Direct Reconstruction of Ultrasound Elastography Using an End-to-End Deep Neural Network

Sitong Wu[1,2], Zhifan Gao[1], Jianwen Luo[3†], Zhi Liu[3], Heye Zhang[1†], and Shuo Li[4]

[1] Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China.
[2] Shenzhen College of Advanced Technology, University of Chinese Academy of Sciences, China
[3] Tsinghua University, China
[4] University of Western Ontario, Canada

Abstract. In this work, we developed an end-to-end convolutional neural network (CNN) to reconstruct the ultrasound elastography directly from radio frequency (RF) data. The novelty of this network is able to infer the distribution of elastography from real RF data by only using computational simulation as the training data. Moreover, this framework can generate displacement and strain field respectively both from ultrasound RF data directly. We evaluated the performance of this network on 50 simulated RF datasets, 42 phantom datasets, and 4 human datasets. The best results of signal-to-noise ratio (SNR) and contrast-to-noise ratio (CNR) in simulated data, phantom data, and human data are 39.5dB and 69.64dB, 32.64dB and 48.76dB, 23.24dB and 46.22dB, respectively. Furthermore, we also compare the performance of our method to the state-of-art ultrasound elastography using normalized cross-correlation (NCC) technique. From this comparison, it shows that that our method can effectively compute the strain field robustly and accurately in the this paper. These results might imply great potential of this deep learning method in ultrasound elastography application.

Keywords: ultrasound elastography, convolutional neural network, tissue displacement and strain, RF data

1 Introduction

The mechanical behaviour of healthy and pathological tissue are different under external compression because the stiffness of tissue will be changed by the diseases [6]. Ultrasound elastography has been one popular technique in clinical practice to examine the distribution of tissue strain in the suspected lesion area, according to the tissue displacement derived from the two-dimensional radio frequency (2D RF) data. Many studies have tried to develop varieties of

[†]Corresponding author: Heye Zhang (hy.zhang@siat.ac.cn) and Jianwen Luo (luo_jianwen@mail.tsinghua.edu.cn)

feature extractors, such as normalized cross-correlation [2] and optical flow [3], for estimating the tissue displacement from 2D RF data and then predicting the tissue strain. However, these attempts are always disturbed by the noises or insufficient feature information. Moreover, most of features in previous works only provide low-level information related to local variation of 2D RF data, rather the high-level semantic information related to the tissue strain.

From the viewpoint of computer vision, ultrasound elastography can be considered to explore the key visual information (tissue strain), which attracts the human attention, in the complex environment (noise-corrupted RF data). Extracting high-level semantic image features from RF data can help to robustly reconstruct the elastography. Recently popular deep neural networks have shown its effectiveness in medical image analysis, as it can effectively extract, represent, and integrate highly semantic features without manual intervention.

In this paper, we develop an end-to-end deep neural network to predict the strain and displacement fields directly from RF data under the condition that tissue compressed by the constant external force (i.e. quasi-static ultrasound elastography). The architecture of our framework includes two convolutional neural networks to extract high-level features without manual intervention, for predicting the tissue displacement and strain, respectively. Furthermore, the novelty of our work is that only the simulated data are used in training this deep neural network. The performance of our approach is validated on simulated data (50 cases), phantom data (42 cases) and real clinical data (4 patients), by comparing with the ground truth (in simulated data) and a state-of-the-art ultrasound elastography method. Our experimental results have proved the generalization of our proposed framework in the reconstruciton of elastography. To the best of our knowledge, it is the first time to reconstruct the ultrasound elastography using the deep learning method with only simulated training data.

2 Methodology

Network Architecture. The architecture of our network is composed of two stages: one is to estimate tissue displacement from 2D RF data; the other is to predict tissue strain from the tissue displacement. The network architecture is illustrated in Figure 1. The deformation of soft tissue can be measured by ultrasound technique. We collect two sets of 2D RF data (denoted by I_1 and I_2) before and after compressing the soft tissue. The phase time interval between these two 2D RF data are highly related to the tissue displacement. We get I_1 and I_2 as the input of our network after giving a small global elongation to the post- compression RF data.

