Displacement Estimation in Ultrasound Elastography using Pyramidal Convolutional Neural Network

Ali K. Z. Tehrani, Hassan Rivaz

***Abstract*—In this paper, two novel deep learning methods are proposed for displacement estimation in ultrasound elastography. Although Convolutional Neural Networks (CNN) have been very successful for displacement estimation in computer vision, they have been rarely used for ultrasound elastography. One of the main limitations is that the Radio Frequency (RF) ultrasound data, which is crucial for precise displacement estimation, has vastly different frequency characteristics compared to images in computer vision. Top-rank CNN methods used in computer vision applications are mostly based on a multi-level strategy which estimates finer resolution based on coarser ones. This strategy does not work well for RF data due to its large high frequency content. To mitigate the problem, we propose Modified Pyramid, Warping and Cost volume Network (MPWC-Net) and RFMPWC-Net, both based on PWC-Net, to exploit information in RF data by employing two different strategies. We obtained promising results using networks trained only on computer vision images. In the next step, we constructed a large ultrasound simulation database, and proposed a new loss function to fine-tune the network to improve its performance. The proposed networks and well-known optical flow networks as well as state-of-the-art elastography methods are evaluated using simulation, phantom and *in vivo* data. Our two proposed networks substantially outperform current deep learning methods in terms of Contrast to Noise Ratio (CNR) and Strain Ratio (SR). Also, the proposed methods perform similar to the state-of-the-art elastography methods in terms of CNR and have better SR by substantially reducing the underestimation bias.**

***Index Terms*—Ultrasound elastography, Displacement estimation, Optical flow, Convolutional neural network, PWC-Net.**

I. Introduction

Ultrasound imaging is being increasingly used as an inexpensive and easy-to-use imaging modality in numerous diagnosis and image-guided intervention applications. Ultrasound elastography (USE) is an imaging technique that reveals viscoelastic properties of tissue, and has been applied to many applications including breast lesion characterization [1] and ablation monitoring [2]–[5]. USE compliments B-mode ultrasound by providing biomechanical properties of the tissue [6].

Among different USE methods, free-hand palpation has gained much popularity due its simplicity, low cost and ease-of-use. The basic idea of free-hand palpation method is that the operator compresses the tissue by the ultrasound probe. The images before and after compression are compared to obtain the displacement of each individual sample. This displacement can be used to obtain strain map which has relative elasticity information [7], [8]. The quality of USE mainly depends on the fidelity of the displacement estimation. Window-based [8]–[12] and optimization-based [13]–[15] methods are two main approaches for displacement estimation in USE. Window-based methods try to find the displacements of each individual sample by considering a window around the sample in pre- and post-compression images and assuming that the displacement within the window is constant. In the next step, a similarity metric such as Normalized Cross Correlation (NCC) is chosen to find the corresponding windows [8], [9]. Optimization-based methods use a regularized cost function to find the displacements, therefore they are more robust to signal decorrelation and out of plane motion [13], [16], [17]. GLobal Ultrasound Elastography (GLUE) is a recent optimization-based method [14] with an implementation available online at code.sonography.ai. GLUE aims to estimate sub-pixel displacement and requires initial estimate of the displacement which is obtained by dynamic programming (DP) [15]. The displacement estimation in USE can also be viewed as a non-rigid registration [18] or optical flow problem [19]–[21].

Convolutional Neural Network (CNN) models have been successfully trained to perform many applications such as classification [22] and segmentation [23]. Recently, CNN has been used for optical flow problem [24]–[27]. FlowNet is among the first attempts to extract optical flow using deep learning architectures [25]. Before FlowNet, patch- and point-based deep learning methods were used. These methods were only able to extract optical flow of a point or a small patch of the images. As such, they were computationally expensive as it was necessary to run them many times to cover the entire image. Two variants of FlowNet were proposed [25]: FlowNetS and FlowNetC. FlowNetS has a U-shape architecture with a contracting and an expanding path, and as such, shares many similarities with U-Net [23]. FlowNetS uses coarse outputs in the refinement section to build the finer outputs and uses multi-scale loss function for optimization. FlowNetC is the other variant of FlowNet that differs from FlowNetS only in the contracting part. Instead of concatenating input images and using a U-shape network, it extracts features of each input separately and exploits a correlation layer to merge information from features of the two images. Although they reported better performance with FlowNetS, Mayer *et al.* [26] show that with better learning schedule and more training data, FlowNetC outperforms FlowNetS.

A. K. Z. Tehrani and H. Rivaz are with the Department of Electrical and Computer Engineering, Concordia University, Canada, e-mail:A_Kafaei@encs.concordia.ca and hrivaz@ece.concordia.ca

Manuscript received; revised

Following the success of FlowNet, Mayer *et al.* [26] stack several FlowNetS and FlowNetC to improve the accuracy of FlowNet. They show that the optimum architecture is to use FlowNetC as the first block, followed by two FlowNet blocks. This architecture is called FlowNet2CSS. In order to improve the network accuracy for small displacements, another FlowNet is used and it is trained on a database with small displacements. A fusion network is then proposed to fuse the outputs of FlowNet (trained to provide small displacements) and FlowNet2CSS. This network called FlowNet2 achieves outstanding performance and is the first CNN architecture that outperforms traditional optical flow algorithms. FlowNet2 performs well but with one drawback. It has many parameters (around 160 million), which makes training difficult and renders inference both computationally expensive and memory exhaustive.

Designing a network with fewer parameters is an active field of research. Ranjan and Black [27] proposed Spatial Pyramide Network (SPynet), which has much lower number of parameters but with an accuracy close to that of FlowNet. Recently, Pyramid, Warping and Cost volume Network (PWC-Net) [24], [28] was proposed, which not only has fewer number of parameters (around 9 million), but also achieves slightly better accuracy compared to FlowNet2. The main idea of PWC-Net is to use pyramidal structure to estimate the optical flow in each level and warp the features by the estimated flow to reduce the search range of the next level. This network utilizes cost volumes (similar to correlation layer) in each pyramid level to extract correlation between features of the two images, and unlike SPynet, warps the features of the second image instead of the image itself. The Table I summerizes the well-known optical flow networks introduced in this paper.

There are two important differences between USE displacement estimation and optical flow that limits the use of optical flow CNN models: 1) Accurate subsample displacement estimation is paramount in USE; 2) RF data is characteristically different from images in computer vision because it has a very large frequency content. Therefore, any optical flow method used for USE must preserve and utilize the information of high frequency RF data for an accurate and robust displacement estimation. USE is a new and less explored deep learning application in medical image processing. Only a few papers tried to apply neural networks for USE [19]–[21], [29], [30].

A deep learning architecture was proposed by Wu *et al.* [29] to estimate displacement and strain. A patch around the sample of interest is fed to the network and the displacement and the strain of the patch are estimated. Gao *et al.* [20] further improved this network by introducing Learning-Using-Privileged-Information (LUPI). LUPI uses displacement as the intermediate loss, and results in better generalization and higher accuracy compared to [29], as well as non-deep learning approaches of DP [15] and optical flow [31]. The main drawback of the networks is that in order to compute the strain and the displacement of an image pair, it is required to run the network many times since this network only takes small patches as the input. In [19], [32], we used FlowNet2 for USE. But since the displacement estimates were not precise even after fine-tuning with Field II simulations [33], [34], they were used as the initial estimator for GLUE, replacing dynamic programming [15] with FlowNet2. In [21], FlowNetCSS is used for USE and it was shown that using simulated images for fine-tuning can be beneficial. The main contribution of our work can be summarized as:

- Two networks, namely Modified Pyramid Warping and Cost volume Network (MPWC-Net) and RFMPWC-Net are proposed for USE, both based on PWC-Net. Both of our proposed networks substantially outperform PWC-Net in USE.
- FlowNet2 has been recently exploited for USE [19], [21]. Our proposed networks are based on PWC-Net, and have more than 10 times fewer parameters compared to FlowNet2 while substantially outperforming it in USE. This is paramount as GPU memory is often a critical bottleneck.
- A fine-tuning strategy and a loss function are proposed to improve the displacement estimation and the corresponding strain quality using simulated data.
- The performance of top optical flow CNNs in USE is presented and analyzed.

We have already put the simulation database generated as part of this manuscript online at data.sonography.ai for reviewers. Similar to our previous work [14]–[16], we will put the network and the tuned weights at code.sonography.ai after acceptance of this manuscript.

