An Unsupervised Approach to Ultrasound Elastography with End-to-end Strain Regularisation

Rémi Delaunay[1,2][0000−0002−0398−4995], Yipeng Hu[1][0000−0003−4902−0486], and Tom Vercauteren[1,2][0000−0003−1794−0456]

[1] Wellcome/EPSRC Centre for Interventional and Surgical Sciences, University College London, Gower Street, London WC1E 6BT, UK
[2] School of Biomedical Engineering & Imaging Sciences, King's College London, Strand, London WC2R 2LS, UK
`remi.delaunay.17@ucl.ac.uk`

Abstract. Quasi-static ultrasound elastography (USE) is an imaging modality that consists of determining a measure of deformation (i.e. strain) of soft tissue in response to an applied mechanical force. The strain is generally determined by estimating the displacement between successive ultrasound frames acquired before and after applying manual compression. The computational efficiency and accuracy of the displacement prediction, also known as time-delay estimation, are key challenges for real-time USE applications. In this paper, we present a novel deep-learning method for efficient time-delay estimation between ultrasound radio-frequency (RF) data. The proposed method consists of a convolutional neural network (CNN) that predicts a displacement field between a pair of pre- and post-compression ultrasound RF frames. The network is trained in an unsupervised way, by optimizing a similarity metric between the reference and compressed image. We also introduce a new regularization term that preserves displacement continuity by directly optimizing the strain smoothness. We validated the performance of our method by using both ultrasound simulation and *in vivo* data on healthy volunteers. We also compared the performance of our method with a state-of-the-art method called OVERWIND [17]. Average contrast-to-noise ratio (CNR) and signal-to-noise ratio (SNR) of our method in 30 simulation and 3 *in vivo* image pairs are 7.70 and 6.95, 7 and 0.31, respectively. Our results suggest that our approach can effectively predict accurate strain images. The unsupervised aspect of our approach represents a great potential for the use of deep learning application for the analysis of clinical ultrasound data.

Keywords: Ultrasound elastography · Time delay estimation · convolutional neural network

1 Introduction

Ultrasound elastography (USE) is an imaging technique that enables the characterization of tissue mechanical properties [21]. Since its introduction in 1991 [19],

strain imaging has shown usefulness in diagnostic applications where pathological alterations induce modification of tissue stiffness, such as lesion detection in liver disease [12] and tumour characterisation in the thyroid [16], breast [7] and prostate cancer [3]. This work focuses on quasi-static, free-hand strain elastography, where a time-varying axial compression is applied with an ultrasound transducer to the targeted tissue [23]. The tissue mechanical behavior is then determined by mapping the relative deformation (i.e. strain) induced by manual compression (i.e. stress).

Various methods have been proposed over the years to measure the strain. The main approach consists of determining the spatial displacement between a pair of radio-frequency (RF) ultrasound image data, acquired before and after applying an axial compression. The displacement estimation, also known as time-delay estimation, has been historically performed by maximizing a correlation function between local frame windows, either in the time or phase domain [18, 20, 2, 22, 1]. More recently, different approaches have added a regularization parameter to account for displacement discontinuity and improve displacement estimates [15, 6]. Although these methods demonstrated the ability to make an accurate displacement estimation, current techniques are sensitive to noise and global decorrelation, i.e. the change of speckle appearance due to out-of-plane motion. Furthermore, real-time imaging is an important feature of elastography, and a trade-off is often made between the precision of standard approaches and their computational cost.

The recent progress of learning-based methods in computer vision for optical flow estimation have inspired new approaches for strain elastography [9]. Those methods demonstrated the ability of neural networks to exploit the ultrasound high-frequency content and to robustly produce accurate displacement estimates [14, 25]. Previous networks have been trained using ultrasound simulation associated with ground truth displacement and strain labels. Accurate ground-truth images can be difficult to obtain for quasi-static elastography because the magnitude of applied stress is unknown. However, real-world ultrasound data often exhibits complex speckle patterns and echogenic features that can be quite challenging to replicate in ultrasound simulation.