In the displacement estimation stage, we firstly use the separable convolution to extract and concatenate the hybrid features of I_1 and I_2. Then we construct a five-layer convolution network to extract the contextual information of the difference between I_1 and I_2. We adopt the locally-connected convolution, instead of the commonly-used fully-connected convolution, to remove the spatial correlation, for increasing the ability of the convolutional kernel to express the

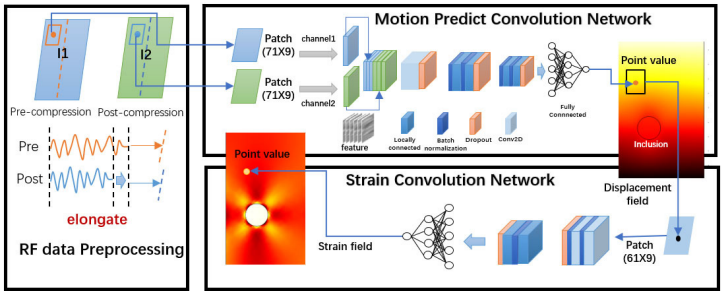

Fig. 1. The flowchart of our approach. The 2D RF data are represented by the first stage of our network after the preprocessing. Then the tissue strain field are computed from the displacement field by the second stage of our network. Note our approach can directly compute both of tissue displacement field and strain field.

high-level information related to spatial location. This will help to differentiate the phase time intervals of tissues with different elasticity. The components to further process the output feature maps in all layers include (from low level to high level) dropout, batch normalization, batch normalization + dropout, batch normalization, batch normalization + dropout, respectively. Then the feature maps generated by the above convolution network are resized into a vector for the subsequent fully-connected network with three layers (64, 32 and 1 units). This fully connected network can get a distribution of the tissue displacements resulting from the compression.

In the strain prediction stage, we use another convolution network with three layers to extract the high-level semantic information of the tissue displacement. The feature maps produced from these three layers are further processed by batch normalization, batch normalization + dropout, batch normalization, and dropout, respectively. Similar to the displacement estimation, a fully-connected network is used to predict the strain from the vectorized feature maps derived from the local displacement field.

Implementation Details. All input images are resized to 2608×128 (pixels) in both training and testing. The patch sizes in the displacement estimation stage and strain prediction stage are 71×9 and 61×9, respectively. In the training process, we use ADAM with the momentum 0.9 as the optimization algorithm. The iteration number is 30 epochs. The learning rate is 10e-4. The loss function is the absolute mean error. All input images are resized to 2608×128 for training and testing.

3 Experiments and Results

3.1 Datasets and Evaluation Indices

We evaluated the performance our approach on simulation data, phantom data and real data. The training set only contains the 40 simulation data. The testing

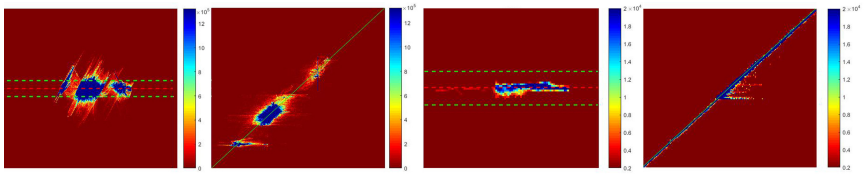

Fig. 2. The results of the Bland-Altman analysis and linear regression (LR) between the ground truth (x-axis in LR) and our approach (y-axis in LR) in all simulation data. The left two plots correspond to the tissue displacement, and the right two plots correspond to the strain percentage with respect to the tissue size. Note that the color indicates the density or the frequency of the scatter points in the two-dimensional BA and LR plane

Table 1. Comparing with the state-of-the-art method in simulation data.

simulation data	displacement		strain		
	SNR_d	RMSE	SNR_e	CNR_e	RMSE
Our approach	96.0746±8.8622	0.2450	28.6832±6.800	62.3249±10.8053	0.2938
AM-Kalman	104.6249±5.3247	0.0852	15.1824±4.1605	31.4127±7.1706	0.9783

set is composed of the remaining simulation, all phantom data and patient data. All of the codes for model training and testing were implemented by TensorFlow on a NIVDIA GTX1070 GPU.