II. Methods

A. PWC-Net

The core ideas of PWC-Net are to utilize pyramid structures, cost volume and a refinement network. This leads to substantial reduction of number of the parameters and improvement in the accuracy. Using the pyramid structure reduces the displacement required to be estimated in each resolution, resulting in a smaller search range. The coarser resolution finds large displacements and removes these displacements by warping the second image features with the estimated displacements, and the finer resolution estimates the smaller displacements from the warped image. Unlike FlowNet2 that warps the moved images, PWC-Net warps the features of the moved images so that fewer number of parameters are required for optical flow estimation. PWC-Net utilizes cost volume in each pyramid level. Unlike FlowNet2 that uses correlation layer (cost volume) only as the first block and reports over fitting by using more correlation layers, PWC-Net uses cost volume in all pyramid levels, substantially reducing the number of parameters. Finally, PWC-Net employs a refinement network which is a post processing stage to improve the quality of the estimated optical flow in the last pyramid level [24], [28]. As shown in Fig. 1, PWC-Net is composed of 4 different blocks: feature extraction, warping and cost volume, optical flow estimation and refinement network.

To compute each pyramid output, first the input images are fed into a CNN in order to extract features from the image pyramid, transforming it to a feature pyramid. Then the warping block warps the second image feature map toward the

TABLE I: Summary of some recent optical flow networks. M represents million.

Network	Publication Date	Description	Number of Learnable Weights (Approximately)
FlowNet [25]	2015	An optical flow CNN with two variants of FlowNetS and FlowNetC.	38M
SpyNet [27]	2016	Uses image warping and pyramid structure to decrease number of parameters (much lighter than FlowNet).	1.2M
FlowNetCSS [26]	2017	Built upon FlowNet by concatenating a FlowNetC and two FlowNetS networks.	124M
FlowNet2 [26]	2017	Composed of a FlowNetCSS and a small displacement FlowNet.	162M
PWC-Net [24], [28]	2017, 2018	Uses feature warping, cost volume and refinement network in a pyramid structure to have high accuracy and a moderate number of parameters.	9M

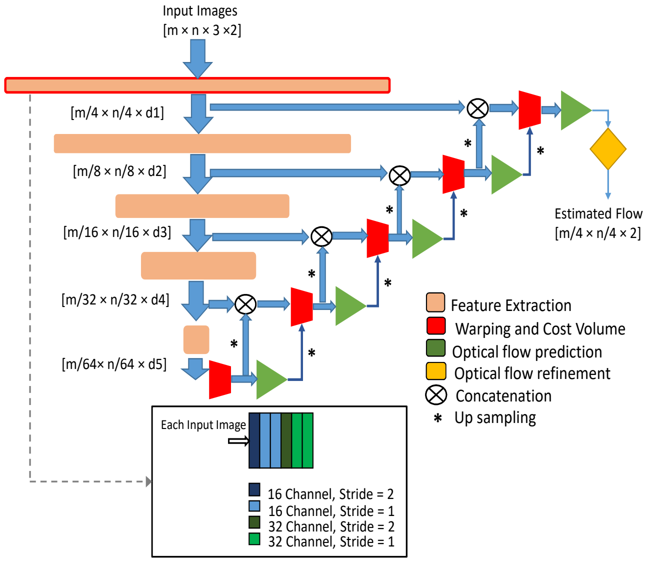


Fig. 1: PWC-Net structure. The feature extraction layer of the final pyramid is outlined by a red box (all kernels in the box are 3×3). m, n, d_x denote image size in axial direction, lateral direction and number of channels of the corresponding layer, respectively.

first one. At the next step, a cost volume is created using the first image feature map and the warped one. This cost volume is then used as an input to the optical flow estimator block in order to estimate the flow. Finally, a refinement network is used to post-process the optical flow. The loss function used in PWC-Net is a multi-scale loss defined in [24]:

$$L(\Theta) = \sum_{l=l_0}^{L} \alpha_l (\|D_\Theta^l(x) - D_{GT}^l(x)\|_q) + \|\Theta\|_2 \qquad (1)$$

where Θ represents the learnable parameters and D_Θ^l and D_{GT}^l denote the estimated and the ground truth flows at the lth level, respectively. This is a regularized loss function where $q < 2$ is chosen to give less penalty to outliers. Also, $\|\Theta\|_2$ is the weight decay which encourages the learnable weights to have small magnitude in order to improve the generalization of the network. For each output resolution, a weight is considered to contribute (α) in the loss function. Generally, higher weights are given to coarser outputs since coarser outputs contribute to build finer ones. The coarse outputs are employed as intermediate losses, and the corresponding ground truths are obtained by down sampling the displacement.

B. *Proposed Methods for USE*

It is common to modify a well-known network for a specific task. As an example, in [35], VGG-16 and ResNet-101 are modified for semantic segmentation by changing the dilations and strides of the convolution layers. In this work, PWC-Net structure is modified for USE wherein accurate subsample displacement estimation using RF data is critical.

PWC-Net contains feature extraction, cost volume and optical flow estimation layer for each pyramid. There are 5 levels and the coarser levels contribute to the estimation of finer resolution levels. As depicted in Fig. 1, the output size is 4 times smaller than the input images. The feature extraction part of the final pyramid level (the first feature extraction layer with red outline, shown in the box) downsamples the input by a factor of 4 using two convolution layers with $stride = 2$. The downsampling of the input images is quite reasonable for computer vision images since there is negligible information in high frequencies. This downsampling reduces the computation complexity, improves the network robustness to noise, and more importantly decreases the displacement and the required search range of the cost volume. However, in USE, accurate subsample displacement estimation is essential and there is valuable phase information in high frequencies, rendering this downsampling detrimental. To cope with this issue, we replace the first two convolution layers with $stride = 2$ with convolution layers with $stride = 1$. This modification provides more information related to displacement estimation for each pyramid level, and useful features can be obtained from high frequency RF data.

An important aspect is the input of the network. RF data, B-mode image and envelope of RF data can be used for displacement estimation. Generally, RF data is the most informative signal for estimation of fine displacements, but using RF data might result in unreliable regions in the pyramidal structures. Envelope and B-mode only contain low frequency information of RF data that can be used for approximation of the displacement but they cannot provide accurate displacement. B-mode and envelope can provide useful information in coarse pyramid levels while RF data contains detailed information for high resolution and high-quality displacement estimation. Consequently, two networks are proposed to exploit RF, B-mode and envelope. In both networks, the downsampling operations (strides = 2) of the final pyramid level are removed.

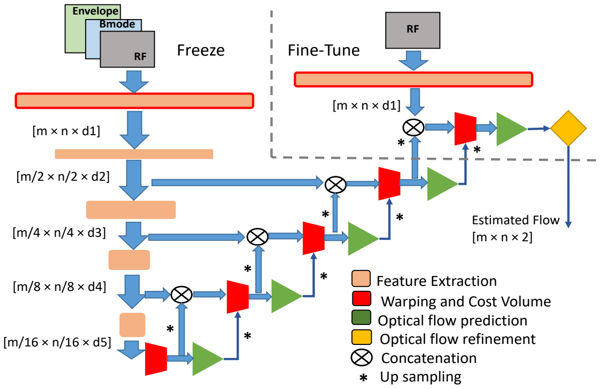


Fig. 2: Proposed RFMPWC-Net structure. A feature extraction layer is added to use features of only RF data to estimate final resolution. The feature extraction layers with red outline have the same weights. The output size of each feature extraction layer, fine-tuned sub-network and frozen network are specified.

In the first network, we concatenate RF data, B-mode and envelope to generate a three-channel input for the network. We name this network Modified PWC-Net (MPWC-Net). This network uses information of B-mode and envelope in low resolutions where RF data cannot provide useful information due to information loss and the network exploits RF data in high resolutions to have high quality subsample displacement estimation.

In the second network, RF data, B-mode and envelope are combined in a different fashion. Concatenated RF, B-mode and envelope is used for displacement estimation of all pyramid levels except for the last pyramid level which has the highest resolution and for that level, only RF data is used for displacement estimation. A feature extraction layer is added to extract useful information of only RF data in the final pyramid level. The block diagram of this method is depicted in Fig. 2. The last layer has the same structure and weights as the feature extraction layer of the main concatenated inputs so no more training is required. This network produces more accurate displacement compared to MPWC-Net, especially in noisy situations because although B-mode and envelope are helpful in low pyramid levels and remove outlier regions, they reduce the accuracy of the network in the final pyramid level. We call this network RF Modified PWC-Net (RFMPWC-Net). Please refer to Supplementary Material for the comparison of the two networks.