In this paper, we proposed an unsupervised method for time-delay estimation that allows a neural network to be trained directly on clinical data and predict tissue displacement. Unlike previous methods, our CNN training procedure is performed without ground truth labels. Instead, the network weights are optimized by minimizing a dissimilarity function between the pre-compression image and warped compressed image. We also introduce a new regularization term that preserves displacement continuity by directly optimizing the smoothness of the strain prediction. We validated our method by applying it on both real ultrasound data and simulations. The performance of our method was evaluated by comparing our displacement field and strain prediction with the ground truth labels and a gold standard USE technique [17]. To the best of our knowledge, this is the first unsupervised deep-learning method applied to strain imaging.

2   Methods

2.1   Problem Statement

Our method follows an approach similar to a standard image registration framework, which aims to find a spatial transformation that maps a moving image into the space of a reference image. In the case of a non-rigid image registration solution, this transformation can be represented as a dense displacement field (DDF). This transformation is generally optimized in an iterative manner, by maximizing an objective function that measures the similarity between the warped moving image and the reference image, and a regularization parameter to ensure displacement continuity. From a learning-based approach perspective, the image mapping is predicted by a neural network instead of being directly optimized. An overview of the method is presented in Fig. 1.

2.2   Displacement Estimation

Our network takes as input a pair of pre- and post-compression 2D RF frames, here named Pre and $Post$, and predicts a DDF u. The network parameters are learned by minimizing a dissimilarity metric L_{sim} between pairs of 2D RF ultrasound data. Our training loss function also includes a regularization term, L_{reg}, acting on the predicted displacement field u and associated with a weighting hyper-parameter α, to ensure balance between the likelihood and smoothness of the predicted transformation. The optimization problem can be written as:

$$\hat{\theta} = \arg \min_{\theta} \ [L_{sim}(Pre, Post; u) \ + \ \alpha \ . \ L_{reg}(u)] \tag{1}$$

where θ represents the network parameters that are optimized through stochastic gradient descent.
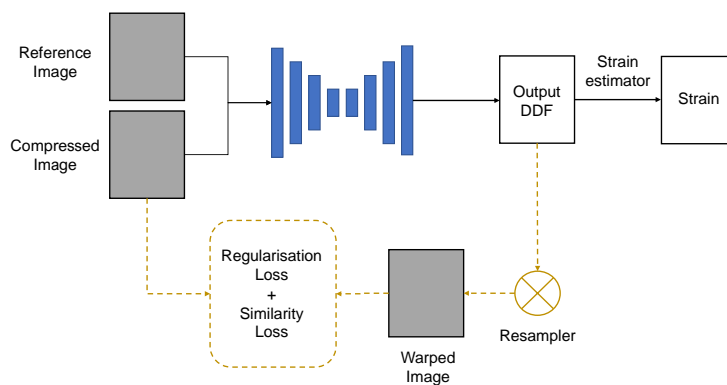

Fig. 1. Overview of the method.

The dissimilarity metric corresponds to a negative local normalized cross-correlation (NCC) which average the NCC score between sliding windows sampled from the pre-compression image and the transformed post-compression image, resampled with the predicted displacement field. The NCC between two local image windows, W_{pre} and W_{post}, with i,j pixel components can be written as:

$$NCC = \frac{1}{N} \sum_{i,j} \frac{\left[W_{pre}(i,j) - \mu_{W_{pre}}\right] \times \left[W_{post}(i,j) - \mu_{W_{post}}\right]}{\sigma_{W_{pre}} \times \sigma_{W_{post}}} \qquad (2)$$

where N is the number of pixels (i,j) and μ and σ correspond to the mean and standard deviation of the images, respectively.

The regularisation term consists of the L1-norm of the second spatial derivatives of the predicted displacement field u. Given that the strain modulus corresponds to the displacement gradient, minimizing its second derivative allows to enforce the strain map smoothness. The regularisation term can be written as:

$$L_{reg} = \sum_{i,j} \left(\mid \partial_x^2 u_{i,j} \mid + \mid \partial_y^2 u_{i,j} \mid \right) \qquad (3)$$

where ∂_x^2 and ∂_y^2 are the second partial derivatives in axial and lateral directions, respectively.