Simulation data. We produced 50 simulation tissues with 3 cm width and 5 cm depth, each of which contains with 10000 scatter points with pre-specified locations and an inclusion with random number of scatter points. The Young's module of the background in tissue was set by 25 kPa and the inclusions were set as four different values (8 kPa, 14 kPa, 45 kPa and 80 kPa). Then, we used finite element method (by commercially available software COMSOL 5.1) to compute the positions of all scattering points after compressing the tissue, and moreover the ground truth of tissue displacement field and strain field. Finally, we computed the 2D RF data of the tissue before and after compression by an ultrasound simulation system (Field II) [1] with 6 MHz central frequency and 40 MHz sampling frequency.

Phantom data. We collected the phantom data from a commercially available elasticity QA phantom (CIRS 049, Norfolk, VA, USA) using a VerasonicsVantage 256 system (Verasonics Inc., WA, USA) equipped with an L12-5 transducer. This transducer was applied to image 42 regions of the phantom with four different inclusions (Young's module are 8 kPa, 14 kPa, 45 kPa and 80 kPa) and the background region (25 kPa). The central frequency and sampling frequency of the scanning were 6.25 MHz and 40 MHz, respectively.

Patient data. We acquired the four real clinical data (one is liver and three are breast) from the public dataset [4].

Evaluation Indices. We used the RMSE and displacement signal-to-noise ratio to measure the quality of the displacement estimate. The SNR is defined

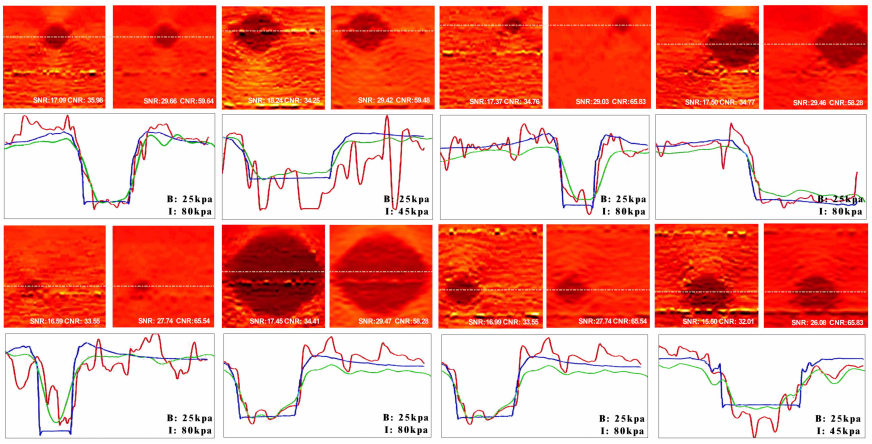

Fig. 3. Sample results of strain estimation in eight simulation data with different Young's module ("B": background; "I": inclusion). In each pair of images, the left is the previous method and the right is our approach. The strain profile in the second and fourth rows corresponds to the white dotted line of the strain field showed in the first and third rows. In the strain profile, the blue, green and red curves corresponds to the ground truth, our approach and AM-Kalman, respectively.

Table 2. Comparing with the state-of-the-art method in phantom and patient data.

	phantom data		patient data	
	SNR_e	CNR_e	SNR_e	CNR_e
Our approach	30.4339±4.0308	43.2689±10.7117	23.2628±3.2032	40.3475±18.3785
AM-Kalman	17.9304±3.3346	29.5817±7.7203	9.6330±12.8452	36.2232±24.2978

by $SNR_d = 10\log(l_d - l_g)^2/l_g^2$, where l_d and l_g are displacement profiles generated by our approach and the ground truth, respectively [5]. Then we apply elastographic signal-to-noise ratio (SNR_e) and elastographic contrast-to-noise ratio (CNR_e) to measure the quality of the strain estimate [5], i.e. $SNR_e = 10\log e_i/\sigma_i$, $\quad CNR_e = 10\log(2(e_b - e_i)^2)/(\sigma_b^2 + \sigma_i^2)$, where e_b and σ_b are the mean value and variance of the background, and e_i and σ_i the mean value and variance of the inclusion.