C. Simulation Dataset

As part of this manuscript, we generate a simulation dataset using Field II [33], [34]. The dataset consists of one or two inclusions with random positions. The Young's modulus of the tissue is randomly set between 18 to 23 kPa, and the Young's modulus of the hard inclusion is randomly set to a value in the range of 40 to 60 kPa. The average strain varies between 0.5 to 4.5 % and displacements are estimated by Finite Element Method (FEM) using the ABAQUS software. The cubic interpolation method is used to obtain the displacements of the scatterers from the nodes obtained by FEM. These scatterers are utilized to simulate ultrasound images using the Field II toolbox [33], [34] with a center frequency of 5 MHz.

24 different phantoms with 10 different average strain values and 10 different random scatterer realizations with different positions are simulated (for each phantom 100 images are simulated with a total of 2400 images). 1000 image pairs are randomly sampled from the mentioned simulated images for training. The test set contains 70 image pairs and it has four different models. The test phantoms have inclusions that differ from training phantom in size, location and shape with average strain values between 1 to 2.5 %. We publicly release this dataset as part of this manuscript at data.sonography.ai.

D. Experimental Phantom and In vivo Data

Phantom data is collected at Concordia University's PERFORM Centre by an E-Cube R12 research ultrasound machine (Alpinion, Bothell, WA, USA) with a L3-12H linear array at the center frequency of 10 MHz and sampling frequency of 40 MHz. A tissue mimicking breast phantom made by Zerdine (Model 059, CIRS: Tissue Simulation & Phantom Technology, Norfolk, VA) is used which has tissue elasticity of $20 \pm 5kPa$ and contains hard inclusions with elasticity at least twice the elasticity of the tissue.

In vivo data was obtained at Johns Hopkins Hospital from a research Antares Siemens system using a VF 10-5 linear array with a center frequency of 6.67 MHz and a sampling frequency of 40 MHz. Data is collected from three patients in open-surgical RF thermal ablation for liver cancer. More experimental details of the procedure can be found in [15]. The study was approved by the institutional review board with consent of all patients.

E. Fine-Tuning of the Network

It is common to fine-tune a network that is already trained on a similar task, as opposed to training it from scratch, a process also known as transfer learning [36], [37]. Therefore, we use the FEM and Field II dataset to fine-tune the proposed networks, which are trained on computer vision data. We tested many settings and fine-tuning strategies and found out that only fine-tuning the final resolution pyramid suffices since the network already performs well and only small improvement to the displacement prediction is required. The fine-tuned sub-network is specified in Fig. 2. Data augmentation is performed by randomly mirroring in lateral direction and adding white Gaussian noise to the RF data. Subsequently, envelope and B-mode images are obtained and used as different input channels of CNNs.

Regarding the loss function selection, due to the fact that displacement error is small, MSE suppresses this small error and amplifies the outlier regions. In practice, we obtained noisier strain by MSE even though the displacement error was reduced (higher displacement variance with lower displacement error). Therefore, we use norm 0.4 similar to FlowNet2 small displacement network [26] as the main loss function since this norm amplifies small error and attenuates large errors obtained by outliers. Another important point is that Total Variation (TV) regularization similar to [16], [38] is used to reduce the displacement variations and improve the quality of the strain. The final loss function used for fine-tuning is:

$$loss = \|D_{GT} - D_{\Theta}\|_{0.4} + \frac{\lambda}{N}\|\Delta D_{\Theta} - \varepsilon\|_1 + \gamma\|\Theta\|_2 \quad (2)$$

where D_{GT} and D_{Θ} are the ground truth and estimated displacements, respectively and $\|.\|_p$ denotes norm p. ΔD_{Θ} is the axial derivative of the predicted axial displacement, N is the number of samples used for TV computation, and λ, γ are regularization weights. To avoid underestimation bias due to regularization, we regularize by average strain (ε) similar to [15], [16]. We fine-tune the weights of the final pyramid of RFMPWC-Net using this simulation dataset. We also fine-tune MPWC-Net, but do not report the results in this manuscript since fine-tuned RFMPWC-Net performed better than MPWC-Net. We set the weight decay to 0.01 and λ to 0.2. NVIDIA Titan V with 12 GB RAM is used for training and the image size is 2048×256, which enforces us to use batch size of 1 due to memory limits. The network is fine-tuned for 50 epochs and the learning rate is set to 2e-9.

III. Results

In this section, the proposed networks are evaluated and compared with existing methods. NCC [9], GLUE [14], FlowNet2 [21], [26], original PWC-Net [24], [28] and our proposed networks (MPWC-Net, RFMPWC-Net and fine-tuned RFMPWC-Net) are evaluated for simulated phantoms, an experimental phantom and *in vivo* data. GLUE is a recent method that has already been extensively used in several challenging simulation, phantom and *in vivo* applications by different research groups [30], [39]–[41].

To make the comparison fair, the input of deep learning methods (PWC-Net, FlowNet, MPWC-Net and RFMPWC-Net) is the concatenation of B-mode, RF and envelope signals. We use the trained FlowNet2 and PWC-Net weights publicly available on the Pytorch framework [24]. The GLUE code is publicly available, and NCC implementation is similar to [16] where we perform 2D cubic interpolation to calculate subsample displacements. Substantially better results are expected with a multi-level stretching NCC technique. In the simulation experiments, the ground truth displacement is known. Therefore, Normalized Root Mean Squared Error (NRMSE) of axial displacement [42] defined in 3 is used as the metric for measuring the displacement prediction accuracy. The results are reported for two different Peak Signal to Noise Ratios (PSNR).

$$NRMSE(\%) = \sqrt{mean((\frac{D_{GT} - D_{\Theta}}{D_{GT}})^2)} \times 100 \quad (3)$$

$$PSNR = 20 \times log_{10}(\frac{I_{max}}{\delta}) \quad (4)$$

where δ denotes standard deviation of noise and I_{max} is the maximum of image intensity. Noise with normal distribution is added to the RF data in order to obtain noisy simulation images. It should be noted that NRMSE is computed for each test phantom, then mean and standard deviation of NRMSE are reported for ideal and noisy simulated phantoms.

Two popular metrics, Contrast to Noise Ratio (CNR) and Strain Ratio (SR) are also used to show the strain quality in the experimental and *in vivo* results, which are defined as [7]:

$$SR = \frac{\overline{s}_t}{\overline{s}_b}, \qquad CNR = \sqrt{\frac{2(\overline{s}_b - \overline{s}_t)^2}{{\sigma_b}^2 + {\sigma_t}^2}}, \quad (5)$$