2.3 Implementation

Network The architecture of our network was presented in a method for medical image registration [8], and consists in an encoder-decoder CNN with skip connections. The encoder part is composed of four down-sampling blocks, which capture the hierarchical features necessary to establish correspondence between the image pair. Each down-sampling block consists of two convolutional layers with a residual network shortcut, batch normalization, and leaky rectified linear unit (Leaky ReLU). Symmetrically, the decoder part is composed of four up-sampling blocks that consists of an additive up-sampling layer summed over a transpose convolutional layer. Finally, each up-sampling block outputs a displacement field that is convoluted and resized to the input size, then summed to output the predicted displacement field. The network was implemented in TensorFlow using NiftyNet [4]. It was trained by using the Adam optimiser, starting at a learning rate of 10-3, with a minibatch of 4. The regularisation weight was set to $\alpha = 2$ to ensure displacement continuity.

Strain Estimation In USE, the strain estimates are obtained by computing the displacement field gradient. However, direct differentiation of the displacement field is rarely used because gradient operations generate a significant amount of noise in the resulting strain map. We used the least-squares strain estimator (LSQSE) to improve the elastogram signal-to-noise ratio [13]. The strain estimation was also implemented in TensorFlow to facilitate efficient parallel computing. In inference, the strain map prediction rate reach a total of 13 images per second on a 12GB NVIDIA GTX-1080ti GPU.

3 Experiments

The performance of our method was evaluated on ultrasound simulations and *in vivo* data. We compared our results with a state-of-the-art strain elastography method called "tOtal Variation Regularization and WINDow-based time delay estimation" (OVERWIND) [17]. The OVERWIND results were obtained by using the publicly available MATLAB implementation, and default parameters were chosen for the simulation experiment. The regularisation coefficients were manually tuned for the *in vivo* data to globally maximize the NCC and Contrast-to-Noise Ratio (CNR) scores of the three cases. The results on simulation were also compared with strain estimates obtained by training our network with a supervised loss function. The supervised loss function was used before for time-delay estimation [25], and corresponded to the mean absolute difference (MAE) between the network prediction and the ground truth labels. The quality of the strain estimates were assessed with the CNR and Signal-to-Noise Ratio (SNR), which are metrics commonly used in USE [24, 17].

Simulation Dataset The Field-II software [11, 10] was used to generate the ultrasound images. Each simulation consisted in a 3D rectangle of size 38x40 mm, containing a cylindrical-shaped inclusion with a randomly assigned diameter (from 8 to 12 mm) and position. The speckle pattern typically observed in ultrasound imaging was obtained by randomly assigning a total of 400,000 scatterers across each digital phantom. The axial compression was assigned randomly to each phantom and represented between 0.5% and 4% of the phantom total length. The Young's modulus of the inclusion was set to different values (i.e. 8, 15, 45 and 75 kPa) while the background was fixed to 25 kPa. Tissue displacements were estimated by finite element method using the Partial Differential Equation Toolbox from MATLAB, and were used to interpolate the scatterers position. A total of 192 RF lines were simulated for each image, with probe central and sampling frequency of 7 and 40 MHz, respectively. The background scatterers were associated with a random intensity value from a Gaussian distribution, to mimic homogeneous tissue. To increase the network's robustness to noise and image intensities, the inclusion scatterers intensities were either assigned to 0 or similar to the background. Moreover, white Gaussian noise with random signal power values, from 5 to 20 dBw, was added on each image to increase robustness to noise. Finally, 160 ultrasound image pairs were simulated, where 100 were used for training and 30 for validation and testing, respectively.

***In Vivo* Dataset** We created our own *in vivo* dataset by collecting images of the arms in three human volunteers. The RF data was acquired from a Cicada 128PX system equipped with a 10 MHz linear probe from Cephasonics (Cephasonics Inc., USA). The images were reconstructed using the delay-and-sum beamformer from SUPRA [5]. Experimental protocol consisted in acquiring sequences of images while slowly applying an axial compression on the volunteer's arm with the handheld ultrasound probe. The *in vivo* dataset included

1300 image pairs for training and 300 pairs for validation and testing. The three image pairs presented in the results section were taken from the testing partition and exhibit one or several blood vessels located in the arm.

3.1   Results

Figure 2 shows axial strain images estimated from three simulated image pairs and obtained with the finite element method, OVERWIND and both supervised and unsupervised models. The averaged CNR and SNR values of the entire testing dataset (i.e., 30 image pairs) are displayed in table 1.

