 65 × 65 pixels, while the throughput is increased by a factor of 3.32. In general, pixelwise processing is more computationally expensive than an encoder-decoder architecture applied to the entire image at once. However, CNNs are inherently efficient for this task, because computations can be shared across overlapping regions during testing [42]. Similar, a whole image fully convolutional training [43] could be used to further speed up the training time. We perform patchwise training, which results in higher batch variance and allows to use different augmentation on image crops from the same ROI. Also, a direct advantage of our approach is that sparse estimates can be performed during inference, e.g., only predicting every nth pixel. This allows to scale our approach effectively to larger FOVs while maintaining similar inference times. Overall, our results demonstrate that global elasticity maps can be estimated in real-time using our deep learning approach. In particular, the use of more powerful hardware will improve the inference time of our method. Although a comparison due to different hardware is difficult, our spatio-temporal CNN approach is more time efficient than conventional methods and can perform predictions on a smaller window size, e.g., Kijanka *et al.* [16] report an inference time of 0.22 ms per estimate for a spatial window size of 4.5 × 4.5 mm while our spatio-temporal CNN achieves a inference time of 0.07 ms for a spatial window size of 0.32 × 0.4 mm.

Inclusion Shapes: Finally, we evaluate our methods on gelatin phantoms with circular stiff inclusions. Our results in Fig. 9 demonstrate that our spatio-temporal CNN approach provides consistent estimates with larger spatio-temporal window sizes for the inclusion and the background similar to our results on homogeneous phantoms. Considering the MAE, performance of estimates inside the inclusion increases by a factor of ∼ 2 and on the background by a factor of ∼ 6 with our spatio-temporal CNN in comparison to ToF. While we perform local estimations, this raises the question how our approach performs on elasticity boundaries with respect to the spatio-temporal window size. Our results in Fig. 9 demonstrate that errors can be seen at elasticity boundaries and the shape of the inclusion is still well defined. While smaller window sizes consider a smaller spatial area, this could lead to more distinct boundaries and less blurring. In general, our results show that larger window sizes lead to more consistent estimates, as seen for background predictions in Fig. 9. However, we find that the general performance drop for smaller window sizes outweighs the potential benefit of reduced blurring. Hence, for larger window sizes the boundary is more distinct visible and the Dice score is higher. Nevertheless, it is noticeable that inclusion boundaries can also be retrieved from small windows sizes, e.g., ∼1 × 1 mm (17 × 17 pixels). In direct comparison with ToF, the Dice coefficient is similar for a window size of 17 × 17 pixels (Fig. 10). Hence, spatio-temporal window sizes smaller than 17 ×17 pixel (∼1 × 1 mm) are not favorable. Our results in Fig. 11 demonstrate that elasticity estimation in chicken heart tissue is also feasible with our deep learning approach. The investigation of other soft tissue samples in an interesting next step for future work. Overall, our spatio-temporal CNN approach shows promising results in the estimation of elasticity in inhomogeneous mediums.

V. Conclusion

We present 3D spatio-temporal CNNs for local elasticity estimation from real ultrasound shear wave data, which demonstrate increased performance compared to conventional approaches. Our findings show that spatio-temporal CNNs can retrieve local elastic properties from small spatio-temporal windows while being independent of the push location, and demonstrating consistent performance across various elasticities and inhomogenities. Further work will include in vitro and in vivo experiments of real soft tissues.

Acknowledgment

The authors declare that they have no conflict of interest.

References

[1] D. Cosgrove *et al.*, "EFSUMB guidelines and recommendations on the clinical use of ultrasound elastography. Part II: Clinical applications," *Eur. J. Ultrasound*, vol. 34, no. 3, pp. 238–253, 2013.

[2] S. Umale *et al.*, "Experimental mechanical characterization of abdominal organs: Liver, kidney & spleen," *J. Mech. Behav. Biomed. Mater.*, vol. 17, pp. 22–33, 2013.

[3] Y.-P. Yang *et al.*, "Qualitative and quantitative analysis with a novel shear wave speed imaging for differential diagnosis of breast lesions," *Sci. Rep.*, vol. 7, no. 1, pp. 1–11, 2017.