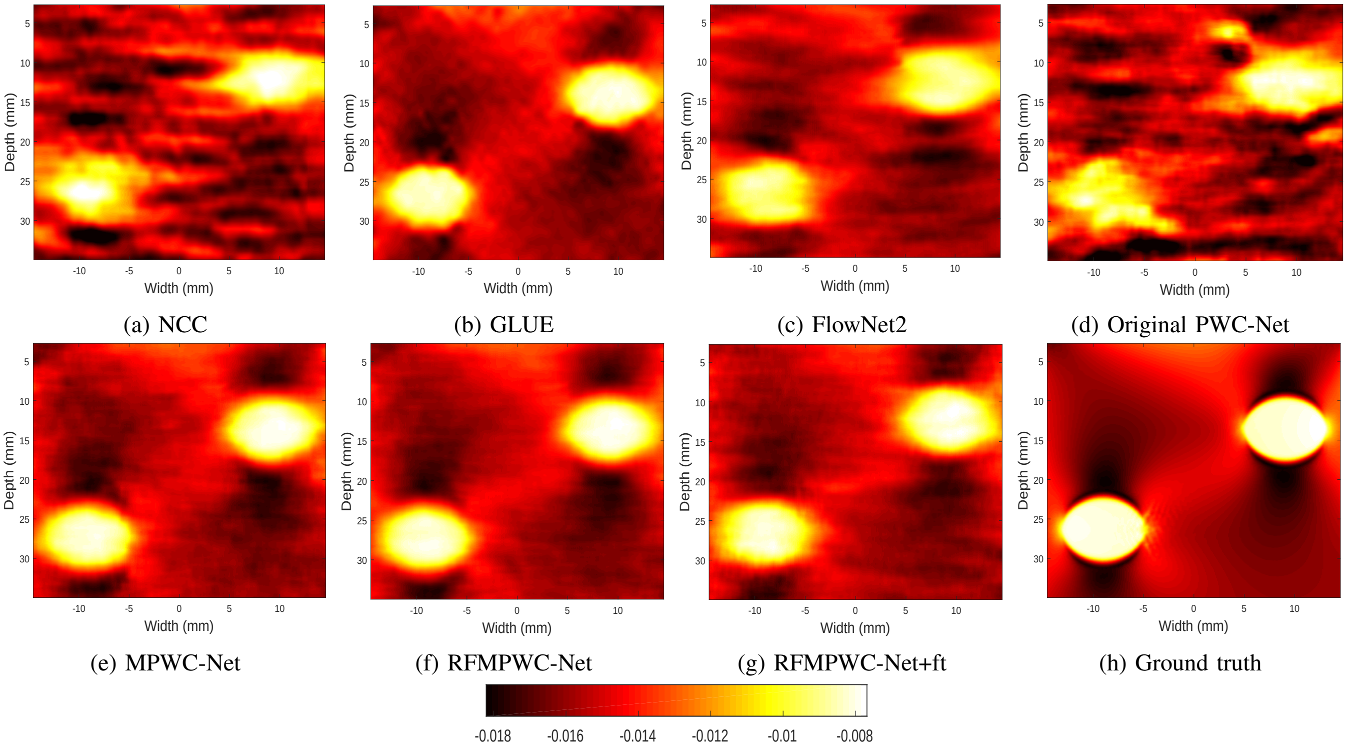
where $\overline{s}_t$ and $\overline{s}_b$ are average values of strain in the target and background regions, and σ_t and σ_b are variance values of strain in the target and background regions, respectively. The selected regions in the target and background must be uniform and large enough to be statistically meaningful. It is important to note that CNR is sensitive to mean and variance of the regions. Whereas, SR only measures the differences in the mean value of the selected region. SR is a proper metric to measure the bias error of the strain. Whereas, CNR shows the combination of bias and variance error of the strain. One basic property of elastography methods is that they estimate lower difference between the tissue and the inclusion due to bias created by different smoothing operations (continuity constraints in GLUE, median filtering or low-pass filtering in NCC and window-based methods, and least squares differentiation). Therefore, in real experiments with unknown ground truth, higher difference usually translates to smaller estimation bias. If a hard inclusion is chosen as the target, the value of SR is less than 1, where lower SR represents higher difference in the strain of the target and background (i.e. lower numbers are generally better). In order to compute reliable CNR and SR, large windows are selected in Fig. 5 (h), Fig. 6 (h) and Fig. 7 (h). The windows are divided into small overlapping patches. CNR and SR are computed for all combination of target and background patches. The mean and standard deviation of the computed CNRs and SRs are reported. To better visualize the results, we show strain images, which are the least squares derivatives of the axial displacement in axial direction.

A. Simulation Results

In this section the results of the simulated phantoms are presented for different methods. The strain image of a simulated phantom with the displacement calculated by the evaluated methods is depicted in Fig. 3 for PSNR= ∞. Our proposed networks perform substantially better than stock deep learning methods in both simulation setups. Please refer to Supplementary Material for more results including a different simulated phantom with added noise.

It is important to note that although the complexity of FlowNet2 is substantially more than the four other networks (both in training and inference), its results are substantially worse than our proposed networks. By closely inspecting the FlowNet2 results, it is evident that there is a substantial underestimation of strain in hard inclusions, which are not as dark as our proposed methods.

Another important point is that all networks except RFMPWC-Net+ft are trained on computer vision images and RFMPWC-Net+ft is fine-tuned by our dataset. Visually, the results of RFMPWC-Net+ft and RFMPWC-Net are close but RFMPWC-Net+ft is smoother. The quantitative results are



(a) NCC (b) GLUE (c) FlowNet2 (d) Original PWC-Net

(e) MPWC-Net (f) RFMPWC-Net (g) RFMPWC-Net+ft (h) Ground truth

Fig. 3: Strain images of a simulated phantom with $PSNR = \infty$.

TABLE II: Comparison of different methods for 70 simulated phantoms.

	PSNR=∞	PSNR=30 dB
Method	NRMSE (%)	NRMSE (%)
NCC	1.88±0.51	1.93±0.53
GLUE	**1.10**±0.53	**1.10**±0.53
FlowNet2	1.65±0.46	1.68±0.46
PWC-Net	1.82±0.74	1.82±0.74
MPWC-Net	1.17±0.54	1.28±0.40
RFMPWC-Net	1.18±0.61	1.19±0.62
RFMPWC-Net+ft	1.15±**0.33**	1.18±**0.34**

given in Table II for 70 simulated phantoms. According to these results, the results of RFMPWC-Net are close to GLUE even without fine-tuning on ultrasound images, which shows the potential of the networks solely trained on computer vision images. Our fine-tuned variant of RFMPWC-Net performs slightly better than RFMPWC-Net. It is important to note that GLUE results remain very similar for no-noise and noisy conditions which indicates the robustness of GLUE due to optimizing all samples simultaneously.

RFMPWC-Net is more robust to noise compared to MPWC-Net which NRMSE increases 0.09 % in noisy conditions. In order to show the effect of fine-tuning, the strain of one line using RFMPWC-Net and RFMPWC-Net+ft is depicted in Fig. 4. As shown, the fine-tuned RFMPWC-Net result (red) has less variations and it is closer to ground truth compared to RFMPWC-Net (blue), which indicates that fine-tuning improves the displacement estimation accuracy. However, the improvements obtained by modifying the structure is more tangible (compare Fig. 3(d), (e) and (f)) than fine-tuning. The main reason is that the networks trained on computer vision images are already performing well in mapping the inputs to the displacement. The modification of the structure brings substantial improvements to the network accuracy by providing more information to the network. Please refer to the supplementary material for fine-tuning of the original PWC-Net.

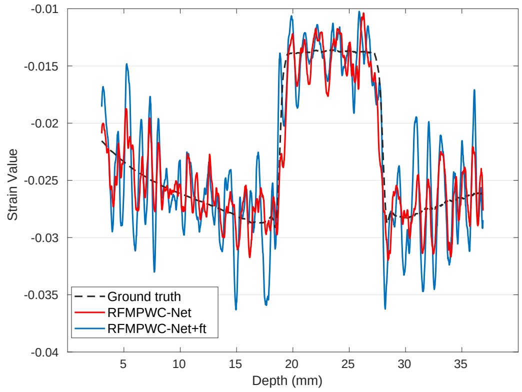
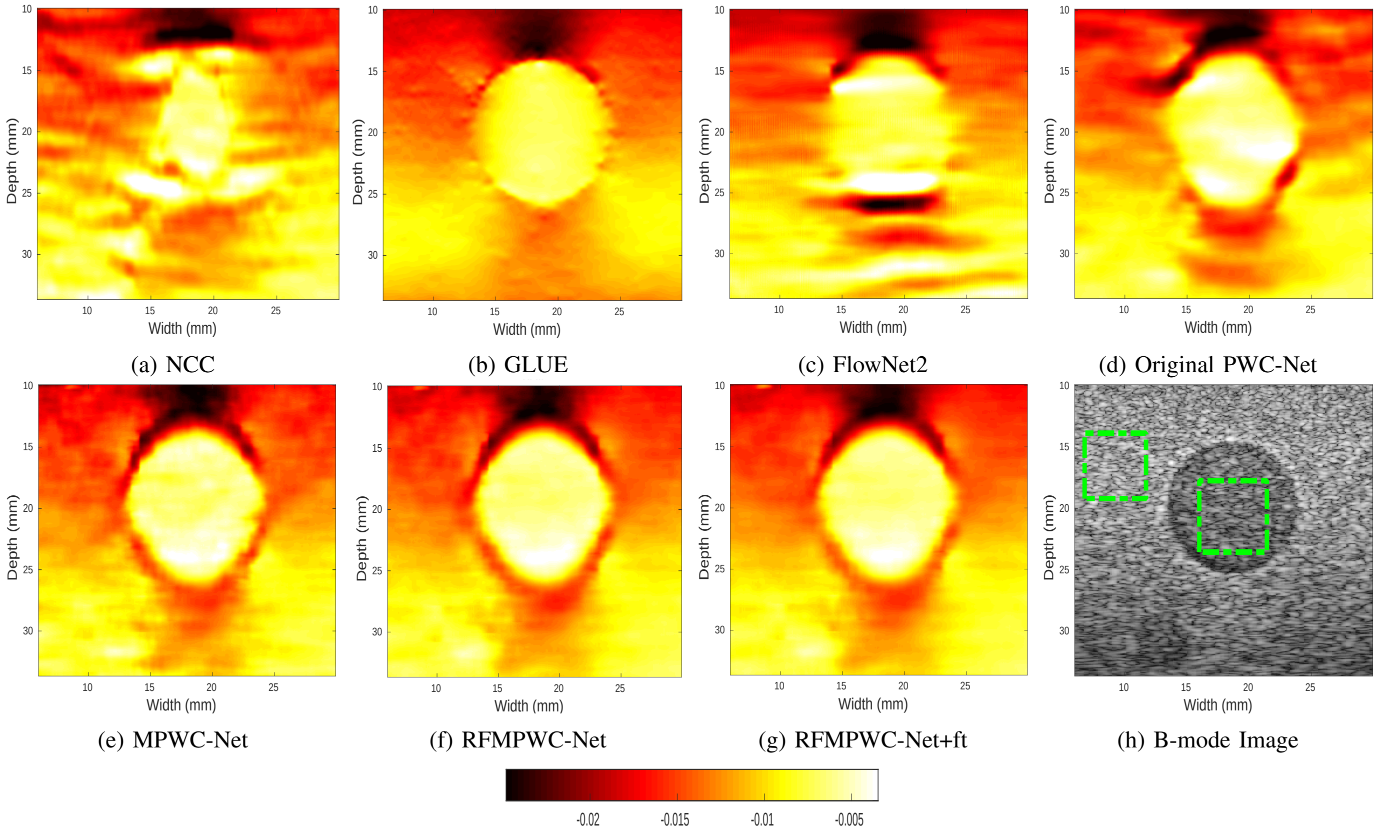