   Ground Truth   Unsupervised   Supervised   Overwind
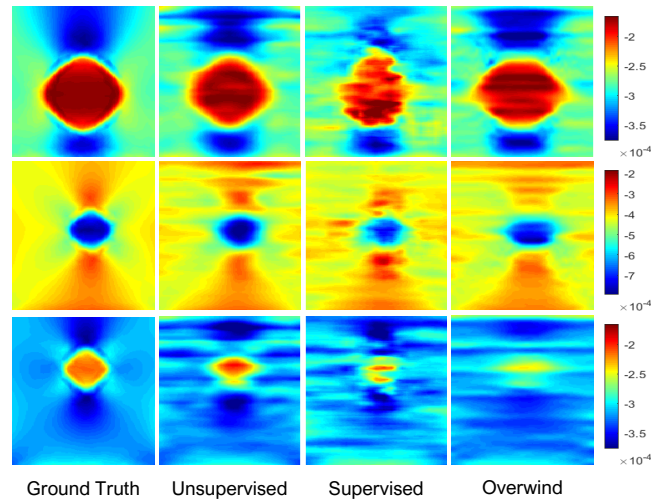
Fig. 2. Comparison of three axial strain fields computed from ultrasound simulations by finite element method (i.e., ground truth), OVERWIND, the unsupervised and supervised network. OVERWIND's regularisation parameters : $\alpha1 = \beta1 = 20$ and $\alpha2 = \beta2 = 8$.

Table 1. Mean and standard deviation of SNR and CNR for the strain images of the simulation testing dataset obtained with finite element method, OVERWIND, the unsupervised and supervised network.

	Ground truth (mean std)	Unsupervised (mean std)	Supervised (mean std)	OVERWIND (mean std)
SNR	7.51 (2.61)	6.95 (2.54)	7.79 (2.01)	9.31 (3.51)
CNR	9.15 (2.73)	7.70 (3.8)	4.22 (2.08)	6.33 (3.6)

Figure 3 shows axial strains estimated from three different *in vivo* image pairs. CNR and SNR values of each case can be found in table 2. The local NCC values between the post-compression and resampled pre-compression images are also displayed, to indicate the quality of the predicted displacement fields. The mean of the three CNR values are 7 and 8.43 for our unsupervised model and OVERWIND, respectively. The average SNR values are 0.31 and 0.7 for our method and OVERWIND, respectively. Displacement field estimates from both the simulation and *in vivo* experiments are available in the supplementary material.
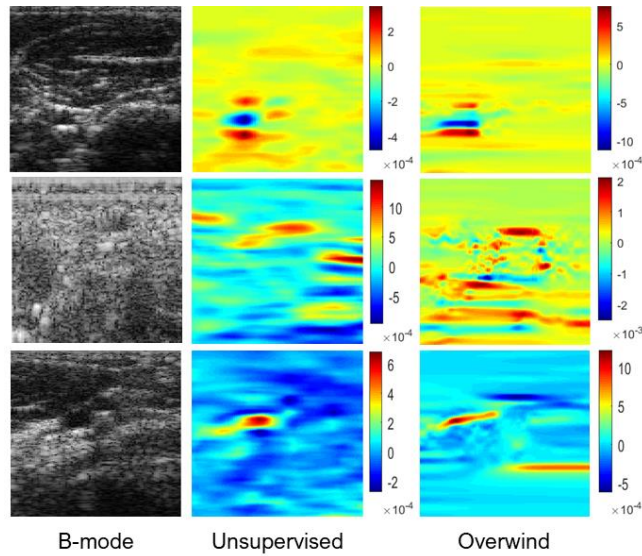
Fig. 3. Comparison of axial strains estimated by our method and OVERWIND on three image pairs from the *in vivo* testing dataset . OVERWIND's regularisation parameters : $\alpha 1 = \beta 1 = 0.2$ and $\alpha 2 = \beta 2 = 0.05$.

Table 2. SNR and CNR of strain estimates from our unsupervised method and OVER-WIND, and local NCC scores for the three image pair results from the *in vivo* dataset.

	Unsupervised			OVERWIND		
	Case 1	Case 2	Case 3	Case 1	Case 2	Case 3
SNR	0.14	0.07	0.1	0.12	0.35	0.23
CNR	6.22	5.2	4.5	4.9	3.15	2.23
LNCC	0.87	0.81	0.94	0.90	0.81	0.93

4   Discussion

In this work, we presented a new deep-learning approach for the estimation of the displacement and strain maps between a pair of ultrasound RF data undergoing an axial compression. We validated our method on both ultrasound simulation and *in vivo* data . Our method is completely unsupervised and ground truth images collected from finite element analysis were only used to assess the performance of our method.