3.2 Comparison with ground truth and other methods

Simulation results. Table 1 shows the performance of our approach in the tissue displacement estimation and the strain prediction. In the displacement, the values of SNR_d and RMSE are 96.0746 ± 8.8622 dB and 0.245, respectively. This indicates the very small difference between the displacement field computed by our approach and the ground truth. In the strain prediction, the value of RMSE (=0.2938) indicates the error of the strain field produced by our approach is at a low level. Besides, the values of SNR_e (28.6832 ± 6.800 dB) and CNR_e (62.3249

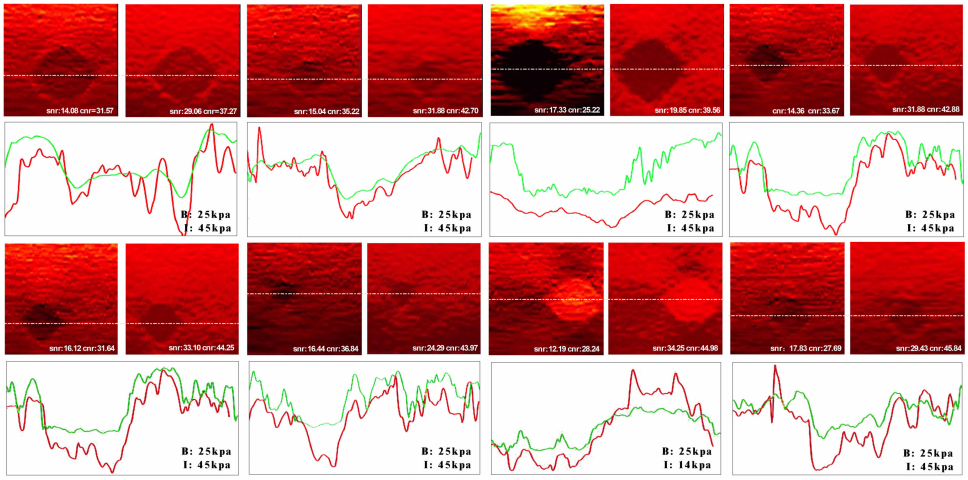

Fig. 4. Sample results of strain estimation in eight phantom data with different Young's module ("B": background; "I": inclusion). The green and red curves shows the strain profile of our approach and AM-Kalman, respectively.

$\pm$ 10.8053 dB) shows not only the strain field in the inclusion less disturbed by noise, but also the strain contrast between the inclusion and background has the high-level signal-to-noise ratio. Then, Figure 2 shows the results of the Bland-Altman analysis (BA) and linear regression (LR) between the ground truth (x-axis in LR) and our approach (y-axis in LR) in all simulation data. The left two plots correspond to the tissue displacement, and the right two plots correspond to the strain percentage with respect to the tissue size. Note that the color indicates the density or the frequency of the scatter points in the two-dimensional BA and LR plane.

Furthermore, we have compared our framework with the a state-of-the-art method proposed by Rivaz et al. [4] (denoted by AM-Kalman). This method applied 2D analytic minimization to obtain the tissue displacement from RF data and then used the Kalman filter to calculate the smooth strain field. The results in Table 1 shows the displacement estimation of our approach is slightly worse than AM-Kalman (our approach is 8.5503 dB lower for SNR_d and 0.1598 higher for RMSE), but our approach shows much better performance than AM-Kalman in the strain prediction (our approach is 13.5008 dB higher for SNR_e, 30.9122 dB higher for CNR_e, and 0.6845 lower for RMSE). The strain is the important parameter concerned in clinical practice, and thus our framework is superior to AM-Kalman in the simulation data.

Figure 3 shows the sample results from eight simulation data for the comparison between our approach and AM-Kalman in the strain estimation. These sample results indicate that our approach can predict the strain field (green curve) closer to the ground truth (blue curve) than AM-Kalman (red curve).