Fig. 4: One line of strain using a small least square window. RFMPWC-Net (blue), RFMPWC-Net+ft (red) and ground truth (black).

B. Experimental Phantom Results

CNR and SR defined in Eq 5 are used as quantitative metrics and the visual results are demonstrated in Fig. 5.



(a) NCC

(b) GLUE

(c) FlowNet2

(d) Original PWC-Net

(e) MPWC-Net

(f) RFMPWC-Net

(g) RFMPWC-Net+ft

(h) B-mode Image

Fig. 5: Strain images of the experimental phantom. The windows used for CNR and SR computation are highlighted in the B-mode image (h). FlowNet2 (c) has high heterogeneity and fails to obtain smooth and high quality strain and the proposed networks have higher contrast compared to GLUE (b).

TABLE III: SR and CNR of the experimental phantom.

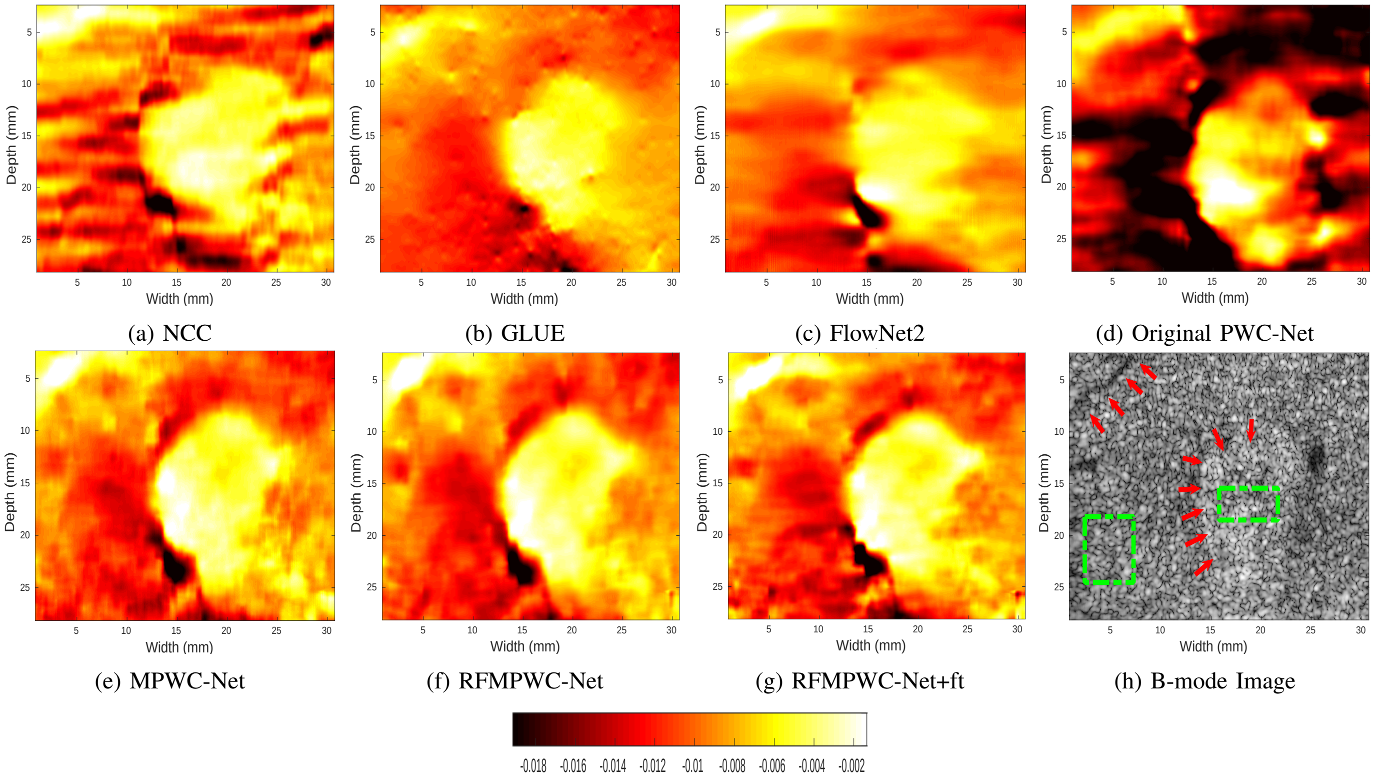
Method	CNR	SR
NCC	10.65±2.69	0.399±0.04
GLUE	26.75±7.86	0.459±0.02
FlowNet2	20.19±3.70	0.48±0.02
PWC-Net	20.28±5.82	**0.376**±0.05
MPWC-Net	17.12±4.59	0.425±0.03
RFMPWC-Net	27.06±4.28	0.410±0.03
RFMPWC-Net+ft	**29.15**±5.77	0.382±0.05

NCC and FlowNet2 fail to obtain acceptable strain and GLUE produces smooth but underestimated strain, which is due to regularization. As such, GLUE result does not have as low SR as the deep learning methods. Nevertheless, GLUE has less variance, which makes the CNR very close to our proposed methods. The quantitative results in Table III. confirm the visual assessments. GLUE has good CNR (26.75) but poor SR (0.459), whereas PWC-Net has the best SR (0.376) with a moderate CNR (20.28). RFMPWC-Net has higher CNR and better SR than MPWC-Net. RFMPWC-Net has higher CNR than GLUE (27.06 compared to 26.75) and better SR (0.41 compared to 0.459) without using any ultrasound images for training, which indicates the strength of the proposed CNN networks. RFMPWC-net+ft produces the most appealing result among deep learning methods and outperforms all evaluated methods in terms of CNR (29.15) and has good SR (0.382). This shows that fine-tuning of trained networks by ultrasound images has a positive impact on the performance of the network.

C. *In vivo Results*

Considering Fig. 6, GLUE estimates low-variance and high quality but blurry strain. The strain estimated by FlowNet2 is too smooth and many details are lost. PWC-Net also fails to estimate an acceptable strain. MPWC-Net has good strain quality but with a few artifacts, and RFMPWC-Net generates the best. RFMPWC-Net+ft further improves strain quality compared to RFMPWC-Net. Regarding Fig. 7, the GLUE result is acceptable but it is over smooth especially in in the top right of image. NCC, FlowNet2 and PWC-Net all fail to estimate strain, and MPWC-Net obtains a high-quality strain compared to PWC-Net. This indicates that our changes in the structure of PWC-Net have substantial impact on the network's performance. RFMPWC-Net has better strain compared to MPWC-Net and most of artifacts are removed in the RFMPWC-Net result. RFMPWC-Net+ft produces a very high-quality strain image and further removes the artifacts.