Our results on ultrasound simulation indicate that our method predicts strain estimates with a significantly better CNR compared to the supervised network, with 7.70 and 4.20 respectively. This suggests that our training loss function, which includes a term that penalizes the strain smoothness, improves the strain contrast. Our experiments showed comparable results to OVERWIND, a state-of-the-art method which has, in terms of CNR, already outperformed a previous classical approach [6] by 27.26%, 144.05%, and 49.90% on average in simulation, phantom, and in-vivo data, respectively, as reported in [17]. In addition, our method is fully automatic in inference while OVERWIND strain estimation relies on the correct adjustment of its regularisation parameters.

Finally, the OVERWIND real-time performance had not been quantitatively reported while our approach reached a strain prediction rate of about 13 frames per second on a 12GB NVIDIA GTX-1080ti GPU. The real-time inference of our network and its ability to be trained without ground truth labels represents a great potential for the use of learning-based methods in ultrasound strain elastography.

References

1. Ara, S.R., Mohsin, F., Alam, F., Rupa, S.A., Awwal, R., Lee, S.Y., Hasan, M.K.: Phase-based direct average strain estimation for elastography. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control **60**(11), 2266–2283 (2013). https://doi.org/10.1109/TUFFC.2013.6644732
2. Chen, H., Shi, H., Varghese, T.: Improvement of elastographic displacement estimation using a two-step cross-correlation method. Ultrasound in Medicine and Biology **33**(1), 48–56 (2007). https://doi.org/10.1016/j.ultrasmedbio.2006.07.022
3. Correas, J.M., Tissier, A.M., Khairoune, A., Khoury, G., Eiss, D., Hélénon, O.: Ultrasound elastography of the prostate: State of the art. Diagnostic and Interventional Imaging **94**(5), 551–560 (2013). https://doi.org/10.1016/j.diii.2013.01.017
4. Gibson, E., Li, W., Sudre, C., Fidon, L., Shakir, D.I., Wang, G., Eaton-Rosen, Z., Gray, R., Doel, T., Hu, Y., Whyntie, T., Nachev, P., Modat, M., Barratt, D.C., Ourselin, S., Cardoso, M.J., Vercauteren, T.: Niftynet: a deep-learning platform for medical imaging. Computer Methods and Programs in Biomedicine **158**, 113–122 (2018). https://doi.org/10.1016/j.cmpb.2018.01.025

5. Göbl, R., Navab, N., Hennersperger, C.: SUPRA: open-source software-defined ultrasound processing for real-time applications: A 2D and 3D pipeline from beamforming to B-mode. International Journal of Computer Assisted Radiology and Surgery **13**(6), 759–767 (2018). https://doi.org/10.1007/s11548-018-1750-6

6. Hashemi, H.S., Rivaz, H.: Global Time-Delay Estimation in Ultrasound Elastography. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control **64**(10), 1625–1636 (2017). https://doi.org/10.1109/TUFFC.2017.2717933

7. Hiltawsky, K.M., Krüger, M., Starke, C., Heuser, L., Ermert, H., Jensen, A.: Freehand ultrasound elastography of breast lesions: Clinical results. Ultrasound in Medicine and Biology **27**(11), 1461–1469 (2001). https://doi.org/10.1016/S0301-5629(01)00434-3

8. Hu, Y., Modat, M., Gibson, E., Li, W., Ghavami, N., Bonmati, E., Wang, G., Bandula, S., Moore, C.M., Emberton, M., Ourselin, S., Noble, J.A., Barratt, D.C., Vercauteren, T.: Weakly-supervised convolutional neural networks for multimodal image registration. Medical Image Analysis **49**, 1–13 (2018). https://doi.org/10.1016/j.media.2018.07.002

9. 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. pp. 2462–2470 (2017). https://doi.org/10.1109/CVPR.2017.179

10. Jensen, J.A.: FIELD: A program for simulating ultrasound systems. Medical and Biological Engineering and Computing **34**, 351–352 (1996)