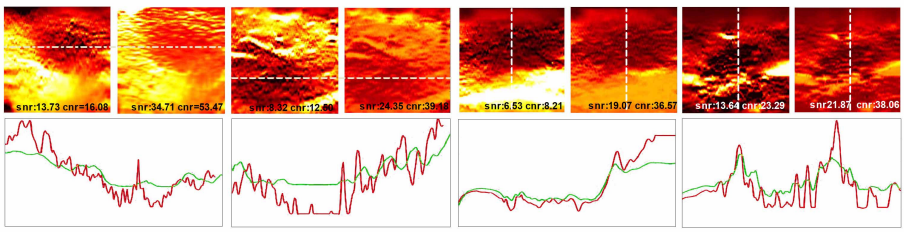

Fig. 5. Sample results of strain estimation in four patient data. The green and red curves shows the strain profile of our approach and AM-Kalman, respectively.

Table 3. Comparing with variants of our network in the displacement estimation.

		simulation				phantom		patient data	
		SNR_d	SNR_e	CNR_e	RMSE	SNR_e	CNR_e	SNR_e	CNR_e
pooling (max)	avg	66.01±6.08	8.95±0.65	6.26±1.11	0.65	5.48±1.04	9.58±8.36	3.58±10.59	7.47±4.45
conv layers (5)	4	66.79±5.00	7.94±1.57	11.53±0.52	0.85	7.39±1.34	14.55±4.06	4.45±2.59	11.13±4.71
	6	65.43±7,25	7.67±3.68	8.46±3.00	1.04	8.23±4.69	3.73±2.68	0.68±1.64	0.14±2.54
FC layers (3)	2	60.09±5.62	3.74±1.74	5.06±4.28	1.17	3.42±1.95	3.89±2.73	2.00±2.95	1.95±0.94
	4	50.43±3.83	3.20±0.58	3.82±1.65	2.13	1.45±0.93	4.04±4.64	1.35±1.24	4.50±1.11
dilation rate (1)	2	65.82±3.66	8.58±1.57	13.16±3.90	0.74	6.28±4.23	17.60±6.06	8.05±3.49	9.45±6.56
	3	75.92±3.96	13.34±2.67	25.45±7.02	0.80	14.33±1.49	24.56±10.27	12.37±4.93	15.57±1.50
patch_size (71x9)	71×5	89.43±8,86	15.35±1.57	29.37±12.48	0.39	19.16±2.08	34.83±19.12	12.59±2.26	29.51±7.53
	45×9	94.33±7.25	16.45±3.35	36.42±9.23	0.27	34.96±4.22	39.16±9.09	19.83±5.11	37.63±14.20
	31×9	55.18±5.10	2.72±0.94	3.78±1.87	1.24	5.56±1.34	3.453±5.39	0.45±1.43	11.46±5.83
Our		96.07±8.86	28.68±6.80	62.32±10.80	0.24	30.43±4.03	43.26±10.71	23.26±3.20	40.34±18.37

Phantom results. Because no ground truth in the phantom experiments, we evaluate the performance of our approach by comparing SNR_e and CNR_e with AM-Kalman. Table 2 shows the performance of our approach is better than with AM-Kalman (SNR_e: 30.4339 dB vs. 17.9304 dB; CNR_e: 43.2689 dB vs. 29.5817 dB). Figure 4 shows the sample results from eight phantom data for the comparison between our approach and AM-Kalman in the strain estimation. These sample results indicate that our approach can predict the strain field with higher SNR_e and CNR_e than AM-Kalman.

Patient data results. Similar to the phantom experiments, we also compared the values of SNR_e and CNR_e with AM-Kalman in the patient data experiments. Table 2 shows the performance of our approach is superior to with AM-Kalman (SNR_e: 23.2628 dB vs. 9.6330 dB; CNR_e: 40.3475 dB vs. 36.2232 dB). Figure 5 shows the results of all four patient data for the comparison between our approach and AM-Kalman in the strain estimation. These sample results show the higher SNR_e and CNR_e of the strain field estimated by our approach than AM-Kalman.