[4] F. Sebag *et al.*, "Shear wave elastography: A new ultrasound imaging mode for the differential diagnosis of benign and malignant thyroid nodules," *J. Clin. Endocrinol. Metab.*, vol. 95, no. 12, pp. 5281–5288, 2010.

[5] J. A. Sande *et al.*, "Ultrasound shear wave elastography and liver fibrosis: A prospective multicenter study," *World J. Hepatol.*, vol. 9, no. 1, pp. 38–47, 2017.

[6] J. Bercoff, M. Tanter, and M. Fink, "Supersonic shear imaging: A new technique for soft tissue elasticity mapping," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 51, no. 4, pp. 396–409, Apr. 2004.

[7] P. Song *et al.*, "Fast shear compounding using robust 2-D shear wave speed calculation and multi-directional filtering," *Ultrasound Med. Biol.*, vol. 40, no. 6, pp. 1343–1355, 2014.

[8] S. Latus *et al.*, "An approach for needle based optical coherence elastography measurements," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, 2017, pp. 655–663.

[9] P. Song *et al.*, "Comb-push ultrasound shear elastography (CUSE): A novel method for two-dimensional shear elasticity imaging of soft tissues," *IEEE Trans. Med. Imag.*, vol. 31, no. 9, pp. 1821–1832, Sep. 2012.

[10] C. A. Carrascal *et al.*, "Improved shear wave group velocity estimation method based on spatiotemporal peak and thresholding motion search," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 64, no. 4, pp. 660–668, Apr. 2017.

[11] A. J. Engel and G. R. Bashford, "A new method for shear wave speed estimation in shear wave elastography," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 62, no. 12, pp. 2106–2114, Dec. 2015.

[12] M. Tanter *et al.*, "Quantitative assessment of breast lesion viscoelasticity: Initial clinical results using supersonic shear imaging," *Ultrasound Med. Biol.*, vol. 34, no. 9, pp. 1373–1386, 2008.

[13] J. Yang *et al.*, "Comparative study on shear wave speed estimation algorithms in ARFI for improving its reliability," in *Proc. 36th Annu. Int. Conf. Proc. IEEE Eng. Med. Biol. Soc.*, 2014, pp. 226–229.

[14] M. Wang *et al.*, "On the precision of time-of-flight shear wave speed estimation in homogeneous soft solids: Initial results using a matrix array transducer," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 60, no. 4, pp. 758–770, Apr. 2013.

[15] P. Kijanka and M. W. Urban, "Local phase velocity based imaging: A new technique used for ultrasound shear wave elastography," *IEEE Trans. Med. Imag.*, vol. 38, no. 4, pp. 894–908, Apr. 2018.

[16] P. Kijanka and M. Urban, "Fast local phase velocity-based imaging: Shear wave particle velocity and displacement motion study," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 67, no. 3, pp. 526–537, Mar. 2020.

[17] M. Kibria *et al.*, "Gluenet: Ultrasound elastography using convolutional neural network," in *Simul., Image Process., Ultrasound Syst. Assist. Diagnosis Navigation*. Springer, 2018, pp. 21–28.

[18] D. Y. Chan *et al.*, "Deep convolutional neural networks for displacement estimation in ARFI imaging," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 68, no. 7, pp. 2472–2481, Jul. 2021.

[19] A. K. Z. Tehrani *et al.*, "Semi-supervised training of optical flow convolutional neural networks in ultrasound elastography," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, 2020, pp. 504–513.

[20] R. Delaunay *et al.*, "An unsupervised learning approach to ultrasound strain elastography with spatio-temporal consistency," *Phys. Med. Biol.*, vol. 66, no. 17, 2021, Art. no. 175031.

[21] F. Q. Jin *et al.*, "Deep learning based quantitative uncertainty estimation for ultrasound shear wave elasticity imaging," in *Proc. IEEE Int. Ultrason. Symp.*, 2021, pp. 1–4.

[22] L. Vasconcelos *et al.*, "Viscoelastic parameter estimation using simulated shear wave motion and convolutional neural networks," *Comput. Biol. Med.*, vol. 133, 2021, Art. no. 104382.