11. Jensen, J.A., Svendsen, N.B.: Calculation of Pressure Fields from Arbitrarily Shaped, Apodized, and Excited Ultrasound Transducers. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control **39**(2), 262–267 (1992). https://doi.org/10.1109/58.139123

12. Jeong, W.K., Lim, H.K., Lee, H.K., Jo, J.M., Kim, Y.: Principles and clinical application of ultrasound elastography for diffuse liver disease. Ultrasonography **33**(3), 149–160 (2014). https://doi.org/10.14366/usg.14003

13. Kallel, F., Ophir, J.: A least-squares strain estimator for elastography. Ultrasonic Imaging **19**(3), 195–208 (1997). https://doi.org/10.1177/016173469701900303

14. Kibria, M.G., Rivaz, H.: Gluenet: Ultrasound elastography using convolutional neural network. In: Stoyanov, D, Taylor, Z, Aylward, S, Tavares, JMRS, Xiao, Y, Simpson, A, et al. (eds.), MICCAI 2018, LNCS. vol. 11070, pp. 21–28. Springer International Publishing (2018). https://doi.org/10.1007/978-3-030-01045-4_3

15. Kuzmin, A., Zakrzewski, A.M., Anthony, B.W., Lempitsky, V.: Multi-frame elastography using a handheld force-controlled ultrasound probe. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control **62**(8), 1486–1500 (2015). https://doi.org/10.1109/TUFFC.2015.007133

16. Kwak, J.Y., Kim, E.K.: Ultrasound elastography for thyroid nodules: Recent advances. Ultrasonography **33**(2), 75–82 (2014). https://doi.org/10.14366/usg.13025

17. Mirzaei, M., Asif, A., Rivaz, H.: Combining Total Variation Regularization with Window-Based Time Delay Estimation in Ultrasound Elastography. IEEE Transactions on Medical Imaging **38**(12), 2744–2754 (2019). https://doi.org/10.1109/TMI.2019.2913194

18. Ophir, J., Cespedes, I., Garra, B., Ponnekanti, H., Huang, Y., Maklad, N.: Elastography: Ultrasonic imaging of tissue strain and elastic modulus in vivo. European Journal of Ultrasound **3**(1), 49–70 (1996). https://doi.org/10.1016/0929-8266(95)00134-4

19. Ophir, J., Céspedes, I., Ponnekanti, H., Yazdi, Y., Li, X.: Elastography: A quantitative method for imaging the elasticity of biological tissues. Ultrasonic Imaging **13**(2), 111–134 (1991). https://doi.org/10.1177/016173469101300201
20. Shi, H., Varghese, T.: Two-dimensional multi-level strain estimation for discontinuous tissue. Physics in Medicine and Biology **52**(2), 389–401 (2007). https://doi.org/10.1088/0031-9155/52/2/006
21. Sigrist, R.M., Liau, J., Kaffas, A.E., Chammas, M.C., Willmann, J.K.: Ultrasound elastography: Review of techniques and clinical applications. Theranostics **7**(5), 1303–1329 (2017). https://doi.org/10.7150/thno.18650
22. Varghese, T., Konofagou, E.E., Ophir, J., Alam, S.K., Bilgen, M.: Direct strain estimation in elastography using spectral cross-correlation. Ultrasound in Medicine and Biology **26**, 1525–1537 (2000). https://doi.org/10.1016/S0301-5629(00)00316-1
23. Varghese, T.: Quasi-Static Ultrasound Elastography. Ultrasound Clinics **4**(3), 323–338 (2009). https://doi.org/10.1016/j.cult.2009.10.009
24. Wang, J., Huang, Q., Zhang, X.: Ultrasound elastography based on the normalized cross-correlation and the pso algorithm. In: 2017 4th International Conference on Systems and Informatics (ICSAI). pp. 1131–1135. IEEE (2017). https://doi.org/10.1109/ICSAI.2017.8248455
25. Wu, S., Gao, Z., Liu, Z., Luo, J., Zhang, H., Li, S.: Direct reconstruction of ultrasound elastography using an end-to-end deep neural network. In: Stoyanov, D, Taylor, Z, Aylward, S, Tavares, JMRS, Xiao, Y, Simpson, A, et al. (eds.), MICCAI 2018, LNCS. vol. 11070, pp. 374–382. Springer International Publishing (2018). https://doi.org/10.1007/978-3-030-00928-1_43