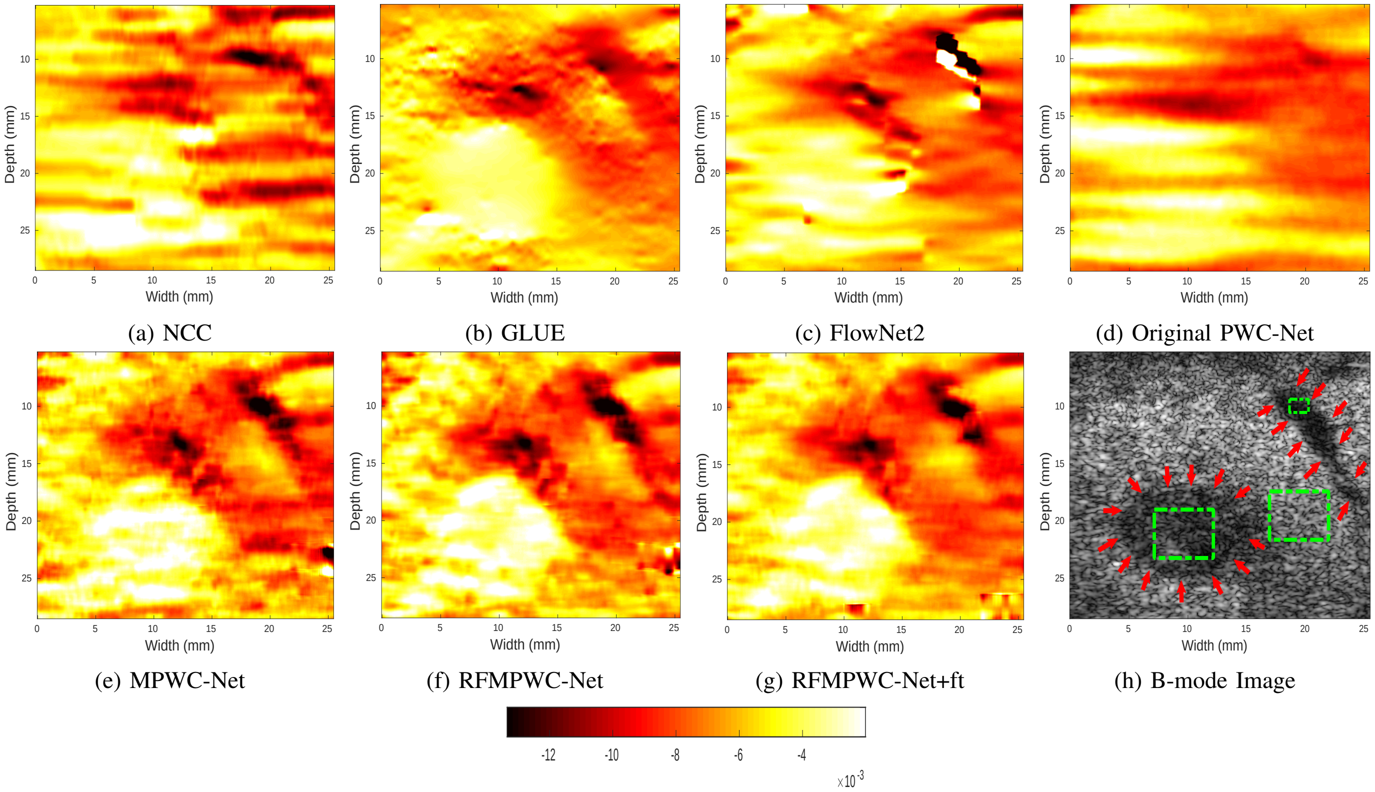
Considering the quantitative results of tumor presented in the first two columns of Table IV, GLUE obtains the high CNR in both patients (19.36 and 15.11) but the SR is poor (0.389 and 0.441). NCC and PWC-Net have poor CNR and



(a) NCC (b) GLUE (c) FlowNet2 (d) Original PWC-Net

(e) MPWC-Net (f) RFMPWC-Net (g) RFMPWC-Net+ft (h) B-mode Image

Fig. 6: *In vivo* strain results of the liver of patient 1 before ablation. The tumors are marked with arrows and the windows used for CNR and SR computation are highlighted in the B-mode image (h).



(a) NCC (b) GLUE (c) FlowNet2 (d) Original PWC-Net

(e) MPWC-Net (f) RFMPWC-Net (g) RFMPWC-Net+ft (h) B-mode Image

Fig. 7: *In vivo* strain results of the liver of patient 2 before ablation. The tumor and the vein are marked with arrows and the windows used for CNR and SR computation are highlighted in the B-mode image (h). the GLUE (b) obtains smooth but blurry strain especially close to the vein on the top right of the image. Fine-tuning reduces the artifacts presented in RFMPWC-Net (compare (f) and (g)).

TABLE IV: Results of *In vivo* data, patient 1 (Fig. 6) and patient 2 (Fig. 7). GLUE has higher CNR for tumor and RFMPWC-Net results in higher CNR for the the vein. The proposed networks perform comparable to GLUE for *in vivo* data.

	Patient 1		Patient 2 (tumor)		Patient 2 (vein)	
Method	CNR	SR	CNR	SR	CNR	$\frac{1}{SR}$
NCC	9.08±3.22	0.29±0.07	3.60±1.41	0.51±0.09	11.84±6.66	0.590±0.14
GLUE	**19.36**±4.51	0.389±0.06	15.11±5.30	0.441±0.02	11.54±6.13	0.795±0.07
FlowNet2	12.86±0.46	0.463±0.049	9.40±2.04	0.415±0.05	fail	fail
PWC-Net	10.79±4.00	0.451±0.09	5.90± 2.45	0.587±0.09	9.19±4.34	0.835±0.07
MPWC-Net	12.11±3.75	**0.376**±0.07	11.66± 2.2	**0.338**±0.03	11.98±5.69	0.610±0.06
RFMPWC-Net	13.55±4.34	0.396±0.06	12.48±3.23	0.409±0.04	**19.88**±9.41	**0.590**±0.05
RFMPWC-Net+ft	16.63±5.53	0.380±0.05	**15.58**±2.58	0.395±0.04	12.52±3.71	0.601±0.08

FlowNet2 has higher CNR compared to them but visually the strain images are not acceptable. MPWC-Net has poor CNR (12.11 and 11.66) but produces the best SR (0.376 and 0.338). This implies that MPWC-Net has high variance in estimation which leads to low CNR but it has low bias in estimation which results in low SR. RFMPWC-Net outperforms MPWC-Net in terms of CNR with slightly worse SR. Fine-tuning improves the CNR with approximately similar SR. RFMPWC-Net+ft produces CNR values very close or even better than GLUE (16.63 and 15.58) with better SR (0.388 and 0.399).

By inspecting the results of the soft target (the vein in up right corner of Fig 7 (h)), it is inferred that our 3 networks substantially outperform GLUE in terms of both CNR and SR. RFMPWC-Net has the highest CNR (19.88) by a large margin, which is 8.34 dB and 7.36 dB better than GLUE and RFMPWC-Net+ft, respectively. The main reason that RFMPWC-Net performs better without fine-tuning is that our database only contains hard inclusions and fine-tuning by this database deteriorates the cases with soft inclusions such as veins. For the vein, $\frac{1}{SR}$ is reported in order to be consistent with other results since the SR value for veins is more than 1. Our networks have the best SR among the compared methods and they have substantially better SR compared to GLUE.

D. Effect of sampling and center frequencies

The sampling and the center frequencies have critical role in displacement estimation accuracy. In the simulation results, the center and sampling frequency are 5 MHz and 50 MHz, respectively. We simulate a phantom with two different center and sampling frequencies. RFMPWC-Net, FlowNet2 and PWC-Net are tested for the center frequencies 5 and 10 MHz and the sampling frequencies 25 and 50 MHz. As shown in Fig. 8, strain obtained by RFMPWCNet (a, d) are high quality and consistent compared to FlowNet2 and PWC-Net. Please refer to the supplementary material for more results.

IV. Discussions

In this paper, two networks based on PWC-Net are proposed for USE. Generally, USE requires high accuracy subsample displacement estimation, which renders efficient use of high frequency information in RF data critical. This is a challenge as stock optical flow networks are not designed to handle RF data.