[23] S. Ahmed *et al.*, "Dswe-net: A deep learning approach for shear wave elastography and lesion segmentation using single push acoustic radiation force," *Ultrasonics*, vol. 110, 2021, Art. no. 106283.

[24] D. Tran *et al.*, "Learning spatiotemporal features with 3D convolutional networks," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2015, pp. 4489–4497.

[25] M. Neidhardt *et al.*, "Deep learning for high speed optical coherence elastography," in *Proc. IEEE 17th Int. Symp. Biomed. Imag.*, 2020, pp. 1583–1586.

[26] M. Neidhardt *et al.*, "4D deep learning for real-time volumetric optical coherence elastography," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 16, no. 1, pp. 23–27, 2021.

[27] G. Huang *et al.*, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4700–4708.

[28] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. learn.*, Jun. 2015, pp. 448–456.

[29] A. P. Sarvazyan *et al.*, "Acoustic waves in medical imaging and diagnostics," *Ultrasound Med. Biol.*, vol. 39, no. 7, pp. 1133–1146, 2013.

[30] A. Samani *et al.*, "Measuring the elastic modulus of ex vivo small tissue samples," *Phys. Med. Biol.*, vol. 48, no. 14, pp. 2183–2199, 2003.

[31] K. Arda *et al.*, "Quantitative assessment of normal soft-tissue elasticity using shear-wave ultrasound elastography," *AJR Amer. J. Roentgenol.*, vol. 197, no. 3, pp. 532–536, 2011.

[32] E. Samur *et al.*, "A robotic indenter for minimally invasive measurement and characterization of soft tissue response," *Med. Image Anal.*, vol. 11, no. 4, pp. 361–373, 2007.

[33] B. C. W. Kot *et al.*, "Elastic modulus of muscle and tendon with shear wave ultrasound elastography: Variations with different technical settings," *PLoS One*, vol. 7, no. 8, 2012, Art. no. e44348.

[34] B.-M. Ahn *et al.*, "Mechanical property characterization of prostate cancer using a minimally motorized indenter in an ex vivo indentation experiment," *Urology*, vol. 76, no. 4, pp. 1007–1011, 2010.

[35] T. A. Krouskop *et al.*, "Elastic moduli of breast and prostate tissues under compression," *Ultrason. Imag.*, vol. 20, no. 4, pp. 260–274,1998.

[36] A. Samani *et al.*, "Elastic moduli of normal and pathological human breast tissues: An inversion-technique-based investigation of 169 samples," *Phys. Med. Biol.*, vol. 52, no. 6, pp. 1565–1577, 2007.

[37] A. Forte *et al.*, "Modelling and experimental characterisation of the rate dependent fracture properties of gelatine gels," *Food Hydrocoll.*, vol. 46, pp. 180–190, 2015.

[38] R. Delaine-Smith *et al.*, "Experimental validation of a flat punch indentation methodology calibrated against unconfined compression tests for determination of soft tissue biomechanics," *J. Mech. Behav. Biomed. Mater.*, vol. 60, pp. 401–415, 2016.

[39] M. Żmudzińska *et al.*, "The assessment of the applicability of shear wave elastography in modelling of the mechanical parameters of the liver," *Acta Bioeng. Biomech.*, vol. 20, no. 4, pp. 59–64, 2018.

[40] T. Loupas, J. T. Powers, and R. W. Gill, "An axial velocity estimator for ultrasound blood flow imaging, based on a full evaluation of the doppler equation by means of a telastic moduli of normal and pathological human breast tissues: An inversion-technique-based investigation of 169 sampleswo-dimensional autocorrelation approach," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 42, no. 4, pp. 672–688, Jul. 1995.

[41] C. A. Trutna *et al.*, "Measurement of viscoelastic material model parameters using fractional derivative group shear wave speeds in simulation and phantom data," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 67, no. 2, pp. 286–295, Feb. 2020.

[42] P. Sermanet *et al.*, "Overfeat: Integrated recognition, localization and detection using convolutional networks," 2013, *arXiv:1312.6229*.

[43] J. Long *et al.*, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3431–3440.