3.3 Comparisons between different configurations of our network

The ablation analysis aims to investigate the effectiveness of our network architecture. Table 3 shows the results of the network components for displacement

Table 4. Comparing with variants of our network in the strain prediction.

		simulation			phantom		patient data	
		SNR_e	CNR_e	RMSE	SNR_e	CNR_e	SNR_e	CNR_e
pooling (max)	avg	28.58±5.08	62.07±15.53	0.35	28.37±5.33	47.59±16.69	15.15±9.19	33.91±15.25
conv layers (5)	3	30.51±3.71	66.99±13.02	0.37	32.64±3.63	48.76±10.06	20.37±3.44	40.65±21.46
	4	30.67±2.13	69.39±14.30	0.37	29.66±3.12	45.39±17.64	23.24±5.97	40.59±22.14
FC layers (3)	2	39.50±4.28	69.64±15.79	0.37	27.5±6.58	44.92±17.77	19.87±3.76	38.76±23.43
	4	30.69±5.24	65.93±16.66	0.35	30.91±2.49	45.46±18.06	16.66±2.82	46.22±15.20
dilation rate (1)	2	30.17±5.22	65.88±21.74	0.36	28.60±4.67	45.00±13.84	17.39±3.66	43.79±15.73
	3	34.84±6.24	68.86±14.49	0.37	28.34±4.23	46.00±17.85	21.49±5.37	40.85±18.40
patch_size (71×9)	(61×5)	32.42±3.02	67.25±10.44	0.34	27.12±7.69	43.56±19.17	16.96±2.99	40.17±18.34
	(41×9)	30.56±4.87	60.76±15.58	0.35	27.83±6.26	40.15±23.58	14.29±1.15	40.17±18.34
	(31×9)	29.06±7.00	62.34±18.94	0.39	27.02±6.67	45.18±18.43	16.79±1.25	39.17±19.44
Our		28.68±6.80	62.32±10.80	0.29	30.43±4.03	43.26±10.71	23.26±3.20	40.34±18.37

estimation in simulation, phantom and patient data, Table 4 displays those for strain prediction. The digits within the parentheses in the first column shows the configuration of our network architecture. The results in these two tables indicate the effectiveness of current configurations of the architecture in the pooling strategy, layer number, dilated rate of the convolution kernel, and patch size.

4 Conclusion

In this study, we have developed an end-to-end deep learning approach to recover the tissue displacement and strain in ultrasound elastography directly from radio frequency (RF) data. The performance of our approach was tested on 50 simulation data, 42 phantom data and 4 patient data, by comparing with the gold standard and a state-of-the-art method. Experimental results show that our method is effective in calculating the strain field, and moreover implies great potential of this deep learning method in ultrasound elastography application.

References

[1] Jensen, J.A.: FIELD: A program for simulating ultrasound systems. In: 10th Nordicbaltic Conference On Biomedical Imaging. pp. 351–353 (1996)

[2] Luo, J., Konofagou, E.E.: A fast normalized cross-correlation calculation method for motion estimation. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control 57(6), 1347–1357 (2010)

[3] Pellot-Barakat, C., Frouin, F., Insana, M.F., Herment, A.: Ultrasound elastography based on multiscale estimations of regularized displacement fields. IEEE Transactions on Medical Imaging 23(2), 153–163 (2004)

[4] Rivaz, H., Boctor, E.M., Choti, M.A., Hager, G.D.: Real-time regularized ultrasound elastography. IEEE Transactions on Medical Imaging 30(4), 928–945 (2011)

[5] Srinivasan, S., Righetti, R., Ophir, J.: Trade-offs between the axial resolution and the signal-to-noise ratio in elastography. Ultrasound in Medicine & Biology 29(6), 847866 (2003)

[6] Zaleska-Dorobisz, U., Kaczorowski, K., Pawluś, A., Puchalska, A., Inglot, M.: Ultrasound elastography - review of techniques and its clinical applications. Advances in Clinical and Experimental Medicine 23(4), 645–655 (2014)