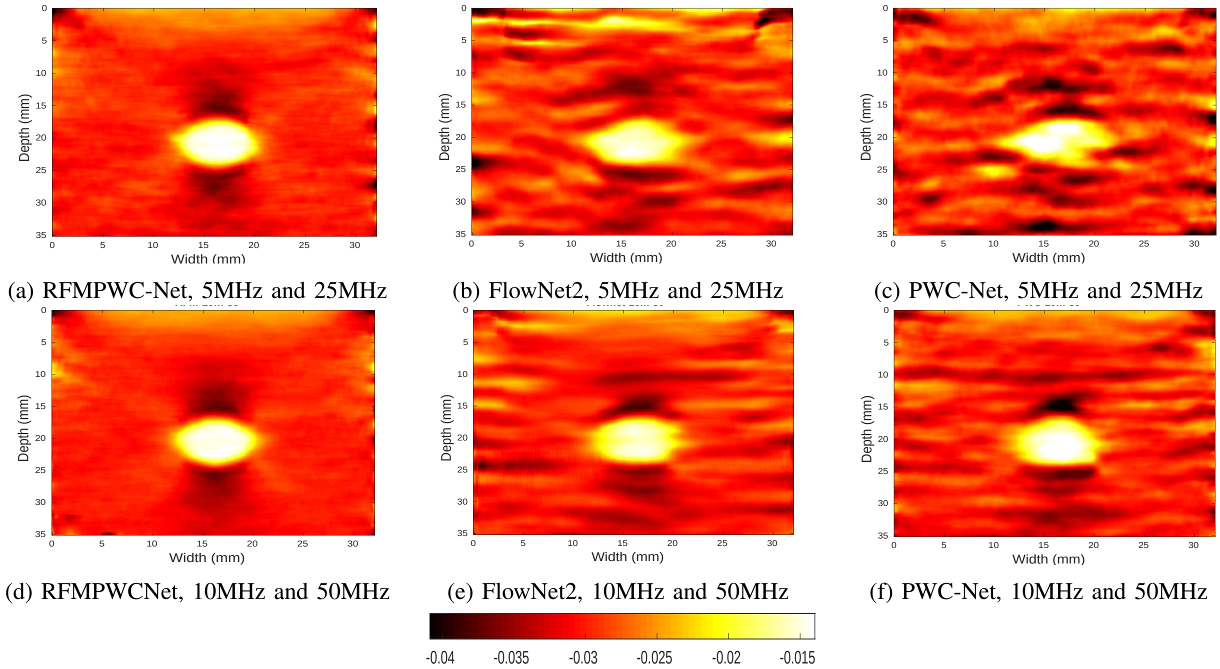
The PWC-Net is modified for USE displacement estimation by: 1) removing downsampling of the first feature extraction layer (this layer is connected to the input directly) to prevent loss of high frequency information; and 2) concatenating RF data, envelope and B-mode images to feed to the network. by doing this, the low-resolution pyramid levels exploit low-frequency B-mode and envelope information and high-resolution pyramid levels use RF data to obtain accurate displacement.

The main drawback of MPWC-Net is that B-mode and envelope contribute to the final resolution displacement estimation. B-mode and envelope are beneficial in low pyramid levels where RF data cannot be used, but they result in less accurate estimated displacement compared to RF data. Hence, in noisy conditions, MPWC results degrade considerably (as given in Table II). RFMPWC-Net is proposed to resolve this problem by adding a separate sub-network to extract and use only RF data for the final pyramid level.

FlowNet2 network, which is extensively used by the researchers, obtains under-estimated strain and fails for *in vivo* data. Although FlowNet2 has 18 times more learning parameters than PWC-Net and achieves high accuracy in computer vision databases such as MPI-Sintel [26], it performs poorly on ultrasound images. This emphasizes that less complex pyramidal and warping structure is more suitable for ultrasound data.

Fine-tuning is another avenue that is investigated in this paper, where the networks are tuned by simulated ultrasound images. In the loss function of fine-tuning, TV regularization is used to reduce the variance of displacement estimation. According to our results, fine-tuning improves the strain quality both qualitatively and quantitatively. All ultrasound simulation training data for fine-tuning the networks contain harder inclusions than the background. Nevertheless, the fine-tuned network performed well in a variety of *in vivo* experiments with different kinds of tissue. In the future, we plan to add data with soft inclusions in our training database and expect this to further improve the results. These new simulations will also strengthen the online database that we released.

Another important point about fine-tuning is that we only consider negative strain (post-compression image is the second image) for fine-tuning. The network can be fine-tuned with



(a) RFMPWC-Net, 5MHz and 25MHz
(b) FlowNet2, 5MHz and 25MHz
(c) PWC-Net, 5MHz and 25MHz
(d) RFMPWCNet, 10MHz and 50MHz
(e) FlowNet2, 10MHz and 50MHz
(f) PWC-Net, 10MHz and 50MHz

Fig. 8: Simulation results of RFMPWC-Net, FlowNet2 and PWC-Net for different center and sampling frequencies. (Network, Center frequency and Sampling frequency). RFMPWC-Net quality remains well when sampling frequency decreased (a) or center frequency increased (d) in comparison to the other methods.

both positive and negative strain to be used for cases which post-compression image is not determined. Please refer to the supplementary material for more information.

It is also worth mentioning that our proposed networks are very close to GLUE in terms of CNR and have better SR. By comparing the quantitative results presented in Table III and IV, it can be seen that GLUE has higher CNR than our proposed methods in the tumor part of patient 1. However, GLUE has lower CNR than the fine-tuned network for the experimental phantom data and data of patient 2. Another interesting conclusion is that RFMPWC-Net outperforms GLUE and fine-tuned network by a large margin for the vein (19.88 compared to 11.52 and 12.54). The reason for outperforming the fine-tuned network can be explained by the fact that we performed transfer learning using simulation data that only has hard inclusions.

In terms of SR, our proposed methods are the best among compared methods. MPWC-Net has the best SR but moderate CNR. In contrast, RFMPWC-Net and the fine-tuned variant of the network have higher CNR and slightly worse SR compared to MPWC-Net. The proposed methods perform similar to recent elastography methods without any need for parameter tuning, and have very small memory footprints and can be implemented on inexpensive GPUs.

V. Conclusion

This paper presents a deep learning approach for displacement estimation of the USE. The structure of PWC-Net is modified for our application. Visual and quantitative assessments of simulated phantoms, experimental phantom and *in vivo* data confirm that the proposed methods are suitable for USE and can compete with current state-of-the-art elastography methods.

Acknowledgment

We thank anonymous reviewers for their constructive feedback. This research is funded by NSERC Discovery grant RGPIN 04136. The authors would like to thank Drs. E. Boctor, M. Choti and G. Hager for allowing us to use the *in vivo* data. We also thank M. Mirzaei for valuable discussions, and NVIDIA for donation of the Titan V GPU.

References

[1] T. J. Hall, Y. Zhu, C. S. Spalding, and L. T. Cook, "In vivo results of real-time freehand elasticity imaging," in *2001 IEEE Ultrasonics Symposium. Proceedings. An International Symposium (Cat. No. 01CH37263)*, vol. 2. IEEE, 2001, pp. 1653–1657.

[2] R. M. Sigrist, J. Liau, A. El Kaffas, M. C. Chammas, and J. K. Willmann, "Ultrasound elastography: review of techniques and clinical applications," *Theranostics*, vol. 7, no. 5, p. 1303, 2017.

[3] A. Kling and J. Jiang, "Potential of determining thermal dose for ablation therapies using ultrasound elastography: An ex vivo feasibility study," in *2018 IEEE International Ultrasonics Symposium (IUS)*, 2018, pp. 1–4.

[4] J. Jiang, T. Varghese, C. L. Brace, E. L. Madsen, T. J. Hall, S. Bharat, M. A. Hobson, J. A. Zagzebski, and F. T. Lee, "Young's modulus reconstruction for radio-frequency ablation electrode-induced displacement fields: a feasibility study," *IEEE transactions on medical imaging*, vol. 28, no. 8, pp. 1325–1334, 2009.

[5] F.-F. Lee, Q. He, J. Gao, A. Pan, S. Sun, X. Liang, and J. Luo, "Evaluating hifu-mediated local drug release using thermal strain imaging: Phantom and preliminary in-vivo studies," *Medical physics*, 2019.

[6] H. Zhi, B. Ou, B.-M. Luo, X. Feng, Y.-L. Wen, and H.-Y. Yang, "Comparison of ultrasound elastography, mammography, and sonography in the diagnosis of solid breast lesions," *Journal of ultrasound in medicine*, vol. 26, no. 6, pp. 807–815, 2007.

[7] J. Ophir, S. K. Alam, B. Garra, F. Kallel, E. Konofagou, T. Krouskop, and T. Varghese, "Elastography: ultrasonic estimation and imaging of the elastic properties of tissues," *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, vol. 213, no. 3, pp. 203–233, 1999.

[8] T. Varghese, E. Konofagou, J. Ophir, S. Alam, and M. Bilgen, "Direct strain estimation in elastography using spectral cross-correlation," *Ultrasound in medicine & biology*, vol. 26, no. 9, pp. 1525–1537, 2000.

[9] A. Nahiyan and M. K. Hasan, "Hybrid algorithm for elastography to visualize both solid and fluid-filled lesions," *Ultrasound in medicine & biology*, vol. 41, no. 4, pp. 1058–1078, 2015.

[10] J. Luo and E. Konofagou, "A fast normalized cross-correlation calculation method for motion estimation," *IEEE Trans ultrasonics, ferroelectrics, and frequency control*, vol. 57, no. 6, pp. 1347–1357, 2010.

[11] J. Jiang and T. J. Hall, "A parallelizable real-time motion tracking algorithm with applications to ultrasonic strain imaging," *Physics in Medicine & Biology*, vol. 52, no. 13, p. 3773, 2007.

[12] M. Mirzaei, A. Asif, M. Fortin, and H. Rivaz, "3d normalized cross-correlation for estimation of the displacement field in ultrasound elastography," *Ultrasonics*, vol. 102, p. 106053, 2020.

[13] T. J Hall, P. E Barboneg, A. A Oberai, J. Jiang, J.-F. Dord, S. Goenezen, and T. G Fisher, "Recent results in nonlinear strain and modulus imaging," *Current medical imaging reviews*, vol. 7, pp. 313–327, 2011.

[14] H. S. Hashemi and H. Rivaz, "Global time-delay estimation in ultrasound elastography," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 64, no. 10, pp. 1625–1636, 2017.

[15] H. Rivaz, E. M. Boctor, M. A. Choti, and G. D. Hager, "Real-time regularized ultrasound elastography," *IEEE transactions on medical imaging*, vol. 30, no. 4, pp. 928–945, 2010.

[16] M. Mirzaei, A. Asif, and H. Rivaz, "Combining total variation regularization with window-based time delay estimation in ultrasound elastography." *IEEE transactions on medical imaging*, 2019.

[17] M. Ashikuzzaman, C. J. Gauthier, and H. Rivaz, "Global ultrasound elastography in spatial and temporal domains," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 66, no. 5, pp. 876–887, 2019.

[18] A. K. Z. Tehrani, M. Mozaffarzadeh, Z. Mardi, S. H. Hozhabr, M. Mehrmohammadi, and B. Makkiabadi, "Application of demons algorithm in ultrasound elastography using b-mode ultrasound images," in *Photons Plus Ultrasound: Imaging and Sensing 2019*, vol. 10878. International Society for Optics and Photonics, 2019, p. 108786P.

[19] M. G. Kibria and H. Rivaz, "Gluenet: Ultrasound elastography using convolutional neural network," in *Simulation, Image Processing, and Ultrasound Systems for Assisted Diagnosis and Navigation*. Springer, 2018, pp. 21–28.

[20] Z. Gao, S. Wu, Z. Liu, J. Luo, H. Zhang, M. Gong, and S. Li, "Learning the implicit strain reconstruction in ultrasound elastography using privileged information," *Medical image analysis*, vol. 58, pp. 11–18, 2019.

[21] B. Peng, Y. Xian, and J. Jiang, "A convolution neural network-based speckle tracking method for ultrasound elastography," in *2018 IEEE International Ultrasonics Symposium (IUS)*. IEEE, 2018, pp. 206–212.

[22] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[23] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[24] D. Sun, X. Yang, M.-Y. Liu, and J. Kautz, "Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8934–8943.

[25] A. Dosovitskiy, P. Fischer, E. Ilg, P. Hausser, C. Hazirbas, V. Golkov, P. Van Der Smagt, D. Cremers, and T. Brox, "Flownet: Learning optical flow with convolutional networks," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 2758–2766.

[26] E. Ilg, N. Mayer, T. Saikia, M. Keuper, A. Dosovitskiy, and T. Brox, "Flownet 2.0: Evolution of optical flow estimation with deep networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2462–2470.

[27] A. Ranjan and M. J. Black, "Optical flow estimation using a spatial pyramid network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4161–4170.

[28] D. Sun, X. Yang, M.-Y. Liu, and J. Kautz, "PWC-Net: CNNs for optical flow using pyramid, warping, and cost volume," *arXiv preprint arXiv:1709.02371*, 2017.

[29] S. Wu, Z. Gao, Z. Liu, J. Luo, H. Zhang, and S. Li, "Direct reconstruction of ultrasound elastography using an end-to-end deep neural network," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018, pp. 374–382.

[30] C. Hoerig, J. Ghaboussi, and M. F. Insana, "Data-driven elasticity imaging using cartesian neural network constitutive models and the autoprogressive method," *IEEE transactions on medical imaging*, vol. 38, no. 5, pp. 1150–1160, 2018.

[31] X. Pan, K. Liu, J. Shao, J. Gao, L. Huang, J. Bai, and J. Luo, "Performance comparison of rigid and affine models for motion estimation using ultrasound radio-frequency signals," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 62, no. 11, pp. 1928–1943, 2015.

[32] M. Kibria and H. Rivaz, "Global ultrasound elastography using convolutional neural network," *arXiv preprint arXiv:1805.07493*, 2018.

[33] J. A. Jensen, "Field: A program for simulating ultrasound systems," in *10TH Norticbaltic Conference On Biomedical Imaging, Vol. 4, Suppliment 1, Part 1: 351–353*. Citeseer, 1996.

[34] J. A. Jensen and N. B. Svendsen, "Calculation of pressure fields from arbitrarily shaped, apodized, and excited ultrasound transducers," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 39, no. 2, pp. 262–267, 1992.

[35] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE Trans. pattern analysis machine intelligence*, vol. 40, no. 4, pp. 834–848, 2017.

[36] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" in *Advances in neural information processing systems*, 2014, pp. 3320–3328.

[37] V. Iglovikov and A. Shvets, "Ternausnet: U-net with vgg11 encoder pre-trained on imagenet for image segmentation," *arXiv preprint arXiv:1801.05746*, 2018.

[38] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *European conference on computer vision*. Springer, 2016, pp. 694–711.

[39] S. Rezajoo and A. R. Sharafat, "Robust estimation of displacement in real-time freehand ultrasound strain imaging," *IEEE transactions on medical imaging*, vol. 37, no. 7, pp. 1664–1677, 2018.

[40] C. Rabin and N. Benech, "Quantitative breast elastography from b-mode images," *Medical physics*, 2019.

[41] T. Ersepke, T. C. Kranemann, and G. Schmitz, "On the performance of time domain displacement estimators for magnetomotive ultrasound imaging," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 66, no. 5, pp. 911–921, 2019.

[42] M. T. Islam, A. Chaudhry, S. Tang, E. Tasciotti, and R. Righetti, "A new method for estimating the effective poisson's ratio in ultrasound poroelastography," *IEEE transactions on medical imaging*, vol. 37, no. 5, pp. 1178–1191, 2018.

Ali K. Z. Tehrani was born in Tehran, Iran. He received his BSc from Azad University south branch and MASc from Amirkabir University, Tehran, Iran. He is currently pursuing his PhD in electrical and computer engineering at IMPACT lab at Concordia University, Montreal, Canada. His research interest includes medical image analysis, machine learning and elastography.

Hassan Rivaz directs the IMPACT lab: IMage Processing And Characterization of Tissue, and is a Concordia University Research Chair in Medical Image Analysis. His research interests are medical image analysis, machine learning and elastography. He is an Associate Editor of IEEE Transactions on Medical Imaging, and IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control. He has served as an Area Chair of MICCAI 2017, 2018, 2019 and 2020, and is a member of the organizing committee of IEEE EMBC 